



HAL
open science

Modélisation de la croissance postnatale des enfants pour l'étude de ses déterminants précoces

Sophie Carles

► **To cite this version:**

Sophie Carles. Modélisation de la croissance postnatale des enfants pour l'étude de ses déterminants précoces. Santé publique et épidémiologie. Université Paris Saclay (COMUE), 2016. Français. NNT : 2016SACLS065 . tel-03177183

HAL Id: tel-03177183

<https://theses.hal.science/tel-03177183>

Submitted on 23 Mar 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.

NNT : 2016SACLS065

THESE DE DOCTORAT
DE
L'UNIVERSITE PARIS-SACLAY
PREPAREE A
"L'UNIVERSITE PARIS-SUD"

ECOLE DOCTORALE N° 570
Ecole Doctorale de Santé Publique (EDSP)

Spécialité de doctorat : Santé publique - épidémiologie

Par

Melle Carles Sophie

Modélisation de la croissance postnatale des enfants
pour l'étude de ses déterminants précoces

Thèse présentée et soutenue à Villejuif, le 22 Mars 2016 :

Composition du Jury :

M, Gajdos, vincent	PU-PH, Université Paris-Sud	Président
Mme, Comets, emmanuelle	CR, INSERM	Rapporteur
Mme, Leclercq-Samson, adeline	PU, Université Grenoble Alpes	Rapporteur
Mme, Costet, nathalie	Ingénieur de recherche, INSERM	Examineur
Mme, Proust-Lima, cécile	CR, INSERM	Examineur
M, Botton, jérémie	MCF, Université Paris-Sud, INSERM	Co-directeur de thèse
Mme, Charles, marie-aline	DR, INSERM	Directrice de thèse

REMERCIEMENTS

En préambule à ce travail, je souhaiterais adresser mes remerciements les plus sincères aux personnes qui m'ont apporté leur aide et qui ont contribué à l'élaboration de cette thèse.

Mes remerciements s'adressent en premier à mes deux directeurs de thèse Marie-Aline Charles et Jérémie Botton. Je tiens à les remercier de m'avoir accueillie dans leur équipe et de m'avoir permis de réaliser cette thèse. Je les remercie de m'avoir confiée ce sujet riche et pour leur encadrement et leur disponibilité tout au long de ces trois années. Trois années qui ont été source d'enrichissement professionnel.

Je remercie également Marie-Aline pour l'organisation de la vie d'équipe qui m'a permis de découvrir un champ de recherche passionnant et de bénéficier des connaissances et de la curiosité de l'ensemble des membres de l'équipe dans une ambiance chaleureuse et bienveillante.

Je remercie également Barbara Heude et Yves Akoli Koudou pour leurs conseils et leur écoute attentive.

Merci également à Ismaïl Ahmed et Jean-Louis Foulley de m'avoir conseillée dans la conduite des analyses bayésiennes.

J'exprime également ma reconnaissance à Emmanuelle Comets, Nathalie Costet, Vincent Gajdos, Adeline Leclercq Samson, Cécile Proust-Lima d'avoir accepté de participer à mon jury de thèse.

Merci à Jean Bouyer et Josiane Warszawski de m'avoir permis de participer à l'animation des travaux dirigés d'épidémiologie et des travaux pratiques. Ce fut une expérience très enrichissante.

Ces trois années ont également été source d'enrichissement personnel et m'ont permis de faire de belles rencontres amicales. Je remercie donc l'ensemble des membres de l'équipe pour leur accueil et leur sympathie. Je remercie particulièrement Eve, pour son côté « Bacrette », Fabienne, pour son amitié et les discussions passionnantes ; Jean, pour sa gentillesse et sérénité, Jonathan, pour son souci des autres et son attachement à faciliter l'intégration de chacun, Manik pour sa joie de vivre, sa générosité et les attentions qui égayaient le quotidien, Pauline pour sa gentillesse et sa solidarité, Wen Lun pour sa spontanéité attachante et les discussions philosophiques, Yves (il semblerait que

l'on puisse remercier mille fois une personne) pour ses jeux de mots légendaires et son soutien. Vous avez fait partie de ceux qui font que venir travailler est un plaisir.

Je remercie également Astrid et Catherine qui ont rendu la co-animation des travaux pratiques stimulantes. Plus personnellement merci Astrid pour ta générosité. Merci Catherine, ma famille Villejuifoise.

Merci aux Sudistes pour leur amitié de longue date qui m'est si précieuse, Camille, Camille, Céline, Elise, Emilie, Laure, Florence et leurs acolytes Stéphane, Morgan et Yoan.

Merci aux Ispédiennes, qui ont rendu l'apprentissage de l'épidémiologie encore plus stimulant, pour leur amitié, Astou, Betty et Marylou.

Merci à ma famille : mon oncle et ma tante pour leur accompagnement ; ma sœur et ma mère pour leur soutien sans faille.

Merci à « tonton Sicap » où qu'il soit!

VALORISATION SCIENTIFIQUE

ARTICLES SOUMIS

Sophie Carles, Marie-Aline Charles, Anne Forhan, Rémy Slama, Barbara Heude, Jérémie Botton and the EDEN mother child study group.

“A novel method to describe early offspring Body Mass Index (BMI) trajectories: application to maternal smoking in pregnancy”. *PlosOne*.

Sophie Carles, Marie-Aline Charles, Ismaïl Ahmed, Barbara Heude, Jérémie Botton and the EDEN mother child study group.

“Joint Bayesian weight and height post-natal growth model to study the effects of maternal smoking during pregnancy”. *Statistics in Medicine*.

Jérémie Botton, Claire Philippat, Antonia M. Calafat, Sophie Carles, Marie-Aline Charles, Rémy Slama, the EDEN mother-child cohort study group.

“Phthalate exposure during pregnancy and male offspring growth from the intra-uterine period to five years of age”. *Pediatric Obesity*.

Cécilia Saldanha-Gomes, Barbara Heude, Marie Charles, Blandine De Lauzon-Guillain, Jérémie Botton, Sophie Carles, Anne FORHAN, Patricia Dargent-Molina, and Sandrine Lioret.

“Prospective associations between energy balance-related behaviors at 2 years of age and subsequent adiposity: the EDEN mother-child cohort”. *International Journal of Obesity*.

Marion Taine, Bénédicte Stengel Anne Forhan, Sophie Carles, Jérémie Botton, Marie-Aline Charles, Barbara Heude; on behalf of the EDEN mother-child cohort study group.

“Rapid early growth modulates the association between birthweight and blood pressure at 5 years in the EDEN cohort study”. *Hypertension*.

COMMUNICATIONS ORALES EN CONGRÈS

Sophie Carles, Marie-Aline Charles, Anne Forhan, Rémy Slama, Barbara Heude, Jérémie Botton and the EDEN mother child study group.

“Maternal Smoking and offspring’s postnatal body mass index (BMI) trajectory between birth and five years from the EDEN mother child cohort study”. *Colloque de la Société Francophone des Origines Développementales de la santé, 6-7 Novembre, 2014, France*.

COMMUNICATIONS AFFICHÉES EN CONGRÈS

Sophie Carles, Marie-Aline Charles, Anne Forhan, Rémy Slama, Barbara Heude, Jérémie Botton and the EDEN mother child study group.

“Maternal Smoking and offspring’s postnatal body mass index (BMI) trajectory between birth and five years from the EDEN mother child cohort study”. *The 2nd International Conference on Nutrition and Growth, January 30-February 1, 2014, Barcelona, Spain.*

TABLE DES MATIERES

Introduction générale	15
Chapitre 1 : Etat de la question	18
1.1 Généralités sur la croissance postnatale.....	20
1.1.1 Les déterminants de la croissance postnatale.....	20
1.1.2 Les indicateurs de croissance	21
1.1.2.1 Croissance du poids et de la taille	21
1.1.2.2 Croissance de l'indice de masse corporelle (IMC)	23
1.2 Les principales méthodes de modélisation de la croissance	25
1.2.1 Principales approches de modélisation de la croissance	26
1.2.1.1 Méthodes de modélisation « individuelles».....	26
1.2.1.2 Méthodes de modélisation « groupales».....	27
1.2.2 Principaux modèles pour la modélisation de la croissance.....	28
1.2.2.1 Les modèles structuraux	29
1.2.2.1.1 Modèle de Jenss et Bayley	29
1.2.2.1.2 Modèle de Count.....	30
1.2.2.1.3 Modèle de Berkey et Reed.....	30
1.2.2.1.4 SITAR (SuperImposition by Translation And Rotation).....	31
1.2.2.2 Les modèles non structuraux.....	33
1.2.2.2.1 Les polynômes fractionnaires	33
1.2.2.2.2 Les splines.....	35
1.2.3 Principales méthodes d'inférences.....	37
1.2.3.1 Méthodes d'inférences fréquentistes.....	37
1.2.3.2 Méthodes d'inférences bayésiennes.....	38
1.3 Objectifs de la thèse.....	46
Chapitre 2 : Population et Données	48
2.1 Présentation générale	50
2.1.1 Inclusion et suivi	51
2.1.1.1 Inclusion	51
2.1.1.2 Suivi	51
2.1.1.2.1 Suivi prénatal.....	51
2.1.1.2.2 Suivi postnatal.....	52
2.2 Données utilisées.....	52
2.2.1 Tabagisme maternel pendant la grossesse	52
2.2.2 Données de croissance.....	52
2.2.2.1 A la naissance	52
2.2.2.2 Données extraites des carnets de santé.....	53
2.2.2.3 Données collectées lors des examens cliniques	55
2.2.2.4 Traitement des données.....	55
2.2.3 Tabagisme passif	56

2.2.4	Autres caractéristiques parentales.....	57
2.2.5	Alimentation de l'enfant.....	57
2.2.5.1	Allaitement maternel.....	57
2.2.5.2	Alimentation dans la petite enfance.....	57
2.3	Caractéristiques de la population d'étude.....	58
	Chapitre 3 : Tabagisme maternel pendant la grossesse et croissance de l'enfant	60
3.1	Contexte.....	61
3.2	Objectifs spécifiques.....	62
3.3	Matériel et Méthodes.....	63
3.3.1	Génération de variables.....	63
3.3.2	Modélisation de la croissance.....	66
3.3.2.1	Mesures anthropométriques (poids et taille) utilisées pour la modélisation des croissances staturale et pondérale.....	66
3.3.2.2	Modèles utilisés pour la modélisation des croissances staturale et pondérale.....	67
3.3.2.3	Logiciel et méthode d'estimation.....	70
3.3.2.4	Etude des associations entre le tabagisme maternel et la croissance.....	70
3.3.2.4.1	Sélection de la population d'étude.....	70
3.3.2.4.2	Stratégie d'analyse.....	71
3.4	Résultats.....	85
3.4.1	Caractéristiques de la population d'étude.....	85
3.4.2	Caractéristiques anthropométriques des enfants.....	90
3.4.2.1	Croissance pondérale.....	90
3.4.2.1.1	Paramètres du modèle de croissance pondérale.....	90
3.4.2.1.2	Prédictions individuelles de poids.....	90
3.4.2.2	Croissance staturale.....	91
3.4.2.2.1	Paramètres du modèle de croissance staturale.....	91
3.4.2.2.2	Prédictions de la taille.....	92
3.4.2.3	Croissance de l'IMC.....	93
3.4.2.3.1	Prédictions individuelles d'IMC.....	93
3.4.2.3.2	Pic d'IMC.....	93
3.4.3	Association entre le tabagisme maternel pendant la grossesse et la croissance des enfants de la naissance à 5 ans.....	96
3.4.3.1	Croissance staturale.....	96
3.4.3.2	Croissance pondérale.....	97
3.4.3.3	Croissance de l'IMC.....	97
3.4.3.4	Relation dose-effet.....	102
3.4.3.5	Analyses de sensibilité.....	106
3.5	Discussion.....	127
3.6	Conclusion.....	131
	Chapitre 4 : Tabagisme maternel pendant la grossesse et croissance de l'enfant : une approche bayésienne	132
4.1	Introduction.....	134
4.2	Objectifs spécifiques.....	134

4.3	Méthodes.....	135
4.3.1	Modèle de croissance.....	135
4.3.2	Implémentation MCMC.....	140
4.3.2.1	Algorithme et logiciel.....	140
4.3.2.2	Evaluation de la convergence.....	140
4.3.3	Application.....	141
4.3.3.1	Application sur données de croissance simulées.....	141
4.3.3.3	Application à l'analyse des associations entre le tabagisme maternel pendant la grossesse et la croissance postnatale.....	143
4.3.3.3.1	Echantillon de données.....	143
4.3.3.3.2	Aspects calculatoires.....	143
4.3.3.3.3	Sélection et évaluation du modèle.....	143
4.4	Résultats.....	145
4.4.1	Etude de simulation.....	145
4.4.2	Associations entre le tabagisme de la mère pendant la grossesse et la croissance de l'enfant 150	
4.4.2.1	Evaluation de la convergence du modèle.....	150
4.4.2.2	Sélection et évaluation du modèle.....	153
4.4.2.3	Analyse des associations.....	156
4.4.2.4	Sensibilité aux lois a priori.....	159
4.4.2.5	Différences de poids, taille et IMC selon le statut tabagique.....	160
4.5	Discussion.....	164
4.6	Perspectives.....	167
	Chapitre 5 : Analyse de la modélisation conjointe entre croissances pondérale et staturale	169
5.1	Introduction et objectifs spécifiques.....	170
5.2	Population et méthodes.....	171
5.2.1	Population d'étude.....	171
5.2.2	Méthode.....	171
5.3	Résultats.....	173
5.4	Discussion.....	178
	Travail exploratoire : étude des associations entre la croissance pondérale et staturale et modélisation de la matrice de variances-covariances	180
6.1	Contexte.....	182
6.2	Objectifs spécifiques.....	183
6.3	Méthodes.....	183
6.3.1	Population d'étude.....	183
6.3.2	Méthode de modélisation de la matrice de variances-covariances.....	184
6.3.3	Application de cette méthode à l'étude de la croissance.....	186
6.3.3.1	Ajustement des modèles.....	188
6.3.3.2	Résultats.....	189
6.4	Discussion.....	200
	Discussion générale	204

Annexe 1 : Comparaison des modèles décrivant la croissance du poids et de la taille selon les modèles d'erreur utilisés	228
Annexe 2 : Code rjags, modèle bayésien décrivant conjointement la croissance du poids et de la taille ..	229
Annexe 3 : Mélange des trois chaînes pour les intercepts des paramètres de croissance.....	233
Annexe 4 : Différences moyennes d'IMC entre les enfants exposés au tabagisme maternel durant le 1 ^{er} trimestre exclusivement ou tout au long de la grossesse et les enfants non exposés de la naissance à 5 ans	234
Annexe 5 : Articles	235

LISTE DES TABLEAUX

Tableau 1 : Distributions des paramètres individuels de croissance obtenus sur un sous-échantillon de la population EDEN.....	70
Tableau 2 : Nombre de mesures de poids et de taille des enfants disponibles pour la modélisation de la croissance pondérale et staturale (n=1666), EDEN, France, 2003-2012	87
Tableau 3 : Caractéristiques des enfants (n=1666) et de leurs parents selon le statut tabagique maternel pendant la grossesse, EDEN, France, 2003-2012	89
Tableau 4 : Distribution des paramètres de croissance du poids des enfants (0-5 ans)	90
Tableau 5 : Prédiction des poids et des vitesses de croissance pondérale instantanées, EDEN, 0-5 ans... 91	91
Tableau 6 : Distribution des paramètres de croissance staturale, EDEN, 0-5 ans	91
Tableau 7 : Prédiction de taille et de vitesses instantanées de croissance staturale, EDEN, 0-5 ans.....	92
Tableau 8 : Prédiction d'IMC et de vitesses instantanées de croissance de l'IMC, EDEN, 0-5 ans	93
Tableau 9 : Age et IMC au Pic d'IMC.....	95
Tableau 10 : Différences moyennes des paramètres des modèles de croissances staturale et pondérale de la naissance à 5 ans pour chaque catégorie de tabagisme maternel et la catégorie de référence (absence de tabagisme pendant la grossesse), méthode en 1 étape, étude EDEN, France, 2003-2012.....	100
Tableau 11 : Différences moyennes des caractéristiques anthropométriques prédites de la naissance à 5 ans pour chaque catégorie de tabagisme maternel et la catégorie de référence (absence de tabagisme pendant la grossesse), méthode en 2 étapes, étude EDEN, France, 2003-2012	101
Tableau 12 : Différences moyennes des paramètres des modèles de croissances staturale et pondérale de la naissance à 5 ans pour une augmentation de 5 cigarettes de la consommation journalière de tabac aux 2 ^e et 3 ^e trimestres, Méthode en 1 étape, étude EDEN, France, 2003-2012.....	103
Tableau 13 : Différences moyennes des caractéristiques anthropométriques prédites de la naissance à 5 ans pour une augmentation de 5 cigarettes de la consommation journalière de tabac aux 2 ^e et 3 ^e trimestres, méthode en 2 étapes, étude EDEN, France, 2003-2012.....	105
Tableau 14 : Différences moyennes des caractéristiques anthropométriques pour chaque catégorie de tabagisme maternel et la catégorie de référence (absence de tabagisme pendant la grossesse), méthode en 2 étapes, enfants nés à terme, étude EDEN, France, 2003-2012	107
Tableau 15 : Différences moyennes des caractéristiques anthropométriques de 3 mois à 5 ans entre les enfants exposés au tabagisme durant la période postnatale exclusivement versus les enfants non exposés, Méthode en 2 étapes, étude EDEN, France, 2003-2012.....	109
Tableau 16 : Différences moyennes des caractéristiques anthropométriques pour chaque catégorie de tabagisme maternel et la catégorie de référence (absence de tabagisme pendant la grossesse), méthode en 2 étapes, chez les enfants dont la taille des parents est connue, étude EDEN, France, 2003-2012	111
Tableau 17 : Modélisation de l'IMC, Polynômes Fractionnaires	112
Tableau 18 : Modélisation de l'IMC, Polynômes fractionnaires de degré 2	114
Tableau 19 : Choix de l'emplacement des nœuds, Splines, Modélisation de l'IMC chez les filles.....	115

Tableau 20 : Différences moyennes des IMC prédits de 3 mois à 5 ans pour chaque catégorie de tabagisme maternel et la catégorie de référence (absence de tabagisme maternel), Méthode en 2 étapes (polynômes fractionnaires, splines, modélisation indirecte) ^a , étude EDEN, France, 2003-2012	124
Tableau 21 : Différences moyennes de taille, poids et IMC obtenues à partir des examens cliniques (1, 3 et 5 ans) et des données prédites par la modélisation dans chaque catégorie de tabagisme maternel et la catégorie de référence (absence de tabagisme), méthode en « 2 étapes »	126
Tableau 22 : Distributions a priori des paramètres du modèle bayésien (notations rjags).....	140
Tableau 23 : Distributions a priori des paramètres impliqués dans le modèle bayésien pour les analyses de sensibilité aux lois a priori	144
Tableau 24 : Valeurs réelles (fixées), biais et probabilité de couverture obtenus à partir du modèle bayésien conjoint de la taille et du poids sur 100 jeux de données simulés	146
Tableau 25 : Coefficients de régressions associés au tabagisme et aux paramètres de croissance ^a	157
Tableau 26 : Sensibilité des résultats aux choix des lois a priori	160
Tableau 27 : Différences moyennes de poids, taille et d'IMC entre les enfants exposés au tabagisme maternel durant le 1 ^{er} trimestre exclusivement ou tout au long de la grossesse et les enfants non exposés de la naissance à 5 ans.....	162
Tableau 28 : Variances inter-sujets et moyennes des variances intra-sujet pour les prédictions des caractéristiques de croissance issues de la modélisation conjointe et séparée.	175
Tableau 29 : Distribution des résidus de taille dans le « sous-échantillon de 20% » obtenue avec la modélisation conjointe poids/taille ou avec la modélisation univariée de la taille.....	178
Tableau 30 : Différences moyennes des paramètres de croissances staturale et pondérale de la naissance à 5 ans entre les enfants en surpoids ou obèses à 5 ans et les enfants de poids normal, étude EDEN, France, 2003-2012	189
Tableau 31 : Moyenne a posteriori et intervalle de crédibilité des paramètres de dépendance entre les effets aléatoires des paramètres de croissances pondérale et staturale, matrice de variances-covariances homogène.....	190
Tableau 32 : Moyenne a posteriori et intervalle de crédibilité des paramètres de dépendance entre les effets aléatoires des paramètres de croissances pondérale et staturale, matrice de variances-covariances dépendant du statut pondéral de l'enfant.	199

LISTE DES FIGURES

Figure 1 : Courbes de croissance pondérales et staturales de la naissance à 3 ans issues du carnet de santé pour les garçons et les filles	22
Figure 2 : Percentiles des courbes de croissance de l'IMC de la naissance à 18 ans chez les filles, figure issue du carnet de santé.....	24
Figure 3 : Illustration de l'effet de chaque paramètre du modèle de croissance sur la trajectoire de croissance de la taille	32
Figure 4 : Extrait de la partie « croissance de l'enfant » des auto-questionnaires	54
Figure 5 : Attrition sur 5 ans.....	58
Figure 6 : Effet de l'inclusion du poids de naissance (ne considérant pas la perte de poids néonatale) sur une trajectoire de poids obtenu avec le modèle de Jenss-Bayley.....	66
Figure 7 : Illustration de la signification des paramètres selon la paramétrisation du modèle de Jenss-Bayley.....	67
Figure 8 : Trajectoires de croissance de poids de la naissance à 5 ans par déciles de paramètres de poids obtenus sur un sous-échantillon de la population EDEN.....	69
Figure 9 : Effet de la variation des paramètres du modèle de croissance du poids sur les trajectoires d'IMC	76
Figure 10 : Effet de la variation des paramètres du modèle de croissance de la taille sur les trajectoires d'IMC.....	77
Figure 11 : Trajectoires de vitesse de croissance du poids de la naissance à 5 ans par déciles de paramètres de poids obtenus sur un sous-échantillon de la population EDEN.....	78
Figure 12 : Trajectoires de croissance d'IMC pour 12 sujets avec un âge au pic d'IMC excédant 2 ans	80
Figure 13 : Diagramme de flux des mesures de poids et de taille	86
Figure 14 Trajectoires de croissance moyennes de taille, poids et IMC (A1 ; B1 ; C1) selon le statut tabagique de la mère et différences de taille, poids et IMC (A2 ; B2 ; C2) moyens entre les enfants exposés durant le 1 ^{er} trimestre exclusivement ou tout au long de la grossesse et les enfants non exposés.....	99
Figure 15 : Trajectoires d'IMC obtenues pour 4 garçons avec les polynômes de degré 2 et 6 et la méthode de modélisation indirecte	113
Figure 16 : Trajectoires d'IMC obtenues pour 4 garçons avec le polynôme de degré 2 retenu, méthode de Royston et Altman	114
Figure 17 : Modélisation de l'IMC des filles et des garçons, splines	116
Figure 18 : Trajectoires d'IMC obtenues pour 4 garçons avec le modèle splines retenu.....	117
Figure 19 : Distributions des résidus d'IMC, polynômes fractionnaires, splines, méthode indirecte	118
Figure 20 : Distribution des résidus de poids et de taille, modélisation indirecte.....	120
Figure 21 : Mesures de taille et trajectoires prédites de taille pour 12 individus sélectionnés aléatoirement	121
Figure 22 : Mesures de poids et trajectoires prédites de poids pour 12 individus sélectionnés aléatoirement	122
Figure 23 : Mesures d' IMC et trajectoires prédites d'IMC pour 12 individus sélectionnés aléatoirement	123
Figure 24 : Mesures de poids (données simulées), trajectoires moyennes ajustées par le modèle conjoint bayésien et intervalles de crédibilité à 95% obtenus pour 3 individus chacun issu d'un jeu de données simulé différent et sélectionnés aléatoirement	148
Figure 25 : Mesures de taille (données simulées), trajectoires moyennes ajustées par le modèle conjoint bayésien et intervalles de crédibilité à 95% obtenus pour 3 individus chacun issu d'un jeu de données simulé différent et sélectionnés aléatoirement	149
Figure 26 : Critère de Gelman pour le coefficient de régression associé au tabagisme maternel tout au long de la grossesse sur le paramètre B du modèle de croissance pondérale.....	150
Figure 27 : Autocorrélation de l'une des 3 chaînes pour le coefficient de régression associé au tabagisme maternel tout au long de la grossesse sur le paramètre B du modèle de croissance pondérale.....	151
Figure 28 : Evolution des chaînes pour le coefficient de régression associé au tabagisme maternel tout au long de la grossesse sur le paramètre B du modèle de croissance pondérale.....	152
Figure 29 : Evolution des quantiles, en fonction du nombre d'itérations, de la distribution du coefficient de régression associé au tabagisme maternel tout au long de la grossesse et le paramètre B du modèle de croissance pondérale	152
Figure 30 : Mesures de discrédance pour le poids et la taille pour les données répliquées en fonction des données observées	153

Figure 31 : Valeurs observées de poids, courbes moyennes ajustées par le modèle conjoint bayésien et intervalles de crédibilité à 95% de 6 enfants sélectionnés aléatoirement	154
Figure 32 : Valeurs observées de taille, courbes moyennes ajustées par le modèle conjoint bayésien et intervalles de crédibilité à 95% de 6 enfants sélectionnés aléatoirement	155
Figure 33 : Distribution a posteriori des coefficients de régression associés de manière modérée au tabagisme maternel	159
Figure 34 : Différences d'IMC, poids, taille moyens entre les enfants exposés tout au long de la grossesse ou durant le 1 ^{er} trimestre et la catégorie de référence	163
Figure 35 : Trajectoire de croissance staturale, modélisation conjointe et modélisation univariée	176
Figure 36 : Comparaison des résidus obtenus avec la modélisation conjointe poids/taille et la modélisation univariée de la taille dans l'échantillon où 20% des tailles ont été supprimés	177
Figure 37 : Association entre le paramètre A de la taille et du poids (association marginale)	192
Figure 38 : Association entre le paramètre C de la taille et C du poids (ajustée sur les paramètres de naissance).....	193
Figure 39 : Association entre le paramètre C de la taille et D de la taille (ajustée sur les paramètres de naissance, et le paramètre C du poids)	194
Figure 40 : Association entre le paramètre C du poids et D du poids (ajustée sur les paramètres de naissance, et les paramètres C et D de la taille)	195
Figure 41 : Association entre le paramètre C de la taille et B du poids (ajustée sur l'ensemble des autres paramètres de croissance)	196
Figure 42 : Association entre le paramètre B de la taille et B du poids (ajusté sur l'ensemble des autres paramètres de croissance)	197

Introduction générale

La croissance physique de l'enfant peut être définie comme l'augmentation de la taille des différents éléments de l'organisme entre la naissance et la fin de l'adolescence. Elle peut être caractérisée quantitativement comme un gain statural, de périmètre crânien, de masse corporelle ou d'autres parties du corps (Koletzko et al., 2014).

La surveillance de la croissance en routine est acceptée par le plus grand nombre, est soutenue par les professionnels de santé et est une composante faisant partie intégrante des services pédiatriques dans le monde (Panpanich and Garner, 2000). Cette surveillance a été définie en 5 points : une mesure régulière de la taille et du poids des enfants ; le report de l'information sur les courbes de croissance ; l'initiation d'investigations en cas de croissance anormale ; en résultant, la possibilité de diagnostiquer un(e) problème (maladie) sérieux(se) précocement ; et ainsi améliorer le pronostic.

Au niveau individuel, la surveillance de la croissance fournit un indicateur important de la santé et du développement de l'enfant. En identifiant les enfants s'écartant du profil de croissance normal, elle permet de détecter, particulièrement chez les plus jeunes, des maladies héréditaires, des difficultés à être allaité, des maladies chroniques de toutes causes incluant les infections respiratoires, le paludisme, la tuberculose et des déficits en hormone de croissance, la détresse sociale dont l'une des principales résultantes est la malnutrition (Garner et al., 2000; Koletzko et al., 2014) et d'avoir ainsi l'opportunité de mettre en place des mesures correctives.

Au niveau populationnel, dans les années 1990, Barker formulait la théorie des origines fœtales des maladies (Barker, 2007). Cette théorie fut évoquée pour la première fois pour expliquer les disparités géographiques des taux de maladies cardiovasculaires en Angleterre. En effet, les modes de vie à l'âge adulte et les facteurs de risque biologiques, ne semblaient que partiellement expliquer celles-ci. Cette théorie, loin de nier l'influence potentielle des modes de vie et des prédispositions génétiques sur l'incidence des maladies, soulignait l'importance des événements marquant la vie fœtale sur l'occurrence des maladies. En effet, après avoir observé un parallélisme entre les différences de taux de décès dans différentes zones d'Angleterre et la mortalité infantile plusieurs décennies auparavant, les travaux de David Barker ont mis en évidence une association positive, au niveau individuel, entre une croissance fœtale altérée et

l'incidence des maladies cardiovasculaires. Cette hypothèse serait sous-tendue par un changement de la composition corporelle, de la physiologie et du métabolisme lié à une malnutrition *in utero*. La période prénatale constituerait donc une période critique pour le développement ultérieur de maladies chroniques.

Cependant, croissance foetale et croissance postnatale dans la petite enfance ne peuvent pas être considérées indépendamment, la seconde étant en grande partie la continuité de la première. Certains ont postulé qu'une croissance postnatale précoce rapide, faisant souvent suite à une période de restriction de la croissance foetale, constituerait le réel facteur de risque des maladies ultérieures (Lucas et al., 1999). En effet, les enfants ayant souffert d'une restriction alimentaire ou ayant été soumis à un déficit en oxygène pendant la période intra-utérine conduisant à une restriction de croissance intra-utérine connaissent souvent, après la naissance, un phénomène de rattrapage de croissance (« catch-up »). Prader et coll. l'ont défini comme le rétablissement nutritionnel suivant une période de restriction de croissance après que celle-ci soit supprimée (Prader et al., 1963). Celui-ci est marqué par une période de croissance accélérée du poids, de la taille et du périmètre crânien. Si 90% des enfants nés petits pour l'âge gestationnel présenteraient ce phénomène de rattrapage lors de la première année, il peut également concerner des enfants nés avec un poids normal (Chrestani et al., 2013). Ce phénomène est également observé plus tard dans l'enfance suite à des maladies ou conditions susceptibles de contraindre la croissance.

Une croissance pondérale postnatale rapide dans les premiers mois de vie puis au-delà de l'âge de 2 ans serait associée à un risque accru de surpoids et d'obésité à l'âge adulte émergeant déjà dans l'enfance (Botton et al., 2008; Dennison et al., 2006; Holzhauer et al., 2009; Karaolis-Danckert et al., 2006; Monteiro and Victora, 2005; Reilly et al., 2005). Une croissance rapide du poids après l'âge de 2 ans serait également associée à un risque accru de maladies cardiovasculaires, d'hypertension et de diabète de type II (Eriksson, 2011). Cependant, si de nombreuses études suggèrent une association positive entre une croissance pondérale rapide dans l'enfance et le risque de maladies chroniques à l'âge adulte, d'autres soulignent les effets bénéfiques à court terme d'une croissance staturo-pondérale rapide entre la naissance et 2 ans sur la santé de l'enfant chez les enfants nés petits pour l'âge gestationnel. En effet, celle-ci serait associée à un risque d'hospitalisations et une mortalité plus faible (Victora et al., 2001). Il a également été montré que chez ces enfants, le prédicteur le plus important de

performances intellectuelles et psychologiques en dessous de la normale était l'absence de catch-up (Lundgren et al., 2001). Il convient donc d'être prudent sur l'interprétation des associations en tenant compte de la balance entre bénéfiques et risques et de les situer dans le contexte d'évolution de l'enfant.

Ces résultats illustrent l'importance de la croissance prénatale et postnatale sur la santé à court terme de l'enfant et sur la santé plus tard dans la vie. La croissance peut être vue à la fois comme une résultante (outcome) de l'environnement de l'enfant au sens large et de sa génétique mais également comme une variable de médiation, étant déterminée par l'environnement et la génétique et influençant elle-même des évènements de santé ultérieurs. Ce concept de programmation des maladies témoigne de l'importance d'identifier les déterminants de la croissance afin d'identifier et de mettre en place des mesures de prévention ciblées et adaptées aux populations.

Chapitre 1 : Etat de la question

1.1 Généralités sur la croissance postnatale

1.1.1 Les déterminants de la croissance postnatale

Outre les facteurs génétiques, des facteurs prénataux, reflets de l'environnement de la mère pendant la grossesse, et des facteurs postnataux, reflets de l'environnement dans l'enfance, peuvent influencer la croissance postnatale. Leur étude permet, pour les facteurs non modifiables, de mieux comprendre les mécanismes sous-jacents, pour les facteurs modifiables d'envisager des actions de Santé Publique pour réduire le risque de suivre des trajectoires de croissance qui auraient été identifiées comme potentiellement à risque.

Dans ce paragraphe, nous ne présentons pas les déterminants de la croissance staturo-pondérale postnatale de manière exhaustive mais donnons quelques exemples de facteurs modifiables (nous ne parlerons pas des facteurs génétiques), s'inscrivant dans les grandes catégories de déterminants décrites ci-dessus, afin de mieux comprendre leur impact potentiel sur la croissance. Les facteurs considérés dans les analyses ultérieures seront principalement ciblés.

Parmi les facteurs prénataux susceptibles d'affecter la croissance postnatale, certains peuvent induire une croissance fœtale altérée. En répercussion d'un retard de croissance intra-utérin, les enfants, ont tendance à avoir une croissance postnatale rapide pour rattraper leur déficit en poids et en taille à la naissance (Prader et al., 1963). Les répercussions sur les cinétiques de croissance postnatale peuvent différer selon les déterminants à l'origine de cette restriction de croissance. Le tabagisme de la mère pendant la grossesse s'inscrit parmi les facteurs responsables d'une croissance fœtale altérée (Andres and Day, 2000; Castles et al., 1999; Cnattingius, 2004) et a été associé à un déficit de poids et de taille à la naissance suivi d'une croissance staturo-pondérale rapide (Chen et al., 2006; Durmus et al., 2011; Howe et al., 2012; Timmermans et al., 2013). D'autres facteurs prénataux peuvent à l'inverse induire une croissance fœtale accélérée et être responsables d'une macrosomie de l'enfant à la naissance (Baptiste-Roberts et al., 2012). Le diabète gestationnel est un de ces facteurs et a été associé à une croissance pondérale ralentie dans les premières années de vie (Baptiste-Roberts et al., 2012; Touger et al., 2005). L'IMC maternel pré-conceptionnel a également été positivement associé à un risque accru de donner naissance à des enfants gros pour l'âge gestationnel (Heude et al., 2012). Regnault et coll. ont mis en évidence une vitesse de

croissance pondérale plus faible dans les premiers mois de vie chez les enfants de mères obèses comparée à celle des enfants de mères de poids normal (Regnault et al., 2010).

Parmi les facteurs postnataux, l'alimentation de l'enfant a un rôle important sur la croissance. De nombreuses études épidémiologiques montrent que la croissance staturo-pondérale des enfants recevant un allaitement prolongé et exclusif est plus lente durant la 1^{re} année de vie (particulièrement après le 3^e mois) que celle des enfants nourris avec des préparations infantiles ou recevant un allaitement moins exclusif ou de plus courte durée (Dewey, 1998; Dewey et al., 1995; Hediger et al., 2000; Kramer et al., 2004; Nielsen et al., 1998; Ong et al., 2009, 2002; Regnault et al., 2010; World Health Organization Working Group on Infant Growth, 1995).

La compréhension des mécanismes à travers lesquels ces facteurs agissent sur la croissance est d'autant plus difficile que ces facteurs n'agissent pas indépendamment les uns des autres et peuvent être au centre d'interactions complexes.

1.1.2 Les indicateurs de croissance

Les profils de croissance des trois variables anthropométriques que sont le poids, la taille et l'indice de masse corporelle, auxquelles nous nous sommes intéressés dans ce travail sont décrits dans cette partie.

1.1.2.1 Croissance du poids et de la taille

D'une manière générale, 3 principales périodes de croissance postnatale peuvent être distinguées (Hauspie et al., 2004) :

- La « petite enfance », débutant avant la naissance et se terminant entre 1 an et demi et 2 ans. Elle constitue une phase d'adaptation entre la croissance intra-utérine vers la croissance extra-utérine avec une vitesse très rapide initialement qui décroît progressivement.
- L'enfance : la croissance pendant cette période est linéaire
- La puberté : qui débute par une accélération de la vitesse de croissance, le pic pubertaire, et qui se termine par la fin de la croissance.

La **Figure 1** illustre les courbes de croissances pondérale et staturale de la naissance à 3 ans figurant dans les carnets de santé français.

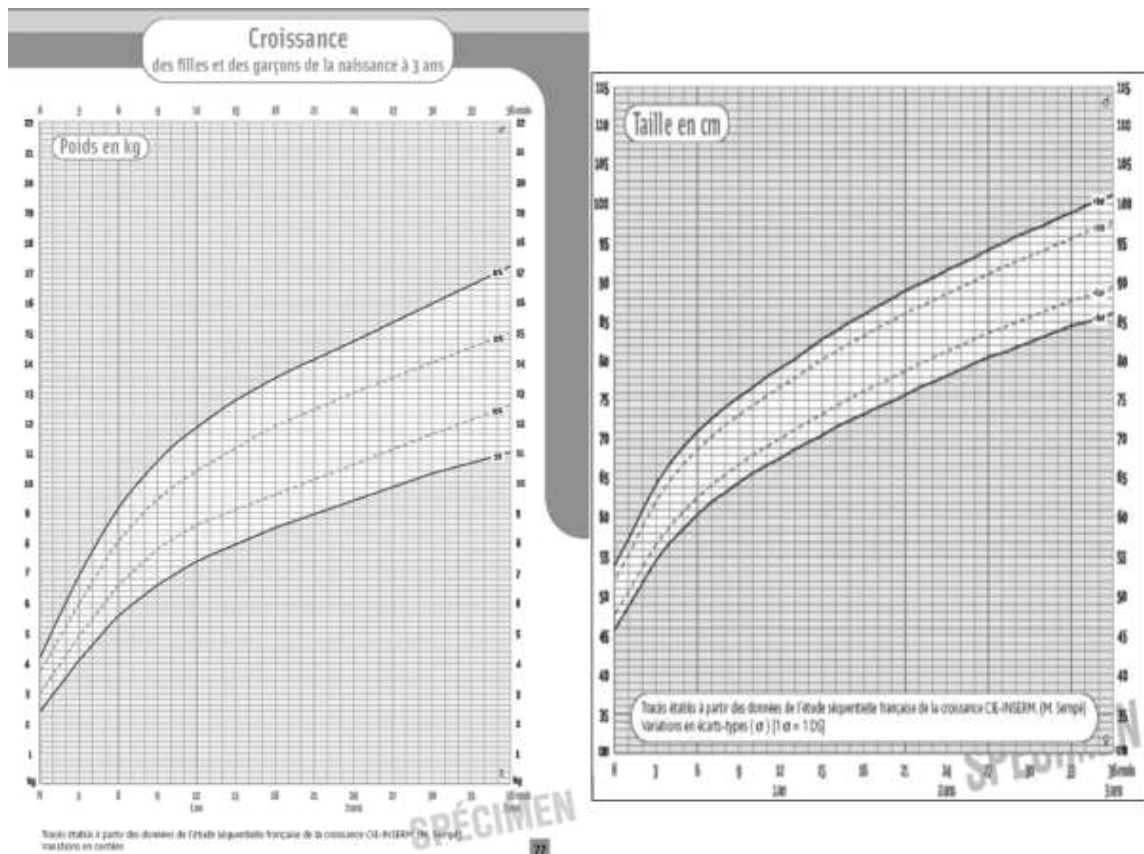


Figure 1 : Courbes de croissance pondérales et staturales de la naissance à 3 ans issues du carnet de santé pour les garçons et les filles

Les profils de croissance du poids et de la taille sont très semblables et par conséquent sont souvent modélisés de manière similaire. Une particularité de la croissance du poids, est l'existence, en général, d'une perte de poids néonatale (non représentée sur les courbes du carnet de santé) qui a lieu dans les premiers jours de vie. Regnault et coll. montrent que le pic de perte de poids néonatale est observé en moyenne entre le deuxième et le troisième jour de vie chez les nouveau-nés à terme (Regnault et al., 2011). Cette courte période est suivie d'une croissance pondérale rapide et globalement monotone. Quant à la taille, elle croît au cours du temps de façon monotone. L'évolution de la croissance, bien que monotone, est non linéaire. En effet, à la naissance, une vitesse de croissance élevée est observée, puis la vitesse de croissance décroît rapidement jusqu'à 2-3 ans environ (expliquant la courbure des trajectoires de croissance dans les premières années de vie). Il s'ensuit une période de croissance plus lente et diminuant moins rapidement (la vitesse de croissance devient presque constante durant cette période, expliquant la linéarité des trajectoires de croissance plus

tard dans l'enfance) jusqu'à la puberté. La puberté est marquée par une vitesse de croissance rapide jusqu'au pic pubertaire puis par une diminution de la croissance jusqu'à la fin de la croissance (Karlberg, 1989).

1.1.2.2 Croissance de l'indice de masse corporelle (IMC)

L'IMC est utilisé comme indicateur de corpulence. Dans les populations adultes, cet indicateur est indépendant de la taille et décrit donc le poids d'un individu corrigé de sa taille. Il est défini comme le rapport du poids en kg sur le carré de la taille exprimé en m². Il est souvent utilisé en épidémiologie comme un proxy de l'adiposité d'un individu, construit à partir de données connues des individus ou mesurables à faible coût. C'est à partir de cet indice que sont définis le surpoids et l'obésité, par le dépassement de valeurs seuil.

Le principal inconvénient de cet indicateur est qu'il ne permet pas de distinguer la masse grasse de la masse maigre, et son utilisation comme proxy de la masse grasse est discutée particulièrement chez l'enfant (Mei et al., 2002).

Rolland-Cachera et coll. discutent le fait que les courbes de croissance de l'IMC illustrent le développement normal de l'adiposité d'un enfant au cours de sa croissance. Différentes phases peuvent être distinguées. Lors de la première année de vie, l'adiposité augmente rapidement. Cette croissance est due à une augmentation en taille et en nombre des adipocytes qui diminuent ensuite et restent stables pendant plusieurs années. Alors que la croissance du tissu adipeux se ralentit, la taille des enfants continue elle à croître ; il en résulte un amincissement physiologique de l'enfant. Cette décroissance est suivie par la reprise d'une croissance rapide de la masse grasse dont le début a été dénommé « rebond d'adiposité » (Rolland-Cachera et al., 1987, 1984).

La **Figure 2** illustre le profil de croissance « habituel » de l'IMC de la naissance à 18 ans chez les filles (le profil d'évolution de l'IMC est similaire chez les garçons).

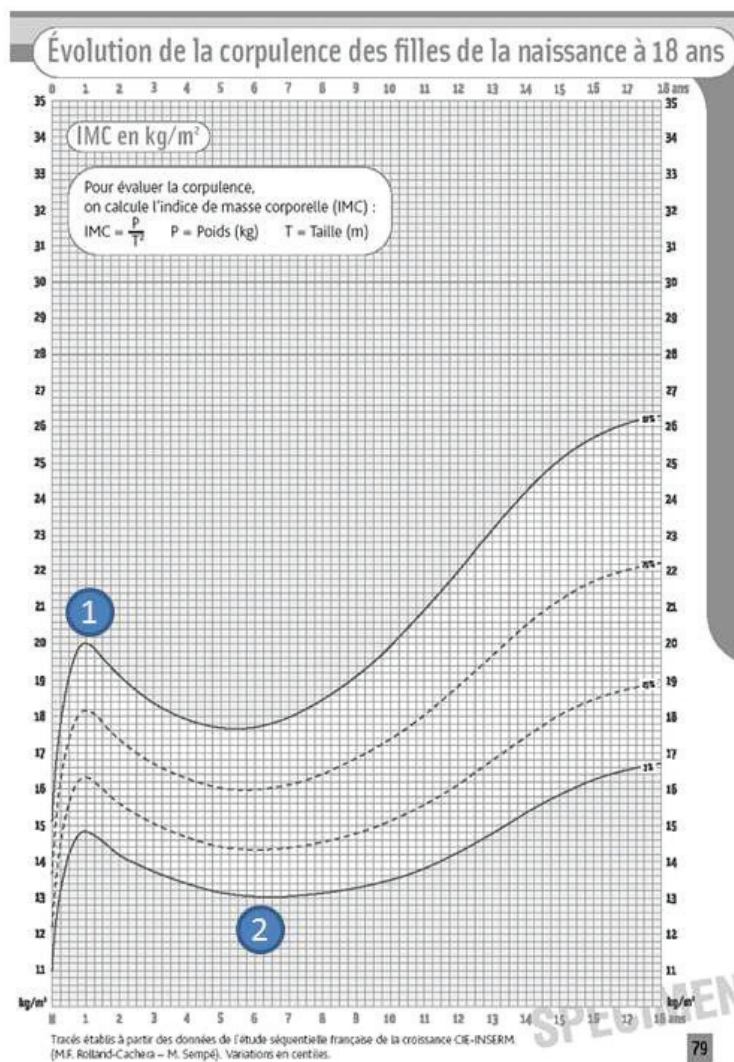


Figure 2 : Percentiles des courbes de croissance de l'IMC de la naissance à 18 ans chez les filles, figure issue du carnet de santé

- 1 L'IMC augmente dans un premier temps de la naissance à environ 9 mois atteignant son maximum appelé le pic d'IMC (Silverwood et al., 2009).
- 2 Après le pic d'IMC, l'IMC décroît jusqu'à atteindre un nadir vers l'âge de 6 ans avant de croître à nouveau, ce « phénomène » est appelé rebond d'IMC ou encore rebond d'adiposité (Silverwood et al., 2009).

Les âges et valeurs d'IMC au pic et au rebond d'adiposité sont des caractéristiques importantes de la croissance du tissu adipeux d'un enfant. En effet, il a été mis en évidence une association entre un rebond d'adiposité précoce et/ou un IMC plus

important au rebond d'adiposité et une augmentation de l'IMC/un risque accru d'obésité dans l'enfance et à l'âge adulte (Chivers et al., 2010, 2009; Rolland-Cachera et al., 1984; Whitaker et al., 1998).

Les associations entre l'âge au pic d'IMC dans l'enfance et l'IMC au pic d'IMC et la croissance ont été moins étudiées. Wen et coll. montrent une forte corrélation entre l'IMC au pic d'IMC et l'IMC au rebond d'adiposité (Wen et al., 2012). Silverwood et coll. ont montré quant à eux une association positive entre l'âge et l'IMC au pic d'IMC et l'IMC entre 5 et 13 ans (Silverwood et al., 2009).

1.2 Les principales méthodes de modélisation de la croissance

L'étude de la croissance des enfants nécessite une collecte de mesures anthropométriques régulière au cours du temps. L'analyse des données longitudinales permet de caractériser la dynamique de croissance des individus afin de pouvoir étudier les déterminants (ou événements de santé ultérieurs) associés à des différences de profils de croissance et de formuler des hypothèses sur les potentiels processus sous-jacents expliquant ces différences.

En pratique, les données de croissance collectées dans le cadre d'études épidémiologiques, peuvent être recueillies dans le cadre d'examens cliniques, organisés pour les besoins des études. Si ces examens sont planifiés à des âges donnés, communs à l'ensemble des enfants, il existe une certaine variabilité inhérente aux contraintes de l'étude, sur l'âge effectif des enfants lors de ces examens. Une autre source de données anthropométriques provient du suivi des enfants dans le cadre de la surveillance de routine de la croissance effectuée par les médecins traitants. De ce fait, les mesures sont variables en fonction des enfants que ce soit en terme de nombre de mesures pour une même tranche d'âge ou en terme d'âge de mesure. Il faut noter que ces mesures sont moins standardisées que celles réalisées dans le cadre des études.

Une limite des études de suivi longitudinal nécessaire pour la collecte des mesures permettant l'analyse de la croissance, est liée aux perdus de vue, ou non-réponses. Dans le cadre d'une analyse des déterminants de la croissance, l'exclusion des enfants ayant des informations manquantes, est susceptible d'induire un biais de sélection si ces enfants présentent des caractéristiques particulières (associées au déterminant et/ou à la croissance).

Hauspie et coll. synthétisent les avantages de la modélisation de la croissance en 6 points, elle permet :

1. d'estimer les trajectoires de croissance (considérant la croissance comme un processus continu) à partir d'un jeu de mesures discontinu de croissance.

Elle permet ainsi d'homogénéiser l'information disponible pour chaque enfant et permettra de disposer d'une mesure prédite du paramètre de croissance d'intérêt à l'âge fixé pour les besoins des analyses ultérieures.

2. d'estimer la croissance entre plusieurs temps de mesure (et peut être vu comme une technique d'interpolation)

3. de résumer la croissance par un nombre restreint de paramètres (de ce fait, elle est une technique de réduction des données)

4. l'estimation de paramètres clés de la croissance

5. d'obtenir la courbe de vitesse de croissance instantanée. Cet autre indicateur donne une autre information sur la dynamique de croissance que la « mesure brute », car elle s'affranchit de l'aspect cumulatif et par conséquent est davantage période-spécifique.

6. d'estimer la courbe moyenne « typique » de croissance dans la population étudiée

Différentes approches et différents modèles ont été utilisés pour la modélisation de la croissance de l'enfant.

1.2.1 Principales approches de modélisation de la croissance

Deux approches, ayant des objectifs distincts, sont utilisées pour la modélisation de la croissance et l'étude de ses déterminants : une approche que l'on peut qualifier de méthode « groupale » et une approche « individuelle ».

1.2.1.1 Méthodes de modélisation « individuelles »

La nature longitudinale des données de croissance, celles-ci étant collectées plusieurs fois à des temps différents sur les mêmes sujets, nécessite d'analyser ces données comme des données corrélées. En effet, les données d'un même sujet ne peuvent pas être considérées comme étant indépendantes, la corrélation entre les mesures d'un même individu étant importante. Les modèles les plus utilisés et qui permettent de faire une inférence au niveau individuel et donc de caractériser les

trajectoires de croissance spécifiques à chaque individu sont les modèles mixtes qu'ils soient linéaires ou non-linéaires en paramètres. Ces modèles incorporent des effets fixes qui sont des paramètres associés à la population étudiée dans son ensemble et des effets aléatoires qui sont associés à chaque unité expérimentale (dans le cas des données de croissance, à chaque enfant) et qui représentent la déviation de cette unité par rapport à la population. En associant des effets aléatoires aux observations d'un même individu, les modèles aléatoires représentent, de manière flexible, la structure de covariance induite par le caractère hiérarchique des données (Pinheiro and Bates, 2000).

1.2.1.2 Méthodes de modélisation « groupales »

Ces méthodes dites « groupales » reposent sur des théories taxonomiques. Elles considèrent qu'une seule forme de trajectoires de croissance ne convient pas à l'ensemble des enfants étudiés. Leur objectif est en effet d'identifier différents groupes d'enfants partageant des profils de croissance similaires et de définir les caractéristiques des individus au sein de ces « clusters ». Les principales méthodes « groupales » décrites dans la littérature sont les méthodes de Nagin et coll. (i.e. group-based trajectory modeling, GBTM) et Muthén et coll. (i.e. growth mixture modeling, GMM). En recherche clinique, l'intérêt de ces approches a été souligné en mettant en avant la coexistence de trajectoires normales et pathologiques au sein des populations. Ces méthodes diffèrent des méthodes individuelles dans la façon dont est modélisée l'hétérogénéité dans les trajectoires de développement au niveau individuel (Nagin et al., 1999).

Modèles de mélange (Growth Mixture Modeling GMM)

Cette méthode a été développée par Muthén et Shedden (Muthén and Shedden, 1999). Ces modèles sont utilisés pour analyser les données d'une population que l'on pense être composée de différents sous-groupes ne pouvant pas être identifiés à l'aide de caractéristiques mesurées au préalable et dont l'évènement d'intérêt est généré par des processus différents. Si l'on se place dans le contexte de l'analyse de trajectoires de croissance, ces sous-populations constitueraient des groupes avec des trajectoires moyennes différentes ne pouvant pas être expliquées par la variabilité inter-individuelle représentée par les effets aléatoires d'une modélisation individuelle. Dans l'approche, ce modèle équivaut donc à l'utilisation d'autant de modèles (identiques à ceux d'une

approche individuelle ; c'est-à-dire considérant des modèles à effets aléatoires) que de sous-groupes et à l'estimation de la proportion de la population suivant chaque modèle. Le nombre de sous-groupes est défini par la personne effectuant les analyses et non pas estimé à partir des données (Nagin et al., 1999).

Modèles basés sur des groupes de trajectoires (Growth Based Trajectory Modeling GBTM)

Cette méthode est une méthode semi-paramétrique. Contrairement aux GMM, les modèles GBTM ne prennent pas position sur le nombre de groupes de trajectoires et s'appuient sur la méthode du maximum de vraisemblance pour définir les groupes et inférer les paramètres du modèle. A la différence des modèles GMM, les modèles GBTM n'incluent pas d'effets aléatoires dans les sous-groupes de trajectoires. Le fait d'autoriser des effets aléatoires au sein des groupes dans l'approche GMM peut diminuer le nombre de groupes en permettant une plus grande variabilité intra-groupe des trajectoires individuelles que ce que ne permet l'approche GBTM.

Les principaux avantages de cette méthode sont détaillées par Nagin et coll, comme étant sa capacité à identifier le nombre de groupes le plus probable sur la base des observations plutôt qu'imposer un nombre de groupes de trajectoires ; la capacité à estimer la proportion de la population suivant la trajectoire de chaque groupe ; la capacité à relier la probabilité d'appartenance à un groupe à des caractéristiques individuelles ; la capacité à utiliser les probabilités individuelles d'appartenance aux groupes pour la création de profils (Nagin et al., 1999).

1.2.2 Principaux modèles pour la modélisation de la croissance

Deux types de modèles peuvent être distingués (Hauspie et al., 2004):

Les modèles non structuraux ne font pas d'hypothèse sur la forme de la courbe de croissance à modéliser. Ils font appel en général à un grand nombre de paramètres n'ayant pas d'interprétation biologique et ont été décrits comme souvent instables aux extrémités de la période sur laquelle la modélisation est effectuée. Les polynômes fractionnaires et les splines font partie de ce type de modèles.

Les modèles structuraux supposent une forme spécifique du modèle de croissance. Avec ces modèles, la courbe de croissance est entièrement décrite par un jeu

de paramètres ayant une interprétation biologique. Ils tendent en général vers une asymptote. L'avantage représenté par la considération d'une forme de croissance « type » peut toutefois être contrebalancé par un désavantage qui est d'imposer une forme de courbe trop rigide et qui peut engendrer un biais dans l'ajustement des courbes de croissance ne correspondant pas à ce profil de croissance.

1.2.2.1 Les modèles structuraux

Plusieurs modèles structuraux ont été décrits pour la modélisation des croissances pondérale et staturale dans la petite enfance, dans l'enfance, au moment de la puberté ou pour l'ensemble de la croissance de la naissance à l'âge adulte.

Les modèles permettant de modéliser la croissance dans les premières années de vie, qui sera la période ciblée dans nos travaux, sont détaillés ici. Les formules sont exprimées telles qu'elles ont été décrites par leurs auteurs. Cependant, dans une approche de modélisation individuelle, la variable anthropométrique étudiée est indiquée par i (pour l'individu) et j (le temps de mesure) tout comme l'est le temps t à la mesure de la variable. Les paramètres spécifiques des modèles sont indicés sur l'individu i .

1.2.2.1.1 Modèle de Jenss et Bayley

Le modèle de Jenss et Bayley a été proposé en 1937 (Jenss and Bayley, 1937). Ce modèle permet de modéliser la croissance des enfants de leur naissance à 8 ans.

L'équation proposée par Jenss et Bayley est la suivante :

$$y = a + bt - \exp(c + dt)$$

Où y représente la taille ou le poids, t l'âge et a , b , c et d sont les quatre paramètres du modèle. La partie $(a + bt)$ du modèle décrit la croissance dans la partie linéaire de la courbe de croissance du poids ou de la taille et le paramètre b représente la vitesse de croissance durant cette période. La partie exponentielle du modèle $\exp(c + dt)$ représente le taux de ralentissement de la croissance observé peu de temps après la naissance (correspondant à la courbure des courbes de croissance du poids et de la taille) (Hauspie et al., 2004).

Une adaptation du modèle de Jenss-Bayley a été proposée pour estimer la croissance des enfants de la naissance à 12 ans (Botton et al., 2008). Ce modèle diffère de l'équation initiale de Jenss-Bayley par l'addition d'un terme quadratique sur l'âge. L'équation de ce modèle est la suivante :

$$y = a + bt - \exp(c + dt) + e.t^2$$

1.2.2.1.2 Modèle de Count

Le modèle de Count est un modèle linéaire en paramètres. Ce modèle a été proposé en 1942-1943.

$$y = a + bt - c \ln(t + 1)$$

Où y représente la taille ou le poids, t l'âge (en jours) et a , b et c sont les 3 paramètres du modèle.

L'interprétation des paramètres est assez similaire à celle des paramètres du modèle de Jentsch-Bayley. Le paramètre a décrit la valeur à la naissance du paramètre modélisé ; b représente la vitesse de croissance et le paramètre c est la composante principale de la croissance rapide observée précocement dans l'enfance (Berkey, 1982).

1.2.2.1.3 Modèle de Berkey et Reed

Berkey et Reed ont ensuite amélioré la flexibilité du modèle de Count en ajoutant un ou plusieurs termes de décélération.

Le modèle de Reed de 1^{er} ordre s'écrit

$$y = a + bt - c \ln(t) + \frac{d}{t}$$

Le modèle de Reed de 2^e ordre s'écrit

$$y = a + bt - c \ln(t) + \frac{d_1}{t} + \frac{d_2}{t^2}$$

Où y représente la taille ou le poids, t l'âge et a , b , c , d et a , b , c , d_1 et d_2 sont les paramètres des modèles respectifs.

Comme le modèle de Count, un des avantages des modèles de Reed-Berkey est qu'ils sont linéaires en paramètres et de ce fait peuvent être utilisés avec des méthodes statistiques plus simples que les modèles non-linéaires.

La performance de ces modèles pour la modélisation de la croissance a été comparée à plusieurs reprises en se basant sur des populations différentes.

Les modèles de Count, de Reed de 1^{er} ordre, de Reed de 2^e ordre, de Karlberg et de Kouchi (ces deux derniers modèles n'ont pas été présentés ci-dessus car ils sont

spécifiques des 2-3 premières années uniquement) ont été comparés pour la modélisation de la croissance pondérale d'enfants Congolais dans les 13 premiers mois de vie (Simondon, 1992; Simondon et al., 1992). Le modèle s'ajustant le mieux aux courbes de croissance de ces enfants était le modèle de 1^{er} ordre de Reed. Les auteurs comparaient également la capacité des différents modèles à prendre en compte le phénomène de perte de poids néonatale observé dans les premiers jours de vie. Les modèles de 1^{er} et 2^e ordre de Reed permettaient de considérer la perte de poids néonatale car les 1 ou 2 paramètres supplémentaires comparativement aux autres modèles permettent 1 ou 2 points d'inflexion supplémentaires (Simondon, 1992; Simondon et al., 1992). Berkey et coll. ont montré que le modèle de Jentsch-Bayley ajustait mieux les courbes de croissance d'enfants américains de la naissance à 6 ans que ne le faisait le modèle de Count (Berkey, 1982).

Chirwa et coll. ont quant à eux comparé les performances des modèles de Berkey-Reed, Count et Jentsch-Bayley pour la modélisation de la croissance d'enfants Sud-Africains de la naissance à 10 ans. Le modèle de Berkey-Reed s'ajustait bien aux données alors que le modèle de Jentsch-Bayley ne s'ajustait pas correctement aux données de croissance de taille dans les 1^{res} années de vie (Chirwa et al., 2014).

Enfin, Regnault et coll. ont comparé les modèles de Jentsch-Bayley, l'adaptation du modèle de Jentsch-Bayley décrite par Botton et coll. (cf. paragraphe 1.2.2.1.1 page 29) et les modèles de 1^{er} et 2^e ordre de Reed pour la modélisation de la croissance du poids et de la taille d'enfants vivants aux Etats-Unis de leur naissance à 9 ans. L'adaptation du modèle de Jentsch-Bayley (Botton et al., 2008), était le modèle s'ajustant au mieux aux données de croissance staturo-pondérale (Regnault et al., 2014).

Ces résultats montrent que selon la nature des données, les populations étudiées, les performances des différents modèles peuvent varier et, dans un but de comparaison, l'adéquation de ces modèles doit donc être étudiée dans le contexte de chaque étude.

1.2.2.1.4 SITAR (SuperImposition by Translation And Rotation)

Beath et coll. ont décrit un modèle « shape invariant » (supposant une forme identique) permettant d'analyser les courbes de croissance (Beath, 2007) qu'ils ont appliqué dans un premier temps pour l'analyse de données de poids dans la petite enfance.

L'idée de base d'un modèle « shape invariant » est qu'une population a une forme de courbe commune qui, par déplacement et « mise à l'échelle », peut représenter la forme de courbe de n'importe quel individu.

Ce modèle permet de modéliser les trajectoires de croissance individuelles en appliquant à une courbe de croissance unique, 3 transformations (2 translations et une rotation) propres à chaque sujet.

Le modèle a ensuite été décrit par Cole et coll. et appliqué à la modélisation de la taille (Cole et al., 2010).

Le modèle s'écrit de la manière suivante :

$$y_{it} = \alpha_i + h\left(\frac{t - \beta_i}{\exp(-\gamma_i)}\right)$$

Où y_{it} représente la variable d'intérêt (poids, taille, IMC) pour un sujet i à l'âge t , $h(t)$ représente la courbe de croissance caractéristique et est ici un spline cubique, $\alpha_i, \beta_i, \gamma_i$ sont les paramètres propres à chaque sujet.

La Figure 3 illustre la signification des 3 paramètres du modèle.

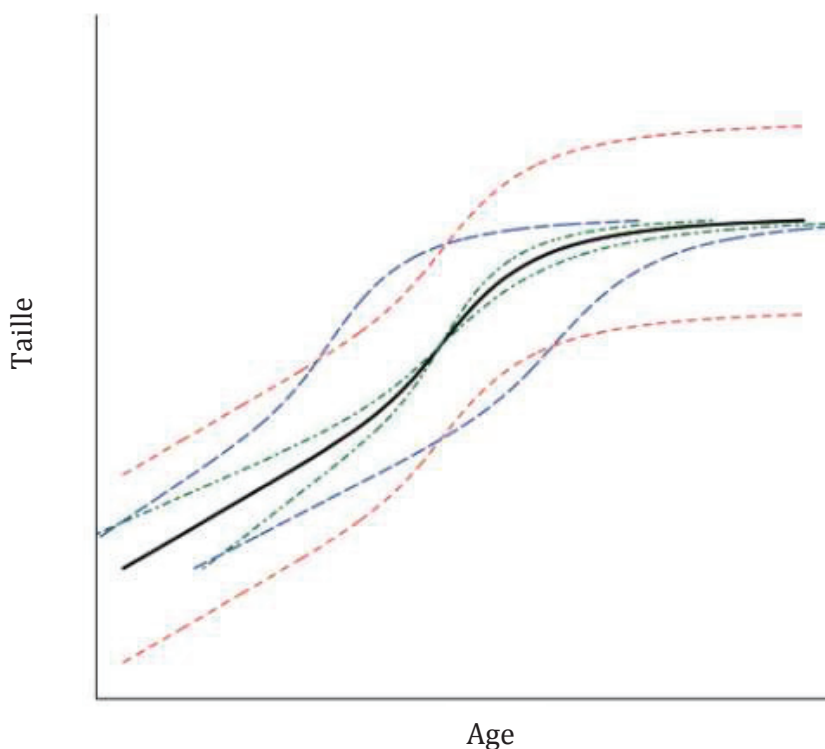


Figure 3 : Illustration de l'effet de chaque paramètre du modèle de croissance sur la trajectoire de croissance de la taille

- La ligne noire représente la courbe initiale (la courbe moyenne au sein de l'échantillon d'étude), elle est décrite par une fonction spline (cf paragraphe 2.2.2.2.2).
- Les lignes rouges illustrent une translation de la courbe vers le haut ou vers le bas, induite par la modification du paramètre α . Ce paramètre peut donc être considéré comme un intercept aléatoire permettant d'ajuster la différence de taille moyenne. Un α_i plus faible permettra de caractériser des individus plus petits dans cet exemple et un α_i plus élevé des individus plus grands.
- Les lignes bleues illustrent l'effet du paramètre β qui correspond à une translation vers la droite ou vers la gauche de la courbe moyenne et correspond à un « timing » de croissance différent.
- Les lignes vertes illustrent l'effet du paramètre γ qui par agrandissement ou rétrécissement de l'échelle de l'âge permet de modéliser des vitesses de croissance plus élevées ou plus faibles.

Ce modèle a, depuis, été utilisé pour la modélisation de la croissance staturale, pondérale et de l'IMC et pour l'étude de ses associations avec différents déterminants (Johnson et al., 2011; Jones-Smith et al., 2013; Pizzi et al., 2014; Willemsen et al., 2015).

1.2.2.2 Les modèles non structuraux

1.2.2.2.1 Les polynômes fractionnaires

Les polynômes fractionnaires ont été décrits pour la première fois par Royston et Altman (Royston and Altman, 1994). Ils ont par la suite été utilisés pour modéliser l'évolution du poids, de la taille et de l'IMC au cours du temps, dans le but d'étudier les déterminants de cette croissance et/ou ses associations avec des caractéristiques (principalement le risque de surpoids et d'obésité) à l'âge adulte (Chivers et al., 2010; Guo et al., 2000; Tilling et al., 2014; Whitaker et al., 1998; Williams et al., 1999). Les polynômes fractionnaires permettent de modéliser la courbure, i.e. la relation non linéaire existant entre ces trois indicateurs et l'âge en incluant un nombre m de fonctions non-linéaires de l'âge dans le modèle (m définissant le degré ou l'ordre du polynôme).

En pratique, le choix est, dans la plupart des cas, à faire entre un polynôme d'ordre 1 ou 2, les polynômes d'ordre 3 ou plus n'étant que rarement utilisés (Royston and Altman, 1994).

Utilisant un polynôme fractionnaire de degré m , l'espérance de la variable d'intérêt Y (poids, taille ou IMC) est modélisée comme une fonction de l'âge (X), comme décrit ci-dessous en accord avec la 1^{re} description de Royston et coll. (Royston and Altman, 1994):

On définit un polynôme de degré m comme la fonction

$$\phi_m(X; \xi; p) = \xi_0 + \sum_{j=1}^m \xi_j X^{(p_j)}$$

Où $p = (p_1, \dots, p_m)$ est un vecteur de puissances avec $p_1 < \dots < p_m$ et $\xi = (\xi_1, \dots, \xi_m)$ sont des coefficients.

$$\text{On a } X^{(p_j)} \begin{cases} X^{(p_j)} & \text{si } p_j \neq 0 \\ \ln X & \text{si } p_j = 0 \end{cases}$$

Lorsqu' une même puissance est répétée plusieurs fois, on a alors pour un polynôme de degré 2 ou plus avec $p = (p_1, \dots, p_m)$

$$\phi_m(X; \xi; p) = \xi_0 + \xi_1 X^{(p_1)} + \sum_{j=2}^m \xi_j X^{(p_j)} \ln(X)^{j-1}$$

Étendant ces deux notations pour un polynôme ayant des puissances arbitraires (toutes différentes, toutes égales ou certaines égales et d'autres différentes), on obtient l'expression suivante :

Pour des puissances arbitraires $p_1 \leq \dots \leq p_m$, on pose $H_0(X) = 1$ et $p_0 = 0$

En combinant les deux expressions ci-dessus on a :

$$\phi_m(X; \xi; p) = \sum_{j=0}^m \xi_j H_j(X)$$

$$\text{Avec } X^{(p_j)} \begin{cases} X^{(p_j)} & \text{si } p_j \neq p_{j-1} \\ H_{j-1}(X) \ln(X) & \text{si } p_j = p_{j-1} \end{cases}$$

Les valeurs candidates du vecteur de puissance sont tous les m -tuples possibles selectionnes avec remise dans un jeu de 8 valeurs candidates 2, -1, -0,5, 0 ou log, 0,5, 1, 2 et 3. Ce jeu de puissance offre une grande flexibilité et un large panel de formes de courbes.

Pour illustration, pour un polynôme d'ordre 3 avec la puissance 2 répétée 3 fois ($p_1 = p_2 = p_3 = 2$) on aura donc :

$$\emptyset_3(X; \xi; p) = \xi_0 + \xi_1 X^2 + \xi_2 X^2 \ln(X) + \xi_3 X^2 (\ln(X))^2$$

Pour un polynôme d'ordre 3 avec la puissance 2 répétée 2 fois et la puissance 3 une fois ($p_1 = p_2 = 2$ et $p_3 = 3$) on aura :

$$\emptyset_3(X; \xi; p) = \xi_0 + \xi_1 X^2 + \xi_2 X^2 \ln(X) + \xi_3 X^3$$

En pratique, pour un même degré de polynôme, on privilégiera le modèle maximisant la vraisemblance. Pour comparer deux polynômes de degrés différents (après avoir choisi la meilleure combinaison de puissances pour chacun des degrés d'intérêt) et juger ainsi de l'apport d'un modèle de degré plus élevé pour modéliser les données, on utilise, dans le cas de modèles emboîtés, un test de rapport de vraisemblance. Ce test comparant un polynôme de degré m et un polynôme de degré $m + 1$ suit une loi du Chi2 à 2 degrés de liberté (1 pour le coefficient de régression supplémentaire à estimer et 1 pour la puissance supplémentaire).

D'une manière générale, les polynômes de faibles degrés n'offrent qu'un faible nombre de formes de courbes possibles. Les polynômes de degrés plus importants s'ajustent mieux aux données d'une façon générale, mais ne s'ajustent souvent pas bien aux extrêmes des intervalles de données et, comme mentionné ci-dessus, ne sont souvent pas meilleurs que les polynômes de degré 2.

Un des désavantages des polynômes fractionnaires, et des modèles non-structuraux d'une manière plus générale, souligné pour l'analyse de la croissance et des déterminants de la croissance en particulier, réside dans la difficulté à interpréter les coefficients des polynômes et, de ce fait, leurs associations avec des covariables d'intérêt pour en tirer une information facilement.

1.2.2.2.2 Les splines

Le mot anglais « spline » désigne une latte flexible utilisée par les dessinateurs pour matérialiser des lignes à courbure variable et passant par des points fixés a priori ou à proximité de ceux-ci (Besse and Thomas-Agnan, 1989). L'idée originale est attribuée à Whittaker (Whittaker, 1922). Les fonctions splines offrent une grande flexibilité et peuvent modéliser des formes de relation très diverses.

Les splines sont des fonctions polynômes par morceaux. Ils modélisent la relation entre une variable Y et une variable X par intervalle. Ainsi, si on considère $k + 1$ intervalles, les bornes des intervalles étant appelés nœuds (à l'exclusion de la valeur

minimale et maximale de X), la relation entre Y et X au sein de chaque intervalle est modélisée par un polynôme de degré m . En général, au niveau des nœuds, les valeurs des fonctions polynômes ainsi que les $m - 1$ premières dérivées sont contraintes à être égales (Pan and Goldstein, 1998).

Une fonction spline s'écrit de la manière suivante :

$$S(x) = \beta_0 + \sum_{i=1}^m \alpha_i x^i + \sum_{j=1}^k \beta_j P_m(x; s_j)^m$$

Où m représente le degré du polynôme, et k représente les nœuds placés aux valeurs de x (aux âges dans notre cas), s_1, \dots, s_k .

$$\text{On a } P_m(x; s_j) = (x - s_j)_+^m = \begin{cases} (x - s_j)^m & \text{si } x \geq s_j \\ 0 & \text{sinon} \end{cases}$$

Dans la plupart des cas, les splines utilisées sont les splines cubiques (Pan and Goldstein, 1998). Ils modélisent la relation entre X et Y par des polynômes de degré 3. Ils offrent un bon compromis entre flexibilité et parcimonie.

Cependant d'autres types de splines peuvent être utilisés. Les splines linéaires, notamment, ont été utilisés à plusieurs reprises pour la modélisation des données de croissance en fonction de l'âge (Anderson et al., 2013; Ben-Shlomo et al., 2008; Howe et al., 2013, 2010; Tilling et al., 2014). Ils représentent une série de droites reliées au niveau des nœuds définis préalablement. C'est le seul cas des modèles par splines où les coefficients sont interprétables (intercept et taux de croissance sur les différentes périodes définies par les intervalles entre nœuds). De ce fait, les associations entre les expositions d'intérêt, par exemple, et la croissance sont plus faciles à exploiter.

L'utilisation des splines linéaires impose toutefois une relation linéaire par morceau entre la croissance et l'âge, ce qui repose sur des hypothèses biologiquement non plausibles. Une fonction spline linéaire avec 3 nœuds s_1, s_2, s_3 s'écrirait :

$$S(x) = \beta_0 + \alpha_1 x + \beta_1(x - s_1)_+ + \beta_2(x - s_2)_+ + \beta_3(x - s_3)_+$$

Avec α_1 , la pente initiale et β_1, β_2 et β_3 représentant les changements de pente à chaque nœud (s_1, s_2, s_3 respectivement).

Choix de l'emplacement et du nombre de nœuds

Plusieurs méthodes ont été proposées pour choisir le nombre et l'emplacement des nœuds. Une des possibilités est de commencer avec un nombre de points importants et de graduellement diminuer ce nombre jusqu'à l'obtention d'une courbe lisse. Une

autre option est de placer des nœuds aux centiles de la distribution de l'âge ou encore d'utiliser des régressions pas à pas en conservant les nœuds c'est-à-dire les coefficients associés aux splines statistiquement significatifs (Howe et al., 2013).

1.2.3 Principales méthodes d'inférences

Les méthodes fréquentistes et bayésiennes sont deux méthodes permettant de faire de l'inférence statistique. En statistique fréquentiste, la probabilité est définie de manière objective et mesure la possibilité physique de réalisation d'un évènement aléatoire (ou fréquence de réalisation d'un évènement). En statistique bayésienne, la définition de la probabilité est subjective et représente le degré de croyance en la réalisation d'un évènement (lié à notre défaut de connaissance).

Le caractère non linéaire des trajectoires de croissance nécessite souvent l'utilisation de modèles non linéaires à effets mixtes (MNLEMs).

Dans cette partie, les principes de ces deux méthodes sont abordés. En ce qui concerne l'approche fréquentiste, plus classique, nous nous attacherons principalement à décrire les algorithmes pouvant être utilisés dans le cas de modèles non linéaires mixtes et principalement l'algorithme SAEM qui sera utilisé dans les analyses.

1.2.3.1 Méthodes d'inférences fréquentistes

Dans la démarche fréquentiste, le vecteur de paramètres θ permettant de modéliser les données observées est considéré fixe mais inconnu. L'estimation des paramètres repose sur la théorie du maximum de vraisemblance. L'intervalle de confiance des paramètres est basé sur la répétition de l'échantillonnage un grand nombre de fois. L'intervalle de confiance à 95% exprime l'idée que si on pouvait répéter l'échantillonnage, 95% des intervalles de confiance contiendraient (au sens de fréquence) la vraie valeur.

Dans les modèles non linéaires à effets aléatoires (NLMEMs), la fonction du modèle n'étant pas linéaire en les effets aléatoires, la vraisemblance n'a, en général, pas d'expression analytique ce qui rend l'estimation des paramètres complexe. Plusieurs approximations de la fonction de vraisemblance ont été proposées. L'algorithme « Espérance-Maximisation » (EM) proposé pour la première fois par Dempster et coll., est un algorithme fréquemment utilisable pour le calcul des estimateurs du maximum de vraisemblance dans ce contexte (Dempster et al., 1977). Des versions stochastiques de

l'algorithme EM comme l'algorithme Monte Carlo EM (MCEM) ont été proposées (Levine and Casella, 2001; Wei and Tanner, 1990). L'algorithme SAEM est un de ces algorithmes, plus efficace que l'algorithme MCEM (ce dernier peut présenter des problèmes calculatoires et converger lentement ou ne pas converger), dont la phase E est divisée en une étape de simulation et une étape d'approximation stochastique. Celui-ci a été implémenté dans R dans le package SAEMIX et a prouvé ses capacités à fournir des estimations précises dans le cas de modèles non linéaires à effets aléatoires (Samson et al., 2007).

1.2.3.2 Méthodes d'inférences bayésiennes

Dans ce chapitre, sont résumés les grands principes de la démarche d'analyse bayésienne décrits dans la littérature (Gelman et al., 2013; Lunn et al., 2012; Makowski, 2012).

Gelman et coll. divisent le processus d'analyse bayésienne des données en 3 étapes :

1. La création d'un modèle de probabilité. Ce modèle doit tenir compte des connaissances sur le problème étudié et du processus de collecte des données.

Les connaissances **a priori** sur les paramètres impliqués dans un modèle sont résumées sous la forme de **distribution/loi a priori** ou **prior**. Les paramètres du modèle sont considérés comme des variables aléatoires et non plus comme ayant une valeur fixe (comme dans l'approche fréquentiste). La prior caractérisera à la fois notre degré de connaissance à leur égard (incertitude) et leur caractère plus ou moins aléatoire (variabilité) avant de prendre en compte les données. La loi a priori sera définie en fonction de notre connaissance sur l'ordre de grandeur et la plage de variation la plus probable des paramètres à estimer (incertitude). Ainsi, les lois a priori seront dites **informatives** si nous disposons d'une information relativement précise sur le paramètre d'intérêt et l'apportons pour la modélisation ou **non informatives** si l'information sur les valeurs possibles du paramètre d'intérêt est assez vague ou si on ne dispose pas d'information sur ce paramètre.

2. Conditionner sur les données observées : en calculant et en interprétant les distributions a posteriori appropriées, i.e. les distributions de probabilité des quantités non observées conditionnellement aux données observées.

La **distribution a posteriori** ou **posteriore** des paramètres du modèle représente la synthèse de l'information a priori et de l'information apportée par les données. Ainsi, la connaissance a priori est mise à jour par l'information apportée par les données et permet l'obtention d'une distribution a posteriori conditionnelle aux données. La loi a posteriori sera plus ou moins modifiée (plus resserrée si les données apportent de la précision, décalée si les données modifient les valeurs des paramètres) par rapport à la loi a priori en fonction de l'importance de l'information apportée par les données. Cette distribution a posteriori permettra de résumer l'information sur les paramètres. On pourra notamment déduire la **moyenne a posteriori** du paramètre et son **intervalle de crédibilité à 95%** représentant une probabilité de 95% que le paramètre soit dans cet intervalle. En effet, contrairement au raisonnement fréquentiste (et donc de l'intervalle de confiance), le raisonnement bayésien permet d'exprimer la probabilité qu'un intervalle contienne une quantité d'intérêt non connue.

Cette mise à jour est possible grâce à l'application du **théorème de Bayes**.

Le théorème de Bayes peut s'exprimer de la façon suivante :

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y, \theta)}{p(y)}$$

Où $p(y) = \sum_{\theta} p(y|\theta)p(\theta)$ ou $p(y) = \int p(y|\theta)p(\theta) d(\theta)$ lorsque θ est continu.

Où p dénote une densité de probabilité plutôt que la simple probabilité d'un événement. $p(\theta)$ est la distribution a priori des paramètres θ du modèle (et exprime l'incertitude sur les valeurs de θ avant de prendre en compte les données) ; $p(\theta|y)$ est la distribution a posteriori de θ (et représente l'incertitude sur θ conditionnellement aux données) ; la loi conjointe des paramètres et des données $p(y, \theta)$ peut être obtenue à partir de ces distributions. La distribution conditionnelle $p(y|\theta)$ décrit de quelle manière les données dépendent des valeurs des paramètres (vraisemblance des données). Il n'est habituellement pas nécessaire de calculer $p(y)$ pour évaluer les propriétés de la posteriore et le théorème peut s'écrire (Lunn et al., 2012):

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

Le calcul des lois a posteriori peut être difficile. Les calculs d'intégrales multiples, liées à la présence de plusieurs paramètres, rendent les calculs difficiles à obtenir explicitement. Les méthodes numériques permettent de les approximer. Elles permettent de générer un échantillon issu de la loi a posteriori sans passer par ce calcul d'intégrales multiples (lorsqu'il n'est pas possible d'échantillonner θ directement dans sa loi a posteriori $p(\theta|y)$). Les méthodes de **Monte Carlo par chaînes de Markov** sont des techniques permettant ces approximations. Par définition une chaîne de Markov est une séquence de variables aléatoires $\theta^1, \theta^2, \dots$ dans laquelle pour n'importe quelle valeur de t , la distribution de θ^t sachant l'ensemble des valeurs précédentes θ^s ne dépend que de la valeur la plus récente θ^{t-1} . Cette méthode itérative permet de tirer des valeurs de θ dans une distribution approximative et de corriger ces valeurs, à chaque simulation (**itération**), pour mieux approcher/converger vers la distribution a posteriori. L'échantillonnage est fait de manière séquentielle, la distribution des paramètres échantillonnés dépendant de la dernière valeur échantillonnée (selon la définition précédemment énoncée). Plusieurs méthodes de simulations par chaîne de Markov existent. Les principales utilisées en analyse bayésienne sont les algorithmes de Metropolis-Hastings et de Gibbs. Avec l'algorithme de Metropolis-Hastings, une loi de proposition (ou loi instrumentale ou candidate) choisie à partir de la densité cible et à partir de laquelle il est assez facile de simuler, permet la simulation d'une réalisation qui est conservée ou rejetée selon la loi a posteriori recherchée. La mise à jour des paramètres n'est donc pas systématique et est soumise à une étape d'acceptation-rejet. L'algorithme de Gibbs est ici présenté plus en détail car il sera utilisé dans les analyses ultérieures.

L'échantillonneur de Gibbs échantillonne selon les lois conditionnelles complètes des paramètres. Dans le cadre de ce processus itératif, après avoir choisi des valeurs de départ pour les composantes du vecteur de paramètres θ , à chaque itération t , chaque paramètre est échantillonné dans la distribution conditionnelle aux valeurs des autres paramètres à l'itération $t - 1$. C'est à dire qu'à chaque itération, chaque sous-vecteur θ_j du vecteur global des paramètres θ (l'ensemble des paramètres à estimer) sera échantillonné conditionnellement aux valeurs les plus récentes des autres

composantes de θ . A l'itération t , la valeur de la 1^{re} composante de θ , θ_1^t , sera donc générée selon la loi $f(\theta_1^t | \theta_1^{t-1}, \dots, \theta_j^{t-1})$ (Gelman et al., 2013; Lunn et al., 2012).

En pratique, l'échantillonneur de Gibbs est utilisé pour générer plusieurs chaînes de Markov initiées avec des valeurs de départ (**valeurs initiales**) différentes pour les paramètres. L'objectif est que les échantillons tirés dans la loi a posteriori de ces chaînes soient utilisés pour réaliser les inférences (résumer la densité a posteriori, les quantiles et toutes autres informations d'intérêt), une fois que celles-ci ont convergé vers la distribution a posteriori. Les premières itérations reflètent davantage les approximations de départ (et dépendent des valeurs initiales) plus qu'elles ne représentent la distribution cible. Il convient donc de définir une **phase de chauffe**, i.e. un nombre d'itérations suffisant pour supprimer l'influence des valeurs initiales et atteindre la convergence des chaînes (la bonne couverture de la loi a posteriori). Les valeurs échantillonnées durant cette phase de chauffe seront supprimées et les chaînes seront prolongées pour récupérer les échantillons reflétant la distribution a posteriori. Les chaînes doivent converger en distribution et vers la même limite. L'utilisation d'une seule chaîne pourrait mener à conclure à tort à la bonne convergence. L'utilisation de plusieurs chaînes va permettre d'évaluer la convergence en inspectant leur mélange (et donc leur couverture de la même distribution) et leur stationnarité.

Le diagnostic de convergence des chaînes peut s'effectuer grâce à :

L'inspection visuelle de la stationnarité et du mélange des 3 chaînes initiées avec des valeurs différentes en observant les valeurs simulées pour chaque paramètre d'intérêt en fonction du nombre d'itérations.

Le diagnostic de Gelman et Rubin (Cowles and Carlin, 1996; Gelman and Rubin, 1992; Lunn et al., 2012). Ce critère est basé sur la comparaison, par leur rapport, des variances intra- et inter-chaînes des valeurs simulées des paramètres et une valeur proche de 1 indique une bonne convergence. En pratique, la convergence peut être supposée atteinte lorsque la valeur de ce critère est inférieure à 1,05.

La vérification de la stabilité par rapport aux itérations de quelques statistiques comme la moyenne et les quantiles.

En parallèle des problèmes de convergence, une corrélation trop importante entre les échantillons tirés dans la loi a posteriori couverte par une même chaîne (problème de mélange de la chaîne), peut diminuer la précision des estimations par comparaison aux estimations réalisées à partir de séquences indépendantes et nécessiter un nombre

d'itérations plus important pour couvrir l'ensemble de la distribution. Il est possible de ne pas échantillonner la chaîne à chaque itération mais de définir un intervalle (un nombre d'itérations) séparant deux échantillonnages successifs.

Le graphique des auto-corrélations permet aussi de détecter la bonne mélangeance des chaînes.

3. Evaluer la qualité d'ajustement du modèle et les implications des distributions a posteriori. Cela inclus la qualité du modèle à s'ajuster aux données ; l'analyse de la plausibilité des conclusions et l'analyse de la sensibilité des résultats aux hypothèses posées à l'étape 1.

- a. Evaluation de la qualité d'ajustement d'un modèle : vérification prédictive a posteriori

Le principe repose sur le fait que si la qualité d'ajustement du modèle est bonne, les données répliquées, générées sous le modèle, devraient ressembler aux données observées. Ceci peut également s'exprimer en disant que les données observées devraient être plausibles sous la distribution (prédictive) a posteriori. Ainsi, pour évaluer la qualité d'ajustement d'un modèle aux données, une technique est de tirer des données dans les distributions a posteriori et de comparer cet échantillon aux données observées. Des différences systématiques entre les données répliquées et les données observées indiquent une faiblesse potentielle du modèle.

Plus précisément, le principe, décrit par Gelman et coll., est de générer des données répliquées à partir de la distribution a posteriori des paramètres θ et de les comparer aux données observées. Nous appellerons y^{rep} les données répliquées basées sur le même modèle et les valeurs des paramètres produisant les données. Pour mesurer la discrédance (divergence) entre le modèle et les données, la p-value bayésienne peut être définie en utilisant une fonction de discrédance du type Chi2.

Nous désignons cette fonction $T(y, \theta)$ et $T(y^{rep}; \theta)$ pour les données observées et les données répliquées respectivement. Gelman et coll. définissent la p-value comme étant la probabilité que les données répliquées soient au moins aussi extrêmes que les données observées et l'expriment comme $p_B = \Pr(T(y^{rep}, \theta) \geq T(y, \theta) | y)$.

En pratique, la distribution prédictive a posteriori est calculée à partir d'un certain nombre de simulations (itérations retenues). Si on note S ce nombre de

simulations, on tirera les données répliquées y^{rep} dans la distribution prédictive pour chaque θ simulé. On aura donc S tirages de la loi a posteriori jointe $p(y^{rep}, \theta|y)$.

La p-value exprime donc ici la proportion de simulations où la statistique de test calculée à partir des données répliquées est supérieure ou égale à celle obtenue à partir des données observées, c'est à dire quand $T(y^{rep s}, \theta^s) \geq T(y, \theta^s), s = 1 \dots S$. Dans notre cas, nous définissons la fonction T comme la somme standardisée des carrés des résidus des données par rapport aux valeurs attendues:

$T(y, \theta) = \sum_{i=1}^N \sum_{j=1}^n \frac{(y_{ij} - E(y_{ij}|\theta))^2}{Var(y_{ij}|\theta)}$ pour la statistique calculée à partir des données observées et $T(y^{rep}, \theta) = \sum_{i=1}^N \sum_{j=1}^n \frac{(y^{rep}_{ij} - E(y_{ij}|\theta))^2}{Var(y_{ij}|\theta)}$ pour la statistique calculée à partir des données répliquées où $E(y_{i,j}|\theta) = f(\phi_i, t_{ij})$ est la valeur prédite pour l'individu i au temps t_{ij} ayant pour paramètres individuels ϕ_i et $Var(y_{ij}|\theta) = \sigma_{e_k}^2$.

Une p-value prédictive a posteriori (« posterior predictive p-value ») proche de 0,50 (50% des statistiques obtenues avec les données répliquées supérieures/inférieures à celles obtenues avec les données observées) sera en faveur d'une bonne qualité prédictive. A l'inverse des valeurs proches de 0 ou 1 indiqueront un possible problème d'ajustement.

b. Comparaison de l'adéquation aux données de plusieurs modèles

Il y a plusieurs manières de comparer des modèles dont l'inférence a été réalisée par une approche bayésienne et d'évaluer leur capacité à s'ajuster aux données. Cette comparaison peut se faire à l'aide du facteur de Bayes (BF) basé sur la vraisemblance marginale. Considérant deux modèles M_0 et M_1 , le facteur de Bayes, testant l'hypothèse nulle que le modèle M_0 est le « vrai modèle », est le rapport des ratios des probabilités a posteriori sur a priori de ces modèles:

$$BF = \frac{p(M_0|y)/p(M_0)}{p(M_1|y)/p(M_1)}$$

En appliquant le théorème de Bayes on a :

$$BF = \frac{p(y|M_0) \int p(y|\theta, M_0)p(\theta|M_0)d(\theta)}{p(y|M_1) \int p(y|\theta, M_1)p(\theta|M_1)d(\theta)}$$

En pratique, ce facteur ne nécessite pas que les modèles comparés soient emboîtés. Un facteur de Bayes sera faiblement, assez, fortement ou de manière décisive en faveur du modèle M_0 s'il est, respectivement, compris entre 1 et 3 ; 3 et 12 ; 12 et 150 ou est supérieur à 150.

L'intégration de la vraisemblance par rapport aux lois a priori peut être complexe et de ce fait le calcul du facteur de Bayes est souvent difficile. Il n'est pas défini dans le cas de loi a priori impropres (c'est-à-dire conduisant à une mesure infinie sur l'espace des paramètres plutôt qu'à une mesure de probabilité) et est sensible aux choix des lois a priori.

Des alternatives au facteur de Bayes peuvent être utilisées pour comparer plusieurs modèles. Spiegelhalter et coll. ont proposé le critère de déviance DIC (« Deviance Information Criterion ») (Spiegelhalter et al., 2002) et Watanabe, le WAIC (« Widely Applicable Information Criterion ») (Watanabe, 2010). Le principe de ces deux critères est de comparer la précision des estimations des modèles d'intérêt. Plus précisément, il s'agit de comparer la précision des prédictions de données « nouvelles » (futures) obtenue avec les modèles comparés. Afin de ne pas avoir besoin, d'un nouvel échantillon de données, ces critères se basent sur l'estimation de la précision des prédictions obtenues à partir de l'échantillon de données disponible (« données observées »), à partir du calcul du log de la densité prédictive ou log-vraisemblance $\log(P(y|\theta))$. La précision des prédictions d'un modèle ajusté à des données futures sera en général plus faible, en moyenne, que la précision des prédictions des données observées obtenue avec ce modèle. Ces deux critères, corrigent ce biais, en calculant la densité prédictive et en soustrayant une correction pour le nombre de paramètres ajustés.

Le DIC peut être vu comme une généralisation du critère d'Akaike et est défini de la manière suivante:

$$DIC = \bar{D} + p_{DIC} = D(\bar{\theta}) + 2p_{DIC}$$

Où p_{DIC} mesure la complexité (pénalité) du modèle et est proche du nombre de paramètres et \bar{D} représente la déviance du modèle (moyenne a posteriori de la déviance correspondant à $-2 * \log(Vraisemblance) = -2 * \log(p(y|\theta))$). $D(\bar{\theta})$ est un estimateur ponctuel de la déviance obtenu en remplaçant θ par les moyennes a posteriori

$(-2 * \log(p(y|\bar{\theta})))$). On a donc $p_{DIC} = 2(\log(p(y|E_{post}(\theta))) - E_{post}(\log(p(y|\theta)))$ où E_{post} désigne l'espérance a posteriori.

La comparaison de deux modèles à partir de leur DIC ne nécessite pas qu'ils soient emboîtés. Le modèle ayant le DIC le plus faible sera considéré comme celui-ci ayant la meilleure adéquation aux données car ayant la plus faible déviance corrigée du nombre de paramètres. En pratique, on considérera une diminution comprise au-delà de 10 comme suffisante pour rejeter le modèle avec le DIC le plus élevé et une différence entre 5 et 10 comme substantielle (Lunn et al., 2012). Le DIC a fait l'objet de critiques, liées au fait qu'il n'est pas un critère complètement bayésien et la difficulté à l'estimer pour certains types de modèles tels que les modèles de mélange (Gilles Celeux, 2006).

Le WAIC est, quant à lui, un critère « complètement » bayésien et peut être exprimé de la manière suivante :

$$WAIC = -2 \sum_{i=1}^N \log(E_{post}(p(y_i | \theta))) + 2p_{WAIC}$$

Où p_{WAIC} est le terme de pénalité et est exprimé comme:

$p_{WAIC} = \sum_{i=1}^N Var_{post}(\log(p(y_i | \theta)))$ où Var_{post} représente la variance a posteriori et par conséquent p_{WAIC} est la somme sur l'ensemble des données (N) de la variance des log vraisemblance des données observées. Le critère WAIC a été décrit comme plus stable que le DIC notamment car le calcul de la variance séparément pour chaque donnée observée et la sommation de l'ensemble de ces éléments assurent une certaine stabilité.

c. Analyse de sensibilité aux valeurs initiales et aux lois a priori

Il est également important de vérifier, pour la conduite des analyses, la sensibilité/la robustesse des conclusions tirées des lois a posteriori, aux choix des valeurs initiales et aux distributions a priori spécifiées. Les choix des lois a priori en terme de familles de distributions et de valeurs peuvent être multiples et il convient de vérifier si la modification du choix réalisé modifie l'estimation des paramètres du modèle.

Une fois le modèle choisi, les paramètres estimés (leurs lois a posteriori) peuvent être utilisés pour prédire de futures observations (**prédictions**) pour des individus dont

les variables explicatives (e.g: caractéristiques socio-démographiques considérées comme covariables dans le modèle) sont connues mais pas les événements d'intérêt (i.e : les variables dépendantes du modèle). Comme cela a été mentionné dans ce résumé, un des intérêts majeurs du raisonnement bayésien est qu'il permet de disposer de la distribution a posteriori des paramètres ou caractéristiques d'intérêt qui tient compte de l'incertitude des estimations (De la Cruz-Mesía and Marshall, 2006). Les prédictions calculées à partir de ces paramètres tiendront donc compte *de facto* de cette incertitude.

1.3 Objectifs de la thèse

Les objectifs de cette thèse étaient d'étudier différentes approches de modélisation de la croissance pour l'étude de ses déterminants précoces à partir des données de la cohorte de naissance EDEN. Dans un 1^{er} temps, nous décrirons plus en détail l'étude EDEN et les données de l'étude exploitées pour ce travail. Dans un 2^e temps, nous présenterons une méthode de modélisation permettant d'étudier les associations entre déterminants précoces et croissance. Nous utiliserons l'exemple du tabagisme maternel pendant la grossesse pour illustrer cette méthode. La 3^e partie s'intéressera à la modélisation conjointe de la croissance du poids et de la taille en utilisant une approche bayésienne. Enfin, nous étudierons, en utilisant des modèles bayésiens, si cette modélisation conjointe nous permet de prédire les principales caractéristiques anthropométriques plus précisément que ne le permettent les modélisations séparées du poids et de la taille. Nous étudierons enfin de quelle manière elle peut être utilisée pour étudier la co-évolution entre la croissance pondérale et staturale.

Chapitre 2 : Population et Données

2.1 Présentation générale

L'étude EDEN (Etude des Déterminants pré- et postnatals du développement et la santé de l'Enfant) est une étude de cohorte mère-enfant, initiée en 2003 dans les centres hospitaliers de Nancy et Poitiers (Heude et al., 2015). Cette étude épidémiologique longitudinale française, prévoyait initialement de suivre les mères et leur enfant pendant 5 ans. L'implication des familles a permis la prolongation de l'étude à 8 et 10 ans.

Entre 2003 et 2006, 2002 femmes enceintes ont été incluses et il a été possible de recueillir des données de naissance pour 1907 enfants. Les autres grossesses n'ont pas été menées à terme ou certaines femmes ont quitté l'étude avant la naissance de leur enfant.

Le large éventail de données recueillies lors des examens cliniques des enfants, les informations recueillies auprès des mères et des pères tout au long du suivi des enfants et les échantillons biologiques collectés ont permis d'avoir un bon reflet des multiples aspects de la vie de l'enfant et de son développement. La pluridisciplinarité des informations collectées offre l'opportunité de caractériser finement les facteurs précoces environnementaux (que ce soit l'environnement familial ou les expositions environnementales à des contaminants) susceptibles d'influencer la santé et le développement de l'enfant.

L'objectif principal de l'étude était précisément d'identifier les facteurs pré et post-natals précoces qui influencent le développement et la santé ultérieure de l'enfant. L'étude peut être décomposée en 3 niveaux permettant de préciser l'enchaînement des événements et de faire des hypothèses sur les chaînes de causalité :

- La femme enceinte et le fœtus : étude des relations entre des facteurs d'exposition maternels et de l'état de santé maternel et le développement du fœtus évalué par des méthodes non invasives (échographie, doppler)
- Le nouveau-né : étude des relations entre facteurs d'exposition et de santé maternels et marqueurs du développement fœtal d'une part et l'état de l'enfant à la naissance
- Le jeune enfant : recherche des facteurs les plus prédictifs de l'état de santé de l'enfant d'âge préscolaire, de son développement psychomoteur, cognitif et comportemental, mais aussi de phénotypes intermédiaires associés aux maladies

de l'adulte parmi les paramètres des deux niveaux précédents et ceux relevés pendant les premiers mois de vie.

2.1.1 Inclusion et suivi

2.1.1.1 Inclusion

Les femmes se présentant pour une visite prénatale avant la 24^e semaine d'aménorrhée (SA) dans les départements d'obstétrique et de gynécologie des centres hospitalo-universitaires de Nancy (entre septembre 2003 et janvier 2006) et de Poitiers (entre février 2003 et juin 2005) ont été invitées à participer à l'étude. Parmi les grossesses non gémellaires, les femmes désirant participer à l'étude devaient, pour être incluses, parler et écrire le français, ne pas présenter de diabète insulino-dépendant, ne pas prévoir de déménagement dans les 3 années à venir et bénéficier d'un régime de sécurité sociale.

2.1.1.2 Suivi

2.1.1.2.1 Suivi prénatal

Pendant la période prénatale, deux examens spécifiques à l'étude ont complété la prise en charge habituelle de la grossesse.

Dans le cadre du 1^{er} examen additionnel, les femmes étaient convoquées entre les 24^e et 28^e SA dans les maternités pour un examen clinique et biologique au cours duquel elles devaient répondre à un questionnaire incluant des renseignements sur leur mode de vie, tel que leur statut tabagique actuel ainsi que lors du 1^{er} trimestre de grossesse. Ces questionnaires étaient administrés par les enquêtrices constituées de sages-femmes recrutées et formées pour l'étude. Un auto-questionnaire destiné au recueil d'informations auprès du père leur était également remis.

Dans le cadre du 2^e examen maternel en suite de couches, les femmes remplissaient un auto-questionnaire incluant des informations sur leur statut tabagique au 3^e trimestre de la grossesse, un examen spécifique du nouveau-né était réalisé, ainsi que des prélèvements biologiques (colostrum, cheveux, sang de cordon).

Les autres informations concernant la grossesse et les caractéristiques de l'enfant à la naissance (notamment des données anthropométriques), collectées en routine au sein des maternités étaient extraites du dossier médical (e.g. données échographiques).

2.1.1.2.2 Suivi postnatal

Les auto-questionnaires envoyés aux parents lors du suivi des enfants de la naissance à 5 ans, permettaient de collecter des informations sur l'environnement de l'enfant et sur sa croissance grâce à la retranscription des données mesurées en routine par les médecins de famille dans le carnet de santé.

Au cours de la première année de vie de l'enfant, deux auto-questionnaires postaux ont été envoyés aux parents aux 4^e et 8^e mois de celui-ci.

Aux auto-questionnaires postaux envoyés chaque année aux parents pour le suivi de l'enfant de 1 à 5 ans, s'ajoutaient des examens cliniques (à 1, 3 et 5 ans) lors desquels étaient notamment collectées des informations anthropométriques permettant le suivi de la croissance des enfants.

2.2 Données utilisées

2.2.1 Tabagisme maternel pendant la grossesse

Durant la grossesse, le statut tabagique des mères a été collecté à deux occasions par auto-questionnaire délivré par les enquêtrices. Lors de l'examen clinique ayant eu lieu entre la 24^e et la 28^e SA de la femme, les femmes devaient rapporter leur statut tabagique actuel, ainsi que, pour les fumeuses, des informations quantitatives sur leur consommation (nombre moyen de cigarettes fumées par jour). Elles déclaraient également leur statut tabagique au début de leur grossesse. Après l'accouchement, lors de la visite en suites de couches, les enquêtrices recueillaient via un auto-questionnaire le statut tabagique déclaré par les mères à la fin du 3^e trimestre de grossesse et, chez les fumeuses, était également collecté le nombre moyen de cigarettes fumées par jour.

2.2.2 Données de croissance

2.2.2.1 A la naissance

Le poids à la naissance ainsi que les poids journaliers entre le 1^{er} et le 5^e jour et la taille à la naissance ont été extraits des dossiers obstétricaux remplis dans le cadre du suivi de grossesse habituel. Les poids étaient mesurés avec une balance électronique Seca 737 à Nancy et Seca 335 à Poitiers, qui offrait une précision de 10 grammes. La taille à la naissance était mesurée avec une toise en bois, Testut, Béthune, France.

2.2.2.2 Données extraites des carnets de santé

Les parents étaient chargés, dans chaque auto-questionnaire qu'ils recevaient régulièrement dans le cadre de l'étude, de reporter les mesures de poids et taille faites par le personnel médical dans le cadre du suivi de routine des enfants et retranscrites dans leur carnet de santé. Les mesures de poids et taille concernées étaient celles effectuées dans les 4 mois précédant la date de remplissage des auto-questionnaires pour les questionnaires envoyés au 4^e et 8^e mois et au 1 an de l'enfant ou des quatre précédents trimestres pour les questionnaires envoyés aux 2 et 3 ans de l'enfant. Ils avaient la possibilité de renseigner quatre mesures de poids et quatre mesures de taille. S'ils disposaient de plus de quatre mesures de taille, il leur était demandé de choisir la 1^{re} mesure de chaque mois/trimestre.

Pour l'auto-questionnaire envoyé aux 2 ans de l'enfant, il était demandé aux parents, de reporter les mesures de poids et de taille effectuées par le médecin ou le pédiatre lors de l'examen conseillé à 2 ans.

Dans l'auto-questionnaire envoyé à 4 ans, deux mesures de poids et de taille étaient retranscrites concernant la croissance entre 3 et 4 ans. Si les parents disposaient de plus de deux mesures, les mesures retranscrites concernaient la dernière mesure de chaque semestre. Il leur était également demandé, s'ils n'avaient pas rempli tous les questionnaires précédents, de retrouver dans le carnet de santé, les mesures effectuées vers l'âge de 3 mois et les âges de 1, 2 et 3 ans (à chaque fois, une mesure pouvait être reportée).

Dans l'auto-questionnaire envoyé à 5 ans, les mêmes informations que celles collectées dans l'auto-questionnaire à 4 ans étaient collectées avec en supplément les mesures de poids et de taille vers l'âge de 4 ans, en cas de remplissage non exhaustif des auto-questionnaires précédents.

La Figure 4 représente des extraits des auto-questionnaires envoyés au 4^e et 8^e mois et à la 1^{re} année de l'enfant et de l'auto-questionnaire envoyé aux 3 ans de l'enfant.

Auto-questionnaires 4 et 8 mois et 1 an

La croissance de votre enfant

Merci de reporter les mesures du poids, de la taille et du périmètre crânien de votre enfant qui ont été notées dans son carnet de santé (en dehors de celles de la naissance) (si plus de 4 mesures ont été notées, choisir la première mesure de chaque mois)

Mesure 1 : date (jj/mm/aa) |_|_| |_|_| |_|_| |_|_| |_|_|
 poids (kg) |_|_| |_|_|
 taille (cm) |_|_| |_|_|
 périmètre crânien (cm) |_|_| |_|_| |_|_|

Mesure 2 : date (jj/mm/aa) |_|_| |_|_| |_|_| |_|_| |_|_|
 poids (kg) |_|_| |_|_|
 taille (cm) |_|_| |_|_|
 périmètre crânien (cm) |_|_| |_|_| |_|_|

Mesure 3 : date (jj/mm/aa) |_|_| |_|_| |_|_| |_|_| |_|_|
 poids (kg) |_|_| |_|_|
 taille (cm) |_|_| |_|_|
 périmètre crânien (cm) |_|_| |_|_| |_|_|

Mesure 4 : date (jj/mm/aa) |_|_| |_|_| |_|_| |_|_| |_|_|
 poids (kg) |_|_| |_|_|
 taille (cm) |_|_| |_|_|
 périmètre crânien (cm) |_|_| |_|_| |_|_|

Auto-questionnaire 3 ans

La croissance de votre enfant depuis l'âge de 2 ans

Merci de reporter les mesures du poids, de la taille et du périmètre crânien de votre enfant qui ont été notées dans son carnet de santé depuis l'âge de 2 ans (si plus de 4 mesures ont été notées, choisir la dernière mesure de chaque trimestre)

Mesure 1 : date (jj/mm/aa) |_|_| |_|_| |_|_| |_|_|
 poids (kg) |_|_| |_|_|
 taille (cm) |_|_| |_|_|
 périmètre crânien (cm) |_|_| |_|_| |_|_|

Mesure 2 : date (jj/mm/aa) |_|_| |_|_| |_|_| |_|_|
 poids (kg) |_|_| |_|_|
 taille (cm) |_|_| |_|_|
 périmètre crânien (cm) |_|_| |_|_| |_|_|

Mesure 3 : date (jj/mm/aa) |_|_| |_|_| |_|_| |_|_|
 poids (kg) |_|_| |_|_|
 taille (cm) |_|_| |_|_|
 périmètre crânien (cm) |_|_| |_|_| |_|_|

Mesure 4 : date (jj/mm/aa) |_|_| |_|_| |_|_| |_|_|
 poids (kg) |_|_| |_|_|
 taille (cm) |_|_| |_|_|
 périmètre crânien (cm) |_|_| |_|_| |_|_|

Figure 4 : Extrait de la partie « croissance de l'enfant » des auto-questionnaires

2.2.2.3 Données collectées lors des examens cliniques

Des mesures du poids et de la taille des enfants ont également été effectuées lors des examens cliniques planifiés dans le cadre de l'étude. Ces examens cliniques ont permis d'obtenir des mesures standardisées des principales caractéristiques anthropométriques d'intérêt.

Lors de l'examen clinique à 1 an, l'enfant était pesé en sous-vêtements dans les bras de sa mère elle-même sans chaussures, et en vêtements légers. La mère était également pesée seule (à 0,1 kg près). Le poids de l'enfant était déterminé en calculant la différence entre ces deux mesures. L'enfant était également pesé seul et cette mesure était utilisée lorsque les mesures décrites précédemment ne permettaient pas de déterminer son poids (mesures manquante(s) ou incohérente(s)).

La taille de l'enfant était mesurée avec une toise pédiatrique. Pour cela, l'enfant était allongé sur son dos, ses jambes maintenues étendues en appuyant d'une main sur ses genoux et sa tête placée au contact de la butée fixe. Il était demandé à la mère de maintenir la tête entre ses deux mains bien au contact. La butée mobile était amenée au contact de ses talons et la taille déterminée à 0,1 cm près.

Lors des examens cliniques des 3 et 5 ans, l'enfant était pesé seul directement, sans chaussures, en sous-vêtements sur la balance électronique.

La taille de l'enfant était mesurée avec une toise murale. Pour cela, l'enfant était placé debout contre le mur, pieds nus ou en chaussettes, les talons joints et collés au mur, les bras le long du corps et le poids bien réparti sur les deux jambes. Après avoir vérifié que ses fesses et omoplates et le crâne touchaient le mur, la toise était descendue au contact de la tête et la taille déterminée à 0,1 cm près. A chaque examen clinique, le poids (excepté pour l'examen clinique à 1 an) et la taille étaient mesurés à deux reprises et la moyenne des deux mesures était calculée.

2.2.2.4 Traitement des données

Les mesures brutes de poids et de taille (collectées à la naissance, à partir des auto-questionnaires et lors des examens cliniques) ont été utilisées en l'état pour les analyses, après qu'elles aient été nettoyées et vérifiées.

Les données de croissance de la naissance à 3 ans ayant déjà été nettoyées et exploitées, mon travail a donc consisté à ajouter les données collectées entre 3 et 5 ans.

Pour cela, les trajectoires de croissance pondérale et staturale de chaque enfant ont été inspectées et les points aberrants (i.e. très éloignés de la trajectoire formée par les autres mesures), non corrigeables aisément (e.g. erreur d'une année), ont été recodés comme valeur manquante.

Lorsque deux mesures de poids ou de taille étaient disponibles pour le même jour de mesure, des règles de décisions ont été adoptées pour ne conserver qu'une mesure.

Ces doublons pouvaient être constitués soit :

- d'une mesure déclarée dans un auto-questionnaire rempli entre 3 et 5 ans et d'une mesure issue d'un examen clinique. Dans ce cas, la mesure effectuée lors de l'examen clinique, considérée comme une mesure standardisée, a été conservée.
- d'une mesure collectée dans les auto-questionnaires remplis de la naissance à 2 ans et d'une mesure collectée dans les auto-questionnaires de 3 à 5 ans. Dans ce cas, la 1^{re} déclaration, considérée plus proche de l'âge de mesure a été conservée.
- de deux mesures collectées dans les auto-questionnaires remplis de 3 à 5 ans.

Dans le cas de mesures de poids, la moyenne des deux mesures a été effectuée si la différence des deux poids n'excédait pas 200 grammes pour une mesure à 3 ans et 1 kg pour une mesure au-delà de 3 ans. Dans le cas inverse, la mesure conservée était déterminée au cas par cas en fonction de la trajectoire de croissance pondérale de l'individu. Dans le cas des mesures de taille, la moyenne des deux mesures était retenue lorsque la différence des deux mesures n'excédait pas 2 cm. Dans le cas contraire, la mesure conservée était déterminée au cas par cas en fonction de la trajectoire de croissance staturale de l'individu.

L'IMC aux différents âges de mesure a ensuite pu être calculé pour chaque enfant lorsque les mesures de poids et de taille étaient toutes deux disponibles.

2.2.3 Tabagisme passif

Les auto-questionnaires remis aux mères durant l'examen planifié entre la 24^e et la 28^e SA adressés aux pères permettaient, de recueillir le statut tabagique paternel durant la grossesse. L'auto-questionnaire envoyé à la mère aux 4 mois de l'enfant permettait de recueillir le statut tabagique maternel et paternel postnatal. Dans cet auto-questionnaire, les mères devaient rapporter leur statut tabagique ainsi que le statut tabagique de leur partenaire au moment du remplissage du questionnaire.

2.2.4 Autres caractéristiques parentales

Pendant l'examen planifié entre la 24^e et la 28^e SA, les mères étaient mesurées à 0,2 cm près par les sages-femmes enquêtrices et déclaraient leur poids pré-conceptionnel.

Les femmes devaient également renseigner le plus haut diplôme qu'elles avaient obtenu à ce jour.

Dans l'auto-questionnaire remis à leur attention aux mères assistant à la visite planifiée entre la 24^e et la 28^e SA, les pères devaient renseigner leur taille et leur poids.

2.2.5 Alimentation de l'enfant

2.2.5.1 Allaitement maternel

Le mode d'alimentation de l'enfant pendant son séjour à la maternité a été recueilli à partir du dossier obstétrical. Les quatre types d'alimentation étaient : lait maternel, mixte, artificiel ou diététique spéciale.

Dans l'auto-questionnaire à 4 mois, les parents devaient renseigner des informations sur l'alimentation liquide et solide de l'enfant à la naissance, durant la 1^{re} semaine, entre la 2^e semaine et la fin du 1^{er} mois et durant les 2^e, 3^e et 4^e mois. Les parents renseignaient notamment des informations sur le type d'alimentation lactée : la fréquence de consommation de lait maternel, le type, la fréquence et la quantité de consommation de préparations infantiles et de lait de vache, ainsi que la date de fin d'allaitement et/ou la durée d'utilisation du lait concerné.

Dans les auto-questionnaires à 4, 8, 12 et 24 mois, les mères renseignaient si elles poursuivaient l'allaitement maternel et si ce n'était pas le cas, renseignaient la date d'arrêt.

2.2.5.2 Alimentation dans la petite enfance

Dans les questionnaires envoyés à 4, 8 et 12 mois, des informations étaient également collectées sur les aliments introduits dans l'alimentation (e.g. viandes, poissons, légumes). A 12 mois, un questionnaire additionnel a permis de collecter chez 1296 familles des informations sur la fréquence d'utilisation, pour l'enfant, d'aliments spécifiques « bébé » vendus dans le commerce, d'aliments ordinaires du commerce ou d'aliments faits maison.

2.3 Caractéristiques de la population d'étude

Sur les 2002 femmes enceintes, 968 (48%) ont été recrutées à Poitiers et 1034 (52%) à Nancy. Sur les 1907 enfants dont la date de naissance était connue, le poids de naissance était connu pour 1899 enfants (3 femmes ont accouché à l'extérieur des centres de recrutement, 1 femme a souhaité sortir de l'étude et il y a eu 4 morts fœtales *in utero*).

Les femmes de l'étude EDEN avaient un âge moyen à l'inclusion de 29,5 ans et avaient un niveau d'études plus élevé que dans la population des femmes françaises en âge de procréer. En effet, 53,2% d'entre elles avaient un niveau supérieur au baccalauréat, ce qui était le cas de 42,6 % des femmes enceintes de l'enquête nationale périnatale (ENP) de 2010 (Blondel and Kermarrec, 2011).

Le suivi des sujets est résumé dans la figure 5.

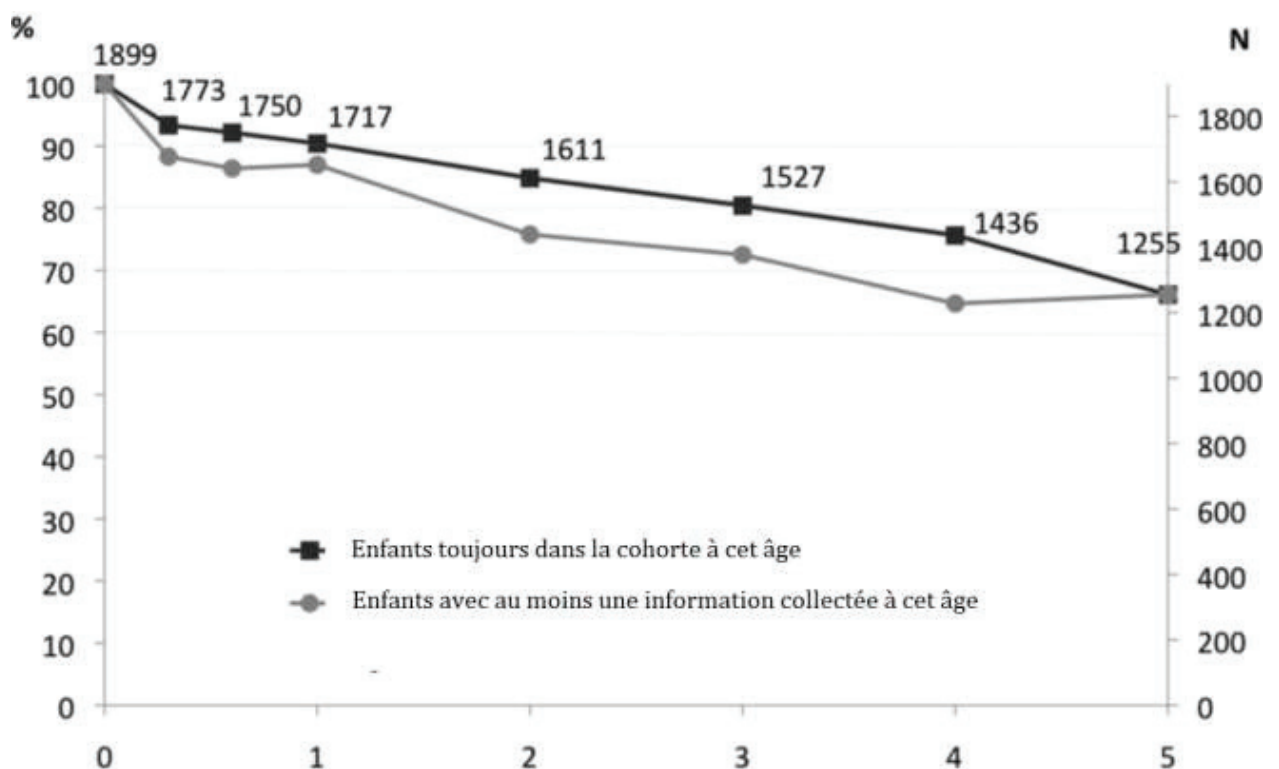


Figure 5 : Attrition sur 5 ans

Chapitre 3 : Tabagisme maternel pendant la grossesse et croissance de l'enfant

3.1 Contexte

Depuis plusieurs décennies, le tabagisme de la mère pendant la grossesse est un facteur de risque établi de restriction de la croissance foétale, de naissance prématurée, et de petit poids de naissance (Andres and Day, 2000; Castles et al., 1999; Cnattingius, 2004).

Ces dernières années, de nombreuses études observationnelles, se sont intéressées aux associations entre le tabagisme maternel et des indicateurs de croissance et d'adiposité que ce soit dans l'enfance ou plus tard dans la vie. Dans une revue de la littérature sur les facteurs précoces associés au surpoids (Brisbois et al., 2012a; Ino, 2010; Oken et al., 2008), le tabagisme maternel pendant la grossesse apparaît comme un facteur associé de manière positive au risque de surpoids à l'âge adulte. Ces études sont en faveur d'un effet à long-terme du tabagisme maternel sur la santé et de plus en plus d'études suggèrent que ces associations émergeraient dès l'enfance (Ino, 2010; Oken et al., 2008; Weng et al., 2012).

L'IMC accru observé chez les individus ayant été exposés *in utero* au tabagisme maternel a été principalement expliqué, dans la littérature, par un déficit de poids à la naissance suivi d'un rattrapage du poids rapide avant l'âge de 1 an pour atteindre ou même dépasser le poids des enfants non-exposés. Alors que les enfants exposés ont tendance à rattraper leur déficit de poids, il semblerait que ces enfants nés plus petits en taille demeurent de plus petite taille dans l'enfance (Chen et al., 2006; Durmus et al., 2011; Howe et al., 2012; Timmermans et al., 2013).

L'existence de confusion résiduelle liée à des modes de vie familiaux délétères partagés (sédentarité, alimentation non équilibrée et statut socio-économique plus faible) ne peut être exclue et a été suggérée par plusieurs études comparant les associations entre tabagisme maternel et paternel (Florath et al., 2014; Laura D Howe et al., 2012; Kwok et al., 2010; Leary et al., 2006; vonKries et al., 2008) et par une étude intra-familiale (Iliadou et al., 2010).

Des incertitudes subsistent également sur l'existence de périodes critiques concernant l'effet du tabagisme pendant la grossesse sur la croissance postnatale de l'enfant. Entre l'âge de 5 et 7 ans, un risque de surpoids plus important a été mis en évidence chez les enfants exposés au 1^{er} trimestre de grossesse (Odds Ratio (OR): 2,65, IC95% [1,26 ; 5,54]) comparé au risque chez les enfants exposés plus tardivement (OR:

1,88, IC95% [0,85 ; 4,15]) suggérant une susceptibilité accrue en début de grossesse (Mendez et al., 2008). Cette hypothèse n'est pas corroborée par une autre étude ayant mis en évidence des risques de surpoids à 5-6 ans comparables chez les enfants exposés durant le 1^{er} trimestre uniquement ou tout au long de la grossesse (Toschke et al., 2003). Des caractéristiques anthropométriques (périmètre crânien, taille, poids et IMC) comparables chez les enfants exposés durant le 1^{er} trimestre uniquement et chez les enfants non exposés ont également été rapportées (Durmus et al., 2011).

De plus, la majorité des études étant transversales, des incertitudes subsistent sur la cinétique des associations entre le tabagisme maternel pendant la grossesse et l'IMC, mais également avec le poids et la taille séparément.

Le faible nombre d'études longitudinales et les limites méthodologiques ne permettent pas de caractériser précisément le profil de croissance des enfants de femmes fumeuses et d'identifier à quel moment les différences d'IMC entre les enfants exposés et les non-exposés émergent.

3.2 Objectifs spécifiques

L'objectif principal de cette partie était de décrire de quelle manière, à partir d'une méthode basée sur la modélisation des courbes de croissance individuelles du poids et de la taille (Botton et al., 2014a), nous pouvions étudier, le plus finement possible, les déterminants de la croissance du poids, de la taille et de l'IMC de la naissance à 5 ans.

Du fait des incertitudes présentées dans le contexte de ce chapitre et car il s'agit de l'un des facteurs précoces probablement impliqué dans la programmation prénatale de l'obésité, le tabagisme maternel pendant la grossesse apparaissait être un déterminant de choix pour la conduite de ce travail. C'est pourquoi, bien que la méthode puisse être utilisée pour tout autre déterminant, nous avons choisi de l'illustrer avec celui-ci.

Notre objectif était donc d'étudier de quelle manière cette méthode pouvait être utilisée pour améliorer les connaissances sur l'existence de périodes critiques d'exposition pendant la grossesse et sur la cinétique d'émergence de potentielles différences d'IMC dans l'enfance.

3.3 Matériel et Méthodes

3.3.1 Génération de variables

Variable d'exposition principale

Nous avons distingué les femmes ne fumant à aucun trimestre de leur grossesse, les femmes fumant au 1^{er} trimestre exclusivement et les femmes fumant lors du 2^e et/ou 3^e trimestre(s). Ces dernières déclaraient très majoritairement fumer également au 1^{er} trimestre (93,5%), nous caractériserons donc cette population, pour plus de simplicité, de « femmes fumant tout au long de la grossesse ». Pour cela, le statut tabagique de la mère en début de grossesse a été considéré comme un proxy du statut tabagique au 1^{er} trimestre ; le statut tabagique de la mère lors de la visite ayant eu lieu entre la 24^e et la 28^e SA a permis d'évaluer le statut tabagique lors du 2^e trimestre de grossesse et le statut tabagique en fin de grossesse a été considéré comme représentatif du statut tabagique au 3^e trimestre de grossesse.

Facteurs de confusion potentiels considérés dans les analyses principales

1. Le centre.

En ne prenant pas en compte le centre dans nos analyses, nous considérerions que les populations des deux centres, mais également les conditions de suivi dans les centres, sont homogènes. La formation des investigateurs et sages-femmes impliqués dans l'étude devrait permettre de standardiser au maximum les pratiques liées à la conduite de l'étude dans les deux centres. Cependant, nous ne pouvons pas exclure des différences de caractéristiques des femmes et des différences inhérentes aux structures hospitalières assurant leur suivi, pouvant entraîner des différences à la fois sur les variables d'exposition et les événements d'intérêt et susceptible d'entraîner un biais dans l'estimation des associations.

Une variable « centre », distinguant les femmes recrutées à Nancy ou Poitiers, a donc été considérée dans les analyses.

2. La durée d'allaitement maternel.

Certaines études épidémiologiques suggèrent que la croissance staturo-pondérale des enfants recevant un allaitement prolongé et exclusif est plus lente durant la 1^{re} année de vie (particulièrement après le 3^e mois) que celle des enfants nourris avec

des préparations infantiles ou recevant un allaitement moins exclusif ou de plus courte durée (Dewey, 1998; Dewey et al., 1995; Hediger et al., 2000; Kramer et al., 2004; Nielsen et al., 1998; Ong et al., 2009, 2002; Regnault et al., 2010; World Health Organization Working Group on Infant Growth, 1995).

Le tabagisme maternel a également été associé aux pratiques d'allaitement : Dennis et coll. montrent, par exemple, que les femmes fumeuses sont moins susceptibles d'initier un allaitement au sein et de continuer à allaiter sur la durée, comparées aux femmes ne fumant pas (Dennis, 2002). Des facteurs psychosociaux, mais également biologiques, semblent responsables de ces associations. Le tabagisme pourrait réduire la sécrétion de prolactine et par conséquent réduire le volume de lait maternel (Bahadori et al., 2013).

Afin de considérer un biais potentiel lié à l'allaitement dans les associations entre le tabagisme maternel pendant la grossesse et la croissance de l'enfant, l'allaitement a été pris en considération. Pour cela, la durée d'allaitement maternel a été calculée et deux groupes ont été distingués selon que l'allaitement maternel excédait 3 mois ou non.

3. L'IMC maternel pré-conceptionnel.

L'IMC maternel pré-conceptionnel est quant à lui associé à la croissance de l'enfant ainsi qu'à l'initiation et à la durée de l'allaitement. Botton et coll. ont notamment mis en évidence une association positive entre l'IMC maternel et la vitesse de croissance pondérale après 3 ans (Botton et al., 2010). L'obésité maternelle semble associée à une fréquence d'initiation et à une durée d'allaitement plus faibles (Turcksin et al., 2014). Il a également été mis en évidence que les fumeurs ont un IMC plus faible que les personnes n'ayant jamais fumé (Plurphanswat and Rodu, 2014).

Le poids pré-conceptionnel déclaré et la taille mesurée ont permis de calculer l'IMC pré-conceptionnel des mères. La classification de l'Organisation Mondiale de la Santé (OMS) a été utilisée pour classer les mères avant grossesse selon leur IMC en 2 catégories : maigre ou de corpulence normale ($IMC < 25$) et en surpoids ou obèse ($25 \geq IMC$).

4. Statut socio-économique :

Le niveau d'études de la mère a été considéré comme un proxy du niveau socio-économique familial. Le niveau socio-économique a été pris en considération, car il a été

montré associé au profil de croissance (Cavelaars et al., 2000; Howe et al., 2011) dans l'enfance et est fortement associé au tabagisme maternel pendant la grossesse (Cnattingius, 2004) : la prévalence du tabagisme est en effet plus élevée chez les femmes ayant un niveau d'études plus faible dans les pays à revenus élevés. Nous avons distingué les femmes ayant un niveau « lycée » ou un plus bas niveau d'études (aucun diplôme, certificat d'études primaires, certificat d'aptitude professionnel (CAP), brevet d'enseignement professionnel (BEP), bac professionnel, technologique ou général); les femmes ayant un diplôme universitaire correspondant à un bac+2 (DUT, BTS, DEUG) ; les femmes ayant un diplôme d'enseignement supérieur (2^e ou 3^e cycle ou grande école).

L'âge de la mère est également un facteur très associé à des différences socio-économiques et a donc été considéré comme un facteur de confusion potentiel et considéré comme une variable continue.

5. Sexe de l'enfant :

S'il n'est probablement pas facteur de confusion en tant que tel, car non associé au tabagisme maternel, le sexe de l'enfant a néanmoins été considéré compte tenu des différences de croissance importantes entre les filles et les garçons. Il est bien établi que la croissance staturo-pondérale diffère chez les filles et chez les garçons (Ayatollahi, 2005). Les garçons auraient un poids de naissance ainsi que des vitesses de croissance du poids à 1 et 3 mois plus élevés que les filles (Regnault et al., 2010). Hosseini et coll. montrent que la différence de poids entre les garçons et les filles se maintient et se renforce même de la naissance à 2 ans (Hosseini et al., 2014). En ce qui concerne la croissance staturale, les garçons auraient une taille à la naissance ainsi que des vitesses de croissance à 1 et 3 mois plus élevées que les filles (Regnault et al., 2010). Ces différences diminueraient après 8 mois et s'inverseraient entre 1 et 4 ans, période pendant laquelle, la vitesse de croissance des filles serait légèrement supérieure à celle des garçons. Les vitesses de croissance des filles et des garçons convergeraient ensuite jusqu'à 9 ans environ (Veldhuis et al., 2005).

3.3.2 Modélisation de la croissance

Nous introduisons dans un premier temps le principe général utilisé pour la modélisation de la croissance.

3.3.2.1 Mesures anthropométriques (poids et taille) utilisées pour la modélisation des croissances staturale et pondérale

Pour la modélisation de la croissance staturale, pour chaque individu, l'ensemble des mesures de taille y compris la taille à la naissance a été considéré. Pour la modélisation de la croissance pondérale, aux mesures de poids collectées à partir du 5^e jour de vie, le poids minimal dans les 4 premiers jours de vie a été inclus et non le poids à la naissance. En effet, comme nous l'avons évoqué dans le chapitre 1, le phénomène de perte de poids néonatal ne peut pas être modélisé par le modèle de Jentsch-Bayley celui-ci ne permettant pas la modélisation de trajectoires non monotones.

Botton et coll. ont précédemment montré les répercussions de l'inclusion du poids à la naissance sur la modélisation de la croissance pondérale (figure 6, Botton et al., 2014a).

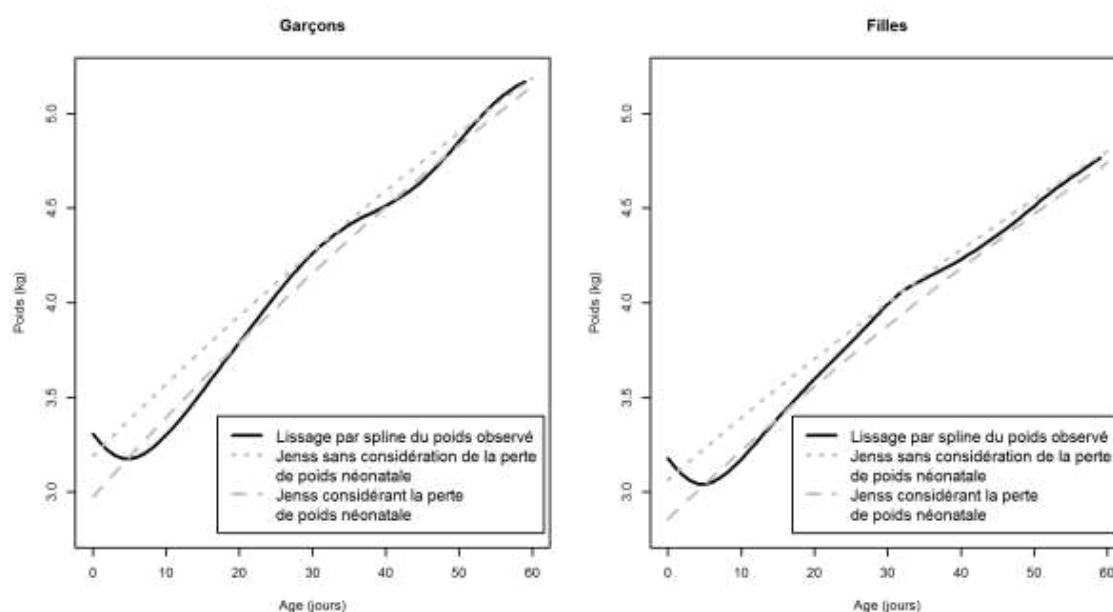


Figure 6 : Effet de l'inclusion du poids de naissance (ne considérant pas la perte de poids néonatale) sur une trajectoire de poids obtenu avec le modèle de Jentsch-Bayley

L'inclusion du poids à la naissance conduisait à une sur-estimation de l'estimation du poids pendant les premières semaines et à une sous-estimation de la vitesse de croissance dans la petite enfance.

3.3.2.2 Modèles utilisés pour la modélisation des croissances staturale et pondérale

Pour la modélisation de la croissance staturale et pondérale, nous avons utilisé une reparamétrisation du modèle de Jenss-Bayley précédemment décrite par Botton et coll. (figure 7).

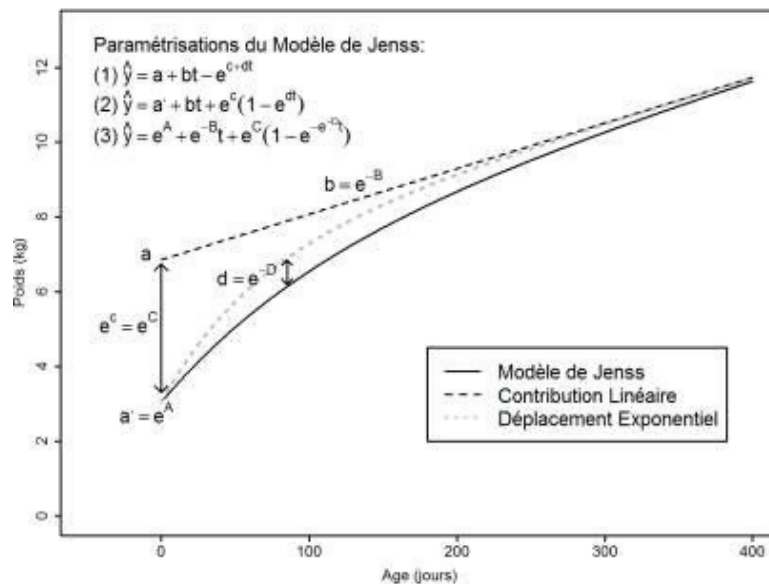


Figure 7 : Illustration de la signification des paramètres selon la paramétrisation du modèle de Jenss-Bayley

L'équation initialement décrite par Jenss-Bayley est la suivante :

$$y = a + bt - e^{-c+dt} \text{ (Equation 1 de la Figure 7)}$$

L'équation a été modifiée sous la forme

$$y = a' + bt + e^c(1 - e^{-dt}) \text{ (Equation 2 de la Figure 7)}$$

Cette équation permet d'exprimer la taille et le poids à $t_0 = 0$ jour avec le paramètre a' .

Afin d'améliorer la convergence des modèles, certaines contraintes ont été appliquées selon l'équation suivante :

$$y = e^A + e^{-B}t + e^c(1 - e^{-e^{-D}t}) \text{ (Equation 3 de la Figure 7)}$$

Une fonction exponentielle a été appliquée à $a'(e^A)$, afin de contraindre la taille à la naissance et le poids dans les premiers jours de vie à être positifs. Cela a également été appliqué à la pente de la croissance sur sa partie linéaire ($b = e^{-B}$). Comme la vitesse de croissance après 1 an est supérieure à la vitesse de croissance dans les premiers mois de vie, a' doit être supérieur à e^A et par conséquent e^c doit être positif (c'est pourquoi un signe positif a été ajouté). Enfin, le taux de décélération d doit être négatif car la vitesse de croissance diminue dans les premiers mois de vie ($-e^{-D}$).

Par facilité d'interprétation des coefficients (sens de variation) nous avons, dans ce travail, utilisé l'équation suivante pour la modélisation de la taille (exprimée en centimètres) ou du poids (exprimé en kilogrammes) d'un enfant i au temps t_{ij} :

$$Poids_{ij} = \exp(A_{P_i}) + \exp(B_{P_i}) t_{ij} + \exp(C_{P_i}) (1 - \exp(-\exp(D_{P_i}) t_{ij})) + e_{P_{ij}}$$

$$Taille_{ij} = \exp(A_{T_i}) + \exp(B_{T_i}) t_{ij} + \exp(C_{T_i}) (1 - \exp(-\exp(D_{T_i}) t_{ij})) + e_{T_{ij}}$$

Où $A_{P_i}, B_{P_i}, C_{P_i}, D_{P_i}$ représentent les paramètres individuels de la croissance du poids et $A_{T_i}, B_{T_i}, C_{T_i}, D_{T_i}$ ceux de la taille et $e_{P_{ij}}$ et $e_{T_{ij}}$ représentent les erreurs résiduelles, respectivement pour le poids et la taille.

Dans ces équations, le paramètre A représente la taille ou le poids à t_0 . Ce paramètre représente donc la taille à la naissance (A_T) ou une approximation du poids minimal dans les 4 premiers jours de vie (A_P). Le paramètre B représente la vitesse de croissance dans la partie linéaire de la trajectoire de croissance, soit après deux ans. Le paramètre C représente la poussée de croissance de t_0 à 2 ans (la différence entre la taille ou le poids à t_0 et l'ordonnée à l'origine de l'asymptote à la trajectoire dans sa partie linéaire). Par facilité, même si t_0 ne représente pas la naissance pour la modélisation du poids, le paramètre C sera dénommé « gain pondéral de la naissance à 2 ans » dans les modèles décrivant la croissance du poids et « gain statural de la naissance à 2 ans » dans les modèles décrivant la croissance de la taille. Le paramètre D représente la courbure de la trajectoire dans les premiers mois et peut être vue comme une combinaison de la vitesse de croissance en début de vie et du taux de décélération de la croissance avant que la croissance ne rejoigne sa partie linéaire. La signification des paramètres est similaire à celle illustrée dans la Figure 7 et l'impact d'une modification des valeurs de ces paramètres sur la forme globale de la courbe de croissance est illustré dans la Figure 8. Les trajectoires illustrées ici sont des trajectoires de poids obtenues pour chaque décile de A_P, B_P, C_P, D_P en attribuant aux autres paramètres la valeur du

paramètre moyen dans la population (Tableau 1). Ceci suppose une indépendance des paramètres ce qui n'est pas le cas en réalité. La covariance non nulle qui existe entre les différents paramètres fait que les formes de trajectoires de croissance ne sont pas représentatives de celles observées dans la population mais cela permet de comprendre la signification des différents paramètres.

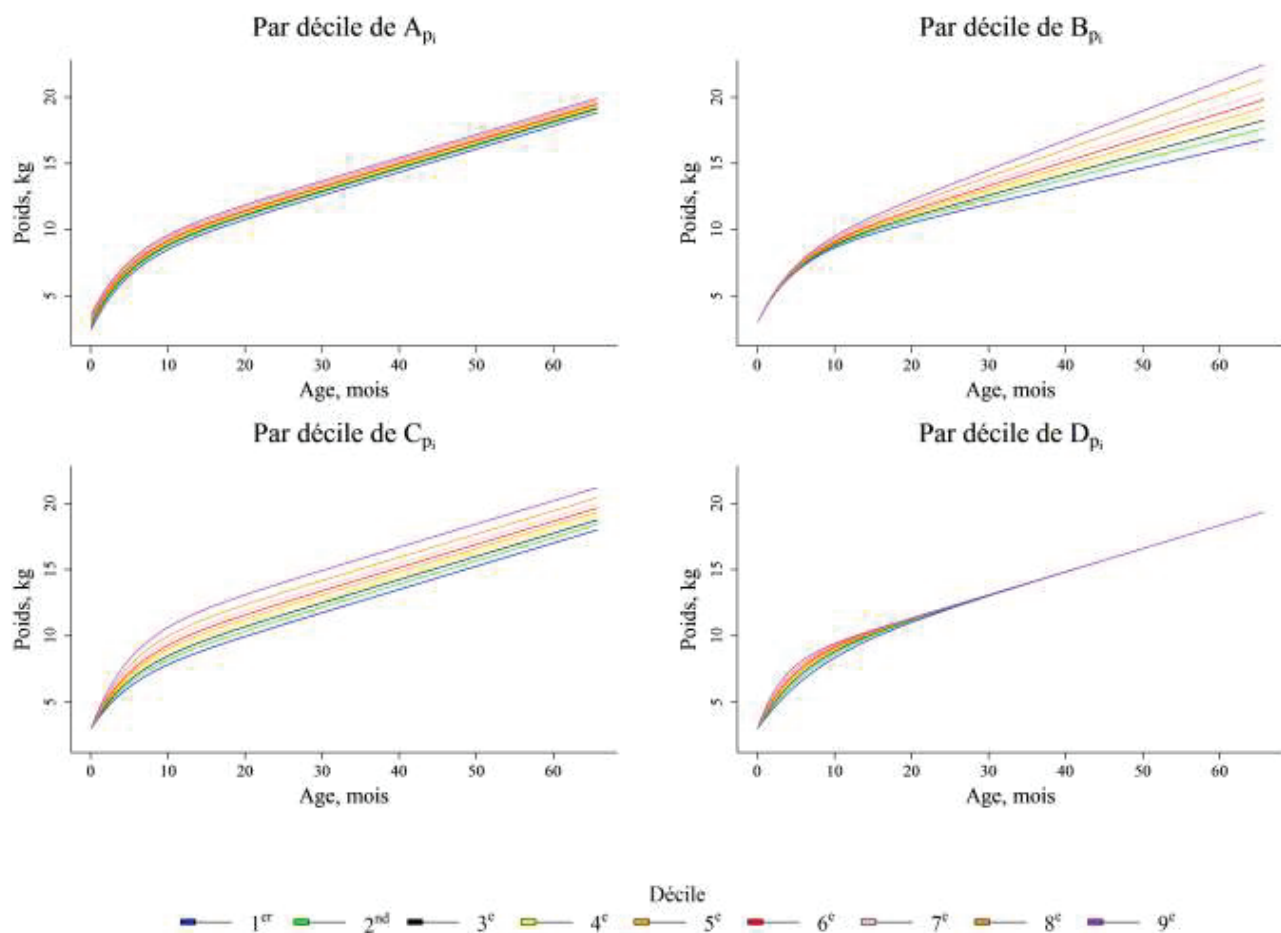


Figure 8 : Trajectoires de croissance de poids de la naissance à 5 ans par déciles de paramètres de poids obtenus sur un sous-échantillon de la population EDEN

Tableau 1 : Distributions des paramètres individuels de croissance obtenus sur un sous-échantillon de la population EDEN.

Paramètres du Poids	Déciles									Espérance
	1 ^{er}	2 ^e	3 ^e	4 ^e	5 ^e	6 ^e	7 ^e	8 ^e	9 ^e	
A_{P_i} Extrapolation du poids de naissance	0,89	0,98	1,02	1,06	1,10	1,13	1,17	1,21	1,26	1,08
B_{P_i} Vitesse de croissance -Enfance-	-5,41	-5,32	-5,26	-5,21	-5,17	-5,12	-5,07	-5,00	-4,92	-5,16
C_{P_i} Poussée de croissance -0-2 ans-	1,27	1,38	1,46	1,53	1,59	1,65	1,71	1,79	1,91	1,59
D_{P_i} Degré de courbure -Premiers mois-	-5,41	-5,25	-5,12	-5,04	-4,96	-4,87	-4,80	-4,70	-4,54	-4,97

3.3.2.3 Logiciel et méthode d'estimation

Pour ce travail, les analyses ont été effectuées avec le logiciel R en faisant appel au package SAEMIX. Ce package a été développé pour les modèles non linéaires à effets mixtes et a été récemment utilisé pour étudier les associations entre l'exposition périnatale au chlordécone et la croissance des enfants (Costet et al., 2015). Il maximise la vraisemblance par un algorithme EM stochastique (SAEM). Cet algorithme calcule l'estimateur du maximum de vraisemblance des paramètres de la population sans approximation du modèle (i.e. sans linéarisation) (Comets et al., 2012).

3.3.2.4 Etude des associations entre le tabagisme maternel et la croissance

3.3.2.4.1 Sélection de la population d'étude

La population d'étude se composait de l'ensemble des enfants ayant au moins deux mesures de poids et deux mesures de taille et pour lesquels le statut tabagique maternel pendant la grossesse (variable d'exposition principale), ainsi que le centre d'inclusion, la durée d'allaitement maternel, l'IMC maternel pré-conceptionnel, le niveau

d'études, l'âge de la mère et le sexe de l'enfant étaient renseignés (cf. diagramme de flux Figure 13 page 86).

3.3.2.4.2 Stratégie d'analyse

Deux méthodes ont été utilisées pour étudier les associations entre l'exposition au tabagisme maternel pendant la vie fœtale et la croissance postnatale du poids, de la taille et de l'IMC : une méthode en « 1 étape » et une méthode en « 2 étapes ». La méthode en « 1 étape » permettait d'étudier si les trajectoires de croissance globales différaient selon le statut tabagique de la mère pendant la grossesse. La méthode en « 2 étapes » permettait d'étudier les associations entre le tabagisme maternel pendant la grossesse et les paramètres anthropométriques à la naissance ; les poids, tailles et IMC ainsi que les vitesses de croissance instantanées correspondantes prédits pour des âges spécifiques, ainsi qu'avec l'âge et l'IMC au pic d'IMC ayant lieu généralement à la fin de la 1^{re} année.

3.3.2.4.2.1 Méthode en « 1 étape »

Dans cette méthode, les variables indicatrices du tabagisme maternel ainsi que les facteurs de confusion potentiels ont directement été introduits dans le modèle. En considérant l'exemple de la modélisation de la taille, l'équation générale précédemment décrite pour la modélisation de la taille :

$$Taille_{ij} = \exp(A_{T_i}) + \exp(B_{T_i}) t_{ij} + \exp(C_{T_i}) (1 - \exp(-\exp(D_{T_i}) t_{ij})) + e_{T_{ij}}$$

se décompose selon l'équation décrite ci-dessous:

$$Taille_{ij} = \exp(\mu_{A_T} + \beta_{A_T} * Z + u_{A_{T_i}}) + \exp(\mu_{B_T} + \beta_{B_T} * Z + u_{B_{T_i}}) t_{ij} \\ + \exp(\mu_{C_T} + \beta_{C_T} * Z + u_{C_{T_i}}) (1 - \exp(-\exp(\mu_{D_T} + \beta_{D_T} * Z + u_{D_{T_i}}) t_{ij})) + e_{T_{ij}}$$

Dans cette équation, $Taille_{ij}$ décrit la taille (en cm) de l'individu i au temps t_{ij} (en jours) ; $\beta_{A_T}, \beta_{B_T}, \beta_{C_T}, \beta_{D_T}$ représentent les effets fixes associés aux covariables du vecteur Z (incluant les variables indicatrices du tabagisme maternel pendant la grossesse et les facteurs de confusion potentiels); $u_{A_T}, u_{B_T}, u_{C_T}$ et u_{D_T} sont les effets aléatoires individuels spécifiques de chaque sujet pour A_T, B_T, C_T , et D_T respectivement; $\mu_{A_T}, \mu_{B_T}, \mu_{C_T}$ et μ_{D_T} représentent donc les effets fixes des paramètres A_T, B_T, C_T , et D_T respectivement pour les enfants dont les covariables du vecteur Z sont nulles ; $e_{T_{ij}}$ est l'erreur résiduelle.

Le poids a été modélisé de la même façon selon l'équation:

$$Poids_{ij} = \exp(\mu_{A_P} + \beta_{A_P} * Z + u_{A_{P_i}}) + \exp(\mu_{B_P} + \beta_{B_P} * Z + u_{B_{P_i}}) t_{ij} \\ + \exp(\mu_{C_P} + \beta_{C_P} * Z + u_{C_{P_i}}) (1 - \exp(-\exp(\mu_{D_P} + \beta_{D_P} * Z + u_{D_{P_i}}) t_{ij})) + e_{P_{ij}}$$

Les modèles de croissance du poids et de la taille peuvent également s'écrire sous la forme générale suivante :

$$y_{ijk} = f(x_{ijk}, \phi_{ik}) + g(x_{ijk}, \phi_{ik}, \xi_k) \varepsilon_{ijk}, 1 \leq i \leq N, 1 \leq j \leq n_i$$

avec $k = (P, T)$ définissant le poids et la taille respectivement, y_{ijk} représente la j^e mesure du sujet i pour la variable k au temps t_{ijk} ; N représente le nombre de sujets et n_i le nombre de mesures du sujet i ; x_{ijk} représente les variables de régression ou variables de design connues (incluant le temps de mesure), ϕ_{ik} représente les paramètres individuels; ε_{ijk} représente les erreurs résiduelles; g est une fonction définissant le modèle d'erreur.

Afin de prendre en compte une possible hétéroscédasticité de la variance résiduelle, une étape préliminaire a consisté à comparer l'AIC et le BIC obtenus pour des modèles de croissances pondérale et staturale ajustés séparément pour les filles et les garçons et n'incluant pas de covariables, supposant des modèles d'erreur différents.

Parmi les modèles d'erreur testés, l'un supposait l'existence d'une variance résiduelle constante ou homoscedastique, ne dépendant pas des valeurs de poids et de taille prédites par les modèles telle que $g = a$ et $\xi_k = a_k$; un autre, dit proportionnel, supposait $g = bf$ et $\xi_k = b_k$ ce qui suppose que l'erreur est proportionnelle à la valeur du poids ou de la taille prédite par les modèles; le dernier dit combiné supposait la forme générale $g = a + bf$ et $\xi_k = (a_k, b_k)$ ce qui suppose que l'erreur augmente de manière linéaire avec la valeur du poids ou de la taille prédite par le modèle.

Les résultats de ces différents modèles en terme d'AIC et de BIC sont présentés en Annexe 1. Parmi les modèles d'erreurs candidats, le modèle d'erreur combiné était celui permettant de minimiser l'AIC/BIC et a, de ce fait, été retenu.

Ce modèle d'erreur, en autorisant la variance résiduelle à augmenter lorsque le poids ou la taille prédite augmente, permet de prendre en compte l'augmentation de la variance résiduelle (hétéroscédasticité) observée au cours du temps (la valeur de la variable modélisée étant très fortement corrélée au temps de mesure).

Les associations longitudinales entre le tabagisme maternel pendant la grossesse et les trajectoires de croissance de la naissance à 5 ans dans leur globalité

respectivement pour la taille et le poids ont été étudiées en comparant les modèles incluant et n'incluant pas les variables de tabagisme à l'aide de tests du rapport de vraisemblance.

L'IMC (en kg/m²) de l'individu i au temps t_{ij} (en jours) a été prédit selon l'équation suivante (poids en kg, taille en cm) :

$$\widehat{IMC}_{ij} = \frac{\widehat{Poids}_{ij}}{(\widehat{Taille}_{ij}/100)^2}$$

Les trajectoires moyennes de poids et de taille ont été calculées et représentées graphiquement par catégorie de tabagisme maternel pour une population choisie arbitrairement (garçons, allaités moins de 3 mois, nés à Poitiers, dont la mère avait un IMC normal, un niveau d'éducation équivalent au lycée ou inférieur et un âge de 30 ans à l'accouchement).

Compte tenu du caractère non-linéaire du modèle, pour un paramètre donné (poids, taille ou IMC), la moyenne des prédictions individuelles à un temps donné n'est pas égale à la prédiction calculée à cet âge en appliquant les coefficients moyens du modèle. Pour calculer les différences entre les trajectoires moyennes de poids, taille et IMC entre les enfants des 3 groupes de tabagisme, les prédictions individuelles de poids, taille et IMC ont été calculées pour chaque individu en fonction de leur groupe de tabagisme et en fixant les valeurs des facteurs de confusion: la valeur du paramètre \hat{A}_{P_i} d'un enfant de femme fumeuse au 1^{er} trimestre exclusivement était donc égale à la somme de l'ordonnée à l'origine du paramètre A_P ($\hat{\mu}_{A_P}$)+30*le coefficient de régression associé à l'âge de la mère ($\hat{\beta}_{age_{A_P}}$)+le coefficient de régression associé au tabagisme au 1^{er} trimestre($\hat{\beta}_{tab1T_{A_P}}$) +l'effet aléatoire individuel sur le paramètre A_P ($\hat{u}_{A_{P_i}}$). Les autres variables d'ajustement n'interviennent pas dans ce calcul car il a été choisi de les fixer à leur valeur de référence.

$$\hat{A}_{P_i} = \hat{\mu}_{A_P} + \hat{\beta}_{age_{A_P}} * 30 + \hat{\beta}_{tab1T_{A_P}} + \hat{u}_{A_{P_i}}$$

La même approche a été utilisée pour l'ensemble des coefficients. Nous pouvons ainsi calculer la prédiction individuelle du paramètre anthropométrique considéré (i.e. du poids ou de la taille) au temps t en ajustant sur les facteurs de confusion puis calculer la moyenne des prédictions individuelles obtenues à un âge donné pour chaque sous-groupe de tabagisme et enfin la différence des moyennes à cet âge entre les enfants

exposés au 1^{er} trimestre uniquement ou exposés tout au long de la grossesse et la catégorie de référence (i.e. les enfants de femmes non fumeuses).

3.3.2.4.2.2 Méthode en « 2 étapes »

3.3.2.4.2.2.1 Etape 1 : Détermination des caractéristiques individuelles de croissance

Détermination des trajectoires individuelles de croissance staturo-pondérale

Dans une 1^{re} étape, les croissances du poids et de la taille ont été modélisées selon le modèle introduit en page 67, sans covariables, séparément chez les filles et les garçons. Ainsi, les équations individuelles caractérisant la croissance pondérale et celles caractérisant la croissance staturale de chaque enfant ont pu être obtenues.

Détermination des trajectoires individuelles de croissance de l'IMC

A partir des équations des trajectoires individuelles de poids et de taille, les équations des trajectoires de croissance de l'IMC ont pu être calculées. L'IMC d'un enfant i au temps t_{ij} est obtenu en faisant le rapport de son poids (exprimé en kilogrammes) à t_{ij} sur le carré de sa taille à ce même temps (exprimée en mètres²).

$$\widehat{IMC}_{ij} = \frac{\widehat{Poids}_{ij}}{(\widehat{Taille}_{ij}/100)^2}$$

$$= \frac{\exp(\widehat{A}_{P_i}) + \exp(\widehat{B}_{P_i}) t_{ij} + \exp(\widehat{C}_{P_i}) (1 - \exp(-\exp(\widehat{D}_{P_i}) t_{ij}))}{[(\exp(\widehat{A}_{T_i}) + \exp(\widehat{B}_{T_i}) t_{ij} + \exp(\widehat{C}_{T_i}) (1 - \exp(-\exp(\widehat{D}_{T_i}) t_{ij}))) / 100]^2}$$

L'effet de la variation des paramètres du modèle poids (Figure 9) ou du modèle taille (Figure 10) sur les trajectoires de croissance de l'IMC sont illustrées ci-dessous. Nous avons procédé de la même manière que pour l'illustration des trajectoires de croissance pondérale, les différences de croissance induites pour chaque paramètre du poids s'expriment donc en considérant les autres paramètres du poids et de la taille constants. Ces figures montrent logiquement que, lorsque le paramètre A du poids est plus élevé (à paramètre A de la taille égal) ou que le paramètre A de la taille est plus faible (à paramètre A du poids égal), l'IMC en début de vie est plus élevé. Lorsque le paramètre C du poids est plus élevé (ou le paramètre C de la taille plus faible), la vitesse de croissance de l'IMC dans les 1^{ers} mois est plus élevée. Les mêmes conclusions peuvent

être faites sur l'importance de la courbure de la trajectoire dans les 1^{ers} mois : elle sera plus importante lorsque le paramètre D du poids augmente et/ou lorsque celui de la taille diminue. Enfin, la vitesse de décroissance de l'IMC après le pic sera plus faible lorsque le paramètre B du poids augmente et lorsque celui de la taille diminue, conduisant à des trajectoires d'IMC plus élevées.

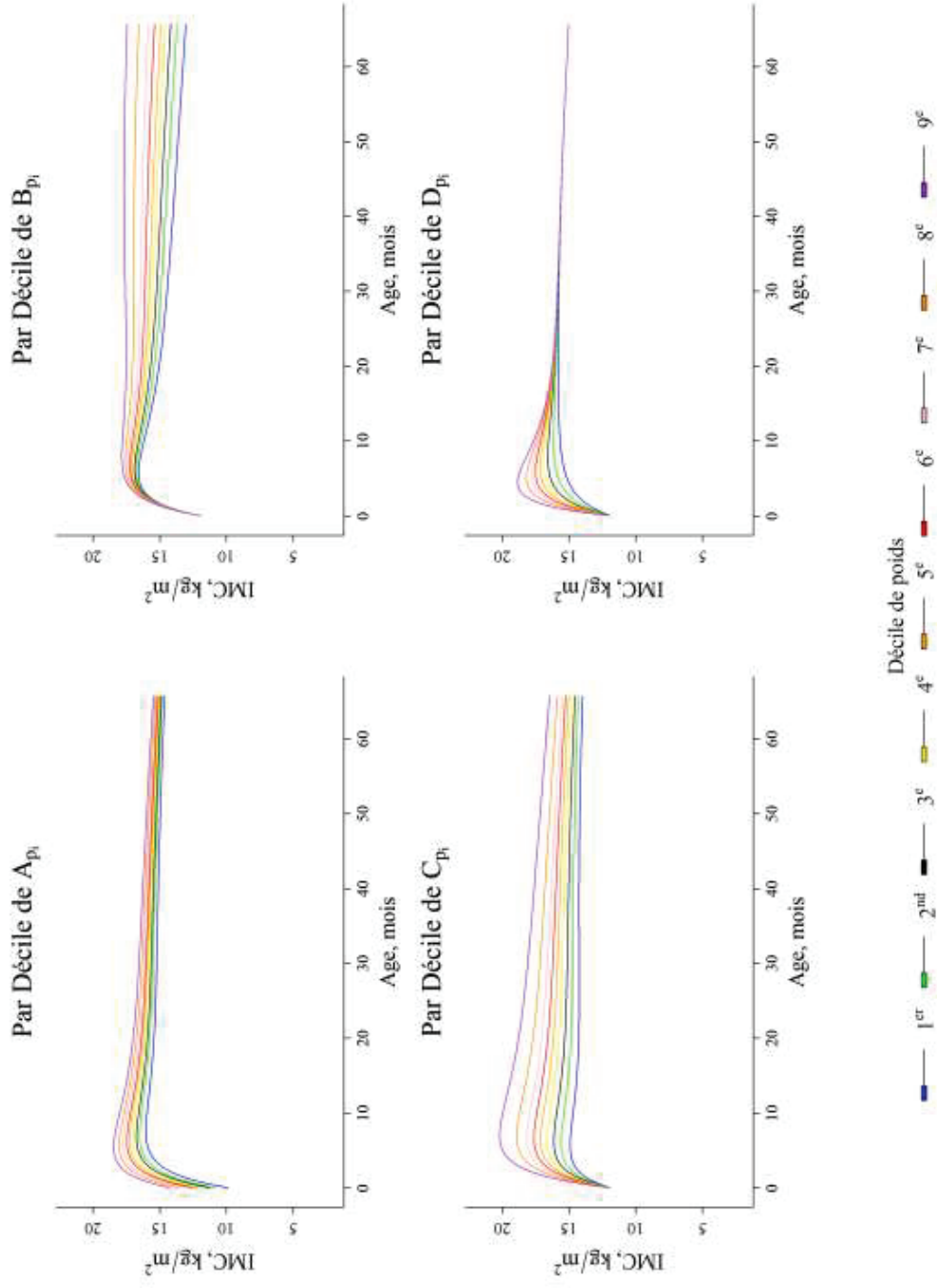


Figure 9 : Effet de la variation des paramètres du modèle de croissance du poids sur les trajectoires d'IMC

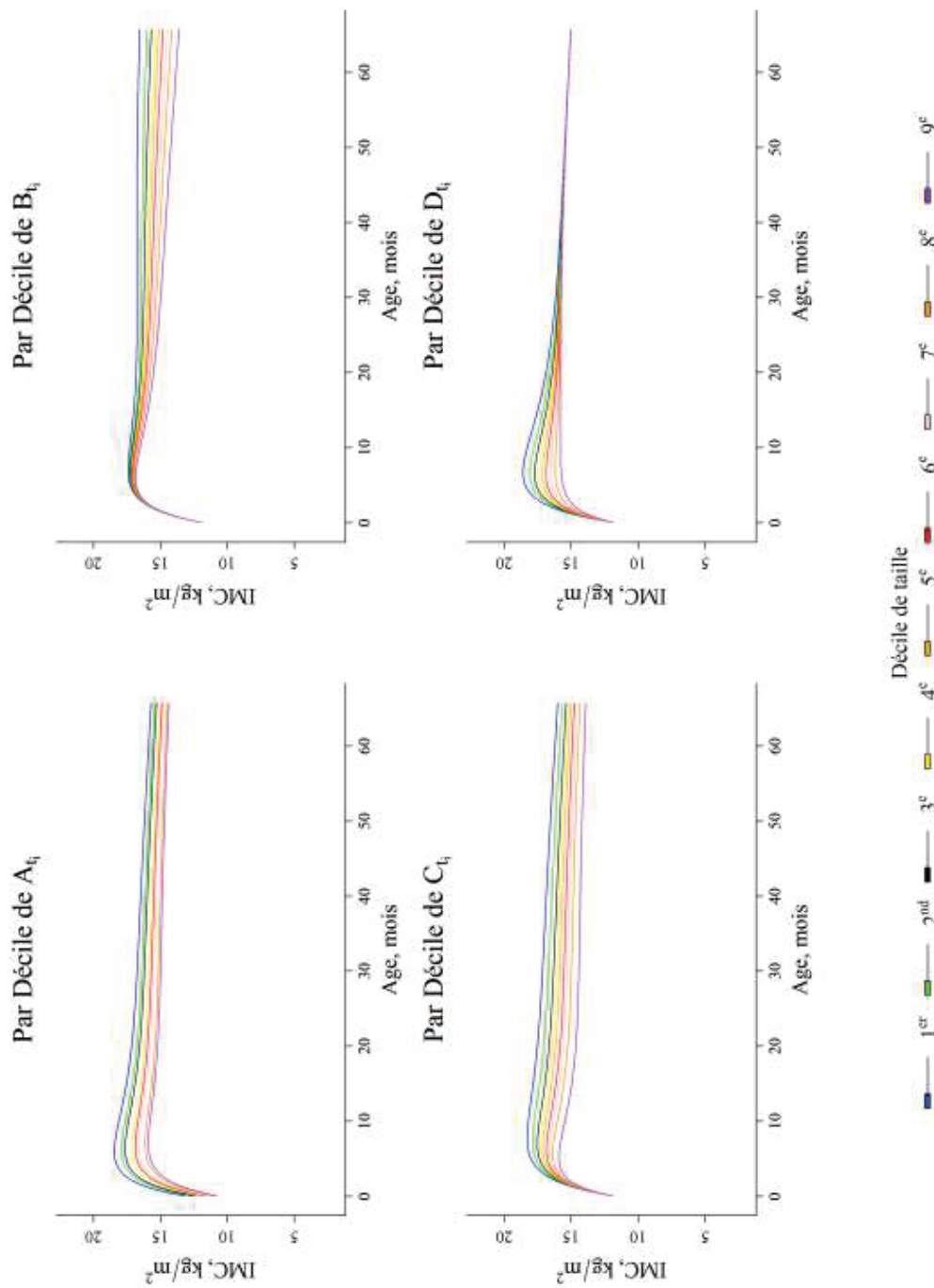


Figure 10 : Effet de la variation des paramètres du modèle de croissance de la taille sur les trajectoires d'IMC

Détermination des trajectoires individuelles des vitesses de croissance du poids, de la taille et de l'IMC

Les trajectoires des vitesses de croissance individuelles du poids, de la taille et de l'IMC ont été calculées en calculant les dérivées premières par rapport au temps des équations individuelles de poids, taille et IMC respectivement.

Les équations obtenues sont les suivantes :

$$\widehat{Vitesse_Poids}_{ij} = \exp(\widehat{B}_{P_i}) + \exp(\widehat{C}_{P_i} + \widehat{D}_{P_i} - \exp(\widehat{D}_{P_i}) * t_{ij})$$

$$\widehat{Vitesse_Taille}_{ij} = \exp(\widehat{B}_{T_i}) + \exp(\widehat{C}_{T_i} + \widehat{D}_{T_i} - \exp(\widehat{D}_{T_i}) * t_{ij})$$

$$\widehat{Vitesse_IMC}_{ij} = \frac{\widehat{Vitesse_Poids}_{ij} * \left(\frac{\widehat{Taille}_{ij}}{100}\right)^2 - 2 * \frac{\widehat{Taille}_{ij}}{100} * \frac{\widehat{Vitesse_Taille}_{ij}}{100} * \widehat{Poids}_{ij}}{\left(\frac{\widehat{Taille}_{ij}}{100}\right)^4}$$

L'effet de la variation des paramètres du modèle poids sur les trajectoires de vitesses de croissance du poids sont illustrées dans la figure 11 (nous avons procédé de la même manière que pour l'illustration des trajectoires de poids, cf. p69).

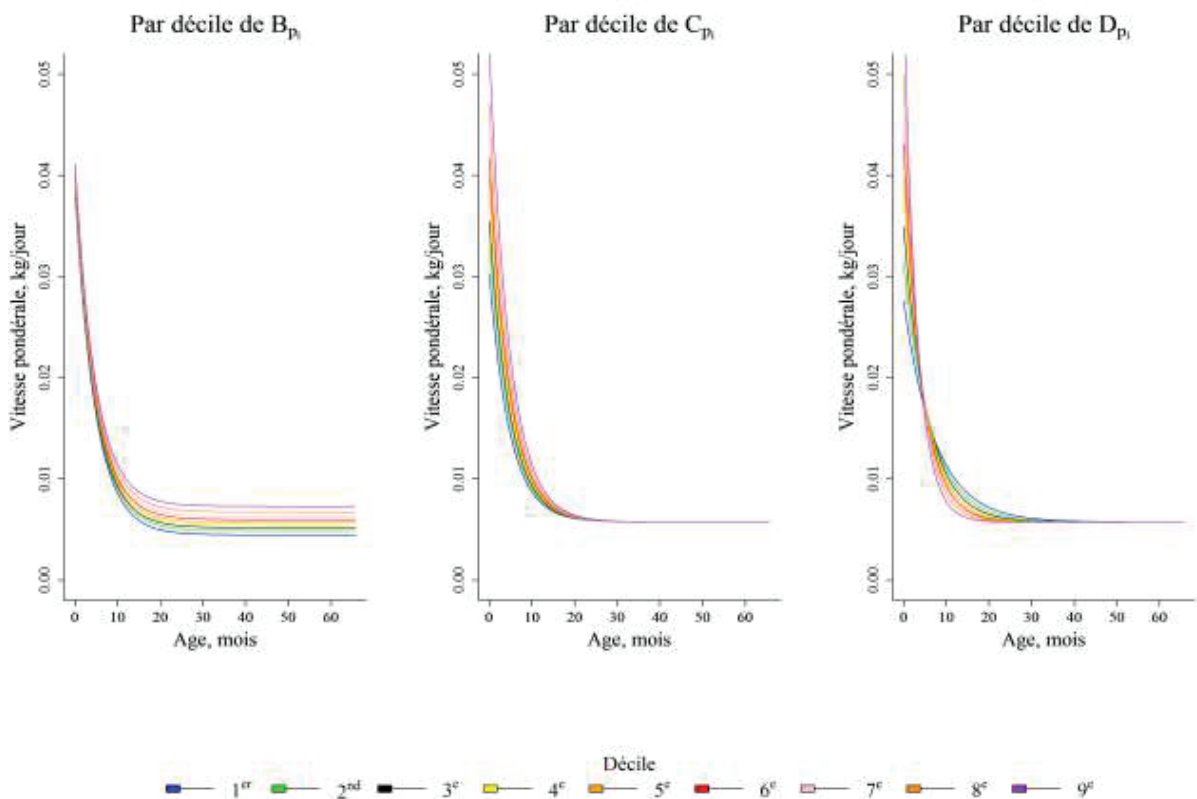


Figure 11 : Trajectoires de vitesse de croissance du poids de la naissance à 5 ans par déciles de paramètres de poids obtenus sur un sous-échantillon de la population EDEN

Prédictions individuelles des poids, taille, IMC et vitesses de croissance du poids, de la taille et de l'IMC

Les poids, taille, IMC et les vitesses correspondantes individuels ont été calculés en remplaçant l'âge de mesure par l'âge d'intérêt (en jours) dans les équations de chaque individu à 1, 3, 6 et 9 mois et 1, 2, 3, 4 et 5 ans.

Détermination du pic d'IMC (âge et IMC)

Les vitesses de croissance journalières de l'IMC de chaque enfant de la naissance à 5 ans ont été calculées. L'âge au pic d'IMC a été défini comme l'âge auquel le signe des vitesses de croissance consécutives s'inversait ($\widehat{Vitesse_IMC}_{ij} / \widehat{Vitesse_IMC}_{ij+1} < 0$).

A partir des équations individuelles de croissance de l'IMC, en remplaçant l'âge de mesure par l'âge au pic d'IMC obtenu pour chaque individu, nous avons pu définir l'IMC correspondant au pic d'IMC.

Pour certains enfants (14 garçons et 35 filles), le pic d'IMC a été identifié au 1^{er} jour après la naissance (1 garçon), n'était pas identifiable (3 filles) ou seulement tardivement après 2 ans (13 garçons et 32 filles). Nous avons identifié ces enfants et inspecté leurs trajectoires. Il ne s'agit pas d'enfants pour lesquels la trajectoire est particulièrement mal modélisée, mais le caractère atypique de ces trajectoires nous a conduits à exclure ces 49 enfants des analyses portant sur l'âge et l'IMC au pic d'IMC.

La Figure 12 donne quelques exemples de trajectoires d'évolution de l'IMC pour des enfants dont le pic d'IMC a été identifié après 2 ans.

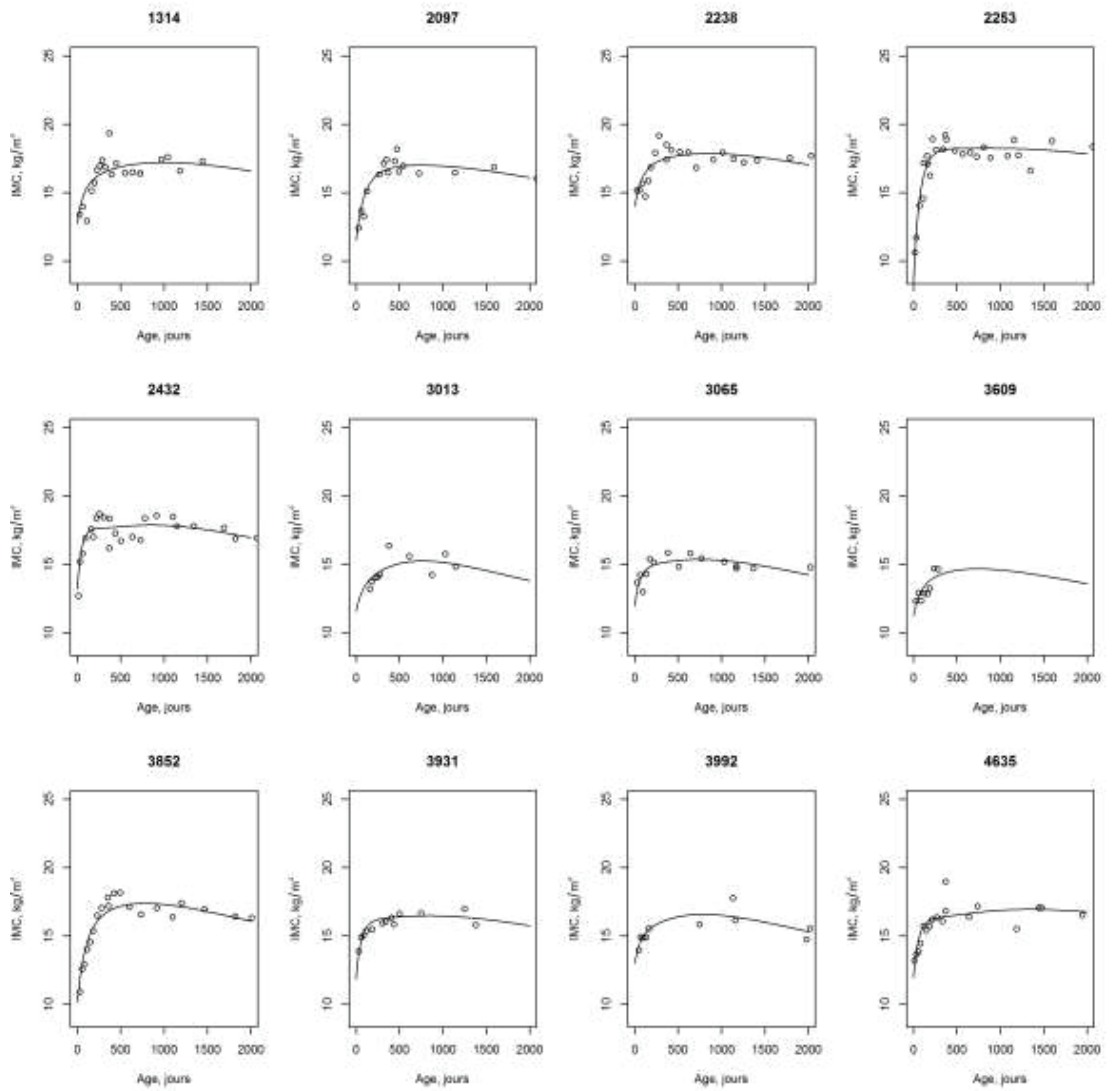


Figure 12 : Trajectoires de croissance d'IMC pour 12 sujets avec un âge au pic d'IMC excédant 2 ans

3.3.2.4.2.2.2 Etape 2 : Analyse des associations avec le tabagisme maternel pendant la grossesse

Les modèles n'ont pas été utilisés pour prédire les paramètres à la naissance (la modélisation du poids ne permet pas de prédire le poids à la naissance), ce sont les tailles et poids mesurés qui ont été considérés dans les analyses.

Les associations entre le tabagisme maternel et les tailles et poids de naissance, les poids, tailles et IMC prédits aux âges sélectionnés et les âges et IMC au pic d'IMC ont été analysées à l'aide de modèles de régression linéaires multivariés ajustés sur les facteurs de confusions potentiellement précédemment décrits.

3.3.2.4.2.3 Analyse de la relation dose-effet

Pour explorer l'existence d'une relation dose-effet entre le tabagisme maternel tout au long de la grossesse et la croissance des enfants, les analyses ont été répétées en considérant le nombre moyen de cigarettes fumées quotidiennement durant les 2nd et 3^e trimestres comme exposition principale.

3.3.2.4.2.4 Analyses de sensibilité

1. Confusion par le risque de prématurité de l'enfant

Le tabagisme de la mère pendant la grossesse a été associé à un risque accru de naissance prématurée (Jaddoe et al., 2008). La prématurité a quant à elle été associée à la croissance (cf. chapitre 1).

Une variable permettant de distinguer les enfants nés prématurément (avant 37 SA) et les enfants nés à terme a été créée et les analyses ont été répétées en excluant les 88 enfants nés avant 37 semaines de grossesse (5%) afin d'évaluer un phénomène de confusion lié à la présence d'enfants prématurés dans l'échantillon analysé.

2. Prise en compte de la confusion résiduelle liée aux modes de vie partagés

De nombreuses études conduites pour étudier les associations entre le tabagisme de la mère pendant la grossesse et le risque de surpoids ou d'obésité de leurs enfants ou avec leurs caractéristiques anthropométriques discutent l'existence d'un phénomène de confusion résiduelle liée à des modes de vie plus obésogènes partagés dans cette association, malgré l'ajustement sur un grand nombre de variables socio-économiques et familiales.

- Exposition au tabagisme en période postnatale

Pour tenter d'explorer l'existence de facteurs familiaux non mesurés susceptibles de biaiser les associations observées entre le tabagisme de la mère pendant la grossesse et la croissance postnatale des enfants, les associations entre l'exposition postnatale au tabagisme et la croissance ont été étudiées. L'hypothèse de cette analyse complémentaire est que, dans le contexte d'un phénomène de confusion résiduelle, le tabagisme serait un proxy du mode de vie et les mêmes associations devraient être observées. Afin de pouvoir explorer les associations entre la croissance et une exposition au tabagisme durant la période postnatale uniquement, les paramètres

prédits de la croissance des enfants de mères ne fumant pas pendant la grossesse mais dont la mère commençait ou recommençait à fumer après la grossesse et/ou dont les pères fumaient ont été comparés à ceux des enfants dont les mères ne fumaient ni pendant la grossesse ni après leur grossesse et dont le père ne fumait pas.

- Confusion par les pratiques alimentaire

Les données d'allaitement et les données alimentaires collectées ont permis, dans un précédent travail, conduit par Betoko et coll. de déterminer des profils de pratiques alimentaires des enfants inclus dans l'étude EDEN (Betoko et al., 2013). Trois profils ont été identifiés. Le premier était caractérisé par une introduction tardive de tous les aliments, par l'utilisation fréquente d'aliments spécifiques « bébé » (légumes, petits pots de fruits et de plat complets) et par une faible utilisation d'aliments faits maison ou d'aliments ordinaires du commerce. Il a été appelé « Diversification tardive et utilisation d'aliments spécifiques pour bébé ». Le 2^e profil était caractérisé par un allaitement long, une introduction tardive de viandes, poissons, fruits, légumes, farines, pommes de terre, laitages, biscuits et jus de fruits, par l'utilisation fréquente d'aliments faits maison (soupes, purées de légumes, compotes, plats à base de viandes et poissons frais) et par une utilisation peu fréquente d'aliments du commerce spécifiques bébé. Il a été nommé « Allaitement maternel long, diversification tardive et utilisation d'aliments faits maison ». Le 3^e profil était caractérisé par l'utilisation fréquente d'aliments ordinaires du commerce (laitages, soupes, compotes, jus de fruits, biscuits) et une utilisation peu fréquente de laitages spécifiques bébé. Il a été appelé « Utilisation fréquente d'aliments ordinaires du commerce ».

Afin d'explorer un biais lié à l'alimentation de l'enfant dans la 1^{re} année de vie, les associations entre le tabagisme maternel et la croissance ont été étudiées en ajustant sur ces profils de pratiques alimentaires en supprimant la durée d'allaitement maternel qui est déjà prise en compte dans la construction de ces profils. Cette analyse a été restreinte à un sous-échantillon de 971 enfants pour lesquels le profil de pratiques alimentaires était connu.

3. Confusion par la taille des parents

Les associations entre le tabagisme maternel et la croissance ont été étudiées en ajustant sur la taille des parents. La taille des mères était disponible pour l'ensemble des

enfants (la présence de l'information sur l'IMC de la mère étant un des critères pour l'inclusion dans les analyses). La taille des pères n'était disponible que pour un sous-échantillon de 1334 enfants.

4. Comparaison de la méthode de « modélisation indirecte » de l'IMC avec d'autres méthodes de modélisation déjà utilisées dans la littérature

Les différences entre l'approche utilisée dans ce travail pour la prédiction des trajectoires de croissance de l'IMC et d'autres approches déjà utilisées pour la modélisation de l'IMC (cf Chapitre 1, p33) ont été explorées. Pour cela, la méthode en « 2 étapes » a été répliquée en modélisant les trajectoires d'IMC par la méthode des polynômes fractionnaires et par l'utilisation de splines cubiques.

La modélisation par polynômes fractionnaires a été réalisée en appliquant la méthode décrite par Wen et coll. dans le cadre de la modélisation de l'IMC (Wen et al., 2012). L'IMC attendu a été modélisé selon l'équation :

$$E(IMC) = b_0 + \sum_{j=1}^m b_j Age^{p_j}$$

Où m désigne le degré (ordre) du modèle et les puissances p_j ont été choisies parmi les 8 valeurs candidates -2, -1, -0,5, 0 ou log, 0,5, 1, 2 et 3. La méthode a été simplifiée par rapport à celle de Royston et Altman décrite dans le chapitre 1. En effet, les répétitions de puissance n'ont pas été considérées en raison du grand nombre de modèles à tester. Toutes les combinaisons de puissances possibles sans remise ont été considérées, soit 247 modèles : 28 modèles de degré 2, 56 modèles de degré 3, 70 modèles de degré 4, 56 modèles de degré 5, 25 modèles de degrés 6, 8 modèles de degré 7 et 1 modèle de degré 8. Comme la trajectoire d'IMC typique d'un enfant de la naissance à 5 ans inclut un point de cassure (le pic d'IMC), le degré minimal a été fixé à 2. Ces modèles ont été ajustés en utilisant des modèles linéaires à effets mixtes (package nlme de R (Pinheiro et al., 2015)).

Les modèles ont d'abord été comparés par degré en retenant le modèle minimisant le critère d'Akaike (AIC) et le critère d'information bayésien (BIC).

Les modèles retenus au sein de chaque degré ont été comparés entre eux (par comparaison des AIC et BIC ainsi qu'en considérant la plausibilité des trajectoires connaissant la forme supposée de la trajectoire d'IMC, et nous avons retenu celui qui

s'ajustait le mieux aux données, sans sur-ajustement aux extrêmes de la distribution d'âge.

En analyse supplémentaire, en fonction du degré retenu et, de ce fait, du nombre de modèles possibles, la plus-value de l'inclusion de répétitions de puissance a été testée, plus conformément à la méthode décrite par Royston et Altman (Royston and Altman, 1994).

La modélisation par splines

Etant donné la forme de la trajectoire d'évolution de l'IMC au cours du temps, le nombre de nœud minimal a été fixé à 2. Toutes les combinaisons de deux nœuds parmi un jeu de valeurs candidates ont été testées. Pour cela, un premier nœud a été fixé à 10 jours et nous avons ensuite fixé les autres nœuds possibles par intervalle de 100 jours (110 jours, 210 jours, etc.).

Le modèle comprenant les deux nœuds minimisant l'AIC et le BIC a été choisi comme point de départ. Il a ensuite été vérifié si des splines incluant plus de nœuds permettaient de mieux s'ajuster aux données. Pour cela, en fixant l'emplacement des deux nœuds identifiés comme minimisant l'AIC/BIC, nous avons testé tous les modèles à 3 nœuds en fixant ce 3^e nœud aux âges restants. A nouveau, le meilleur modèle a été identifié et un nœud a été ajouté selon la même procédure. Nous avons testés jusqu'à l'inclusion de 6 nœuds et déterminé le meilleur modèle. Ces modèles ont été ajustés dans R à l'aide du package `sme` ajustant des modèles à effets mixtes utilisant des splines cubiques de lissage (Berk, 2013).

D'autres méthodes de sélection de la position des nœuds ont été évaluées (comme analyse de sensibilité, chez les filles uniquement) en plaçant de 2 à 10 nœuds aux percentiles de la distribution de l'âge ($100/k+1$; $2*100/k+1$ $k*100/k+1$ où k représente le nombre de nœuds), ou en écartant les nœuds les plus extrêmes, comme recommandé par Harrell pour mieux modéliser la courbe pour les petites et grandes valeurs de l'âge dans notre cas (Harrell, 2001), en proposant, pour 3 nœuds, de les placer aux 10^e, 50^e et 90^e percentiles ; pour 4 nœuds aux 5^e, 27,5^e, 50^e, 72,5^e, 95^e percentiles; pour 5 nœuds aux 5^e, 23^e, 41^e, 59^e, 77^e et 95^e percentiles et pour 7 nœuds aux 2,5^e, 18,33^e, 34,17^e, 50^e, 65,83^e, 81,67^e et 97,5^e percentiles.

4. Comparaison des associations obtenues avec les données prédites et les données observées

Afin de comparer les résultats obtenus en considérant les données observées et les données issues de la modélisation, les analyses ont été répétées à chaque âge où un examen clinique avait été réalisé (1, 3 et 5 ans). Le nombre d'enfants présents à ces examens était respectivement de 1513, 1204 et 1084. Même si les examens cliniques avaient lieu à des âges fixés a priori, il existe une certaine variabilité entre les âges des enfants pour un même examen clinique. Les modèles de régressions linéaires permettant d'étudier les associations ont donc été ajustés sur l'âge de l'enfant au moment de l'examen clinique.

3.4 Résultats

3.4.1 Caractéristiques de la population d'étude

Parmi les 1907 enfants dont la date de naissance était connue, 19 n'avaient aucune mesure de poids ou de taille et 130 et 80 n'avaient qu'une seule mesure de taille et/ou de poids respectivement. Sur les 1764 enfants ayant au moins 2 mesures de taille et de poids, 98 avaient au moins une information manquante concernant les facteurs de confusion potentiels. L'échantillon disponible pour les analyses était de 1666 enfants.

Le nombre d'individus et le nombre de mesures disponibles pour la modélisation de la croissance du poids et de la taille sont résumés dans la Figure 13.

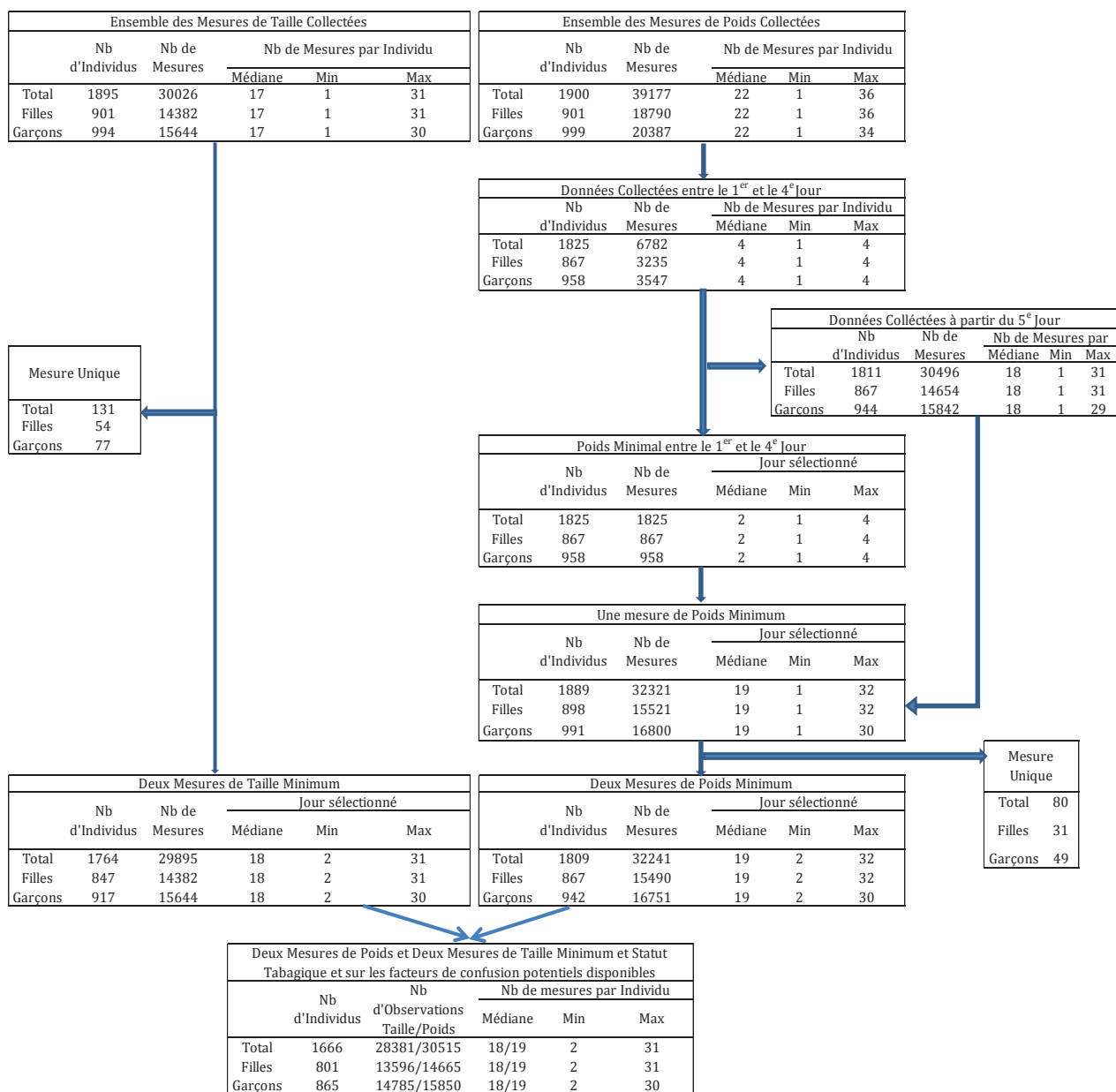


Figure 13 : Diagramme de flux des mesures de poids et de taille

Le nombre total de mesures de poids était de 30515 et le nombre de mesures de taille de 28381. La médiane (min-max) par sujet du nombre de mesures de poids et de taille était de 19 (2-32) et 18 (2-31) pour le poids et la taille respectivement (tableau 2).

Tableau 2 : Nombre de mesures de poids et de taille des enfants disponibles pour la modélisation de la croissance pondérale et staturale (n=1666), EDEN, France, 2003-2012

Tranche d'âge	Nb. d'Individus ^a			Nb. de Mesures			Nb. de Mesures par Sujet ^b		
	Garçons	Filles	Total	Garçons	Filles	Total	Garçons	Filles	Total
Poids^c									
<6 mois	864	801	1665	5198	4864	10062	6 (1-9)	6 (1-9)	6 (1-9)
6 mois - 1 an	814	754	1568	3263	2987	6250	4 (1-9)	4 (1-8)	4 (1-9)
1 - 2 ans	788	721	1509	2838	2686	5524	4 (1-9)	4 (1-9)	4 (1-9)
2 - 4 ans	750	694	1444	3055	2820	5875	4 (1-9)	4 (1-10)	4 (1-10)
>= 4 ans	635	573	1208	1496	1308	2804	2 (1-5)	2 (1-4)	2 (1-5)
Ensemble de la période d'étude	865	801	1666	15850	14665	30515	20 (2-30)	19 (2-32)	19 (2-32)
Taille									
<6 mois	863	801	1664	4765	4449	9214	6 (1-9)	6 (1-8)	6 (1-9)
6 mois - 1 an	808	753	1561	3045	2756	5801	4 (1-9)	4 (1-8)	4 (1-9)
1 - 2 ans	784	715	1499	2697	2555	5252	3 (1-9)	4 (1-9)	4 (1-9)
2 - 4 ans	748	691	1439	2819	2576	5395	4 (1-9)	4 (1-10)	4 (1-10)
>= 4 ans	633	570	1203	1459	1260	2719	2 (1-5)	2 (1-4)	2 (1-5)
Ensemble de la période d'étude	865	801	1666	14785	13596	28381	18 (2-30)	18 (2-31)	18 (2-31)

^aSujets ayant plus d'une mesure

^bMédiane (min-max)

^cExcluant le poids de naissance et incluant la mesure correspondant au poids minimal dans les 4 premiers jours

Parmi les femmes incluses dans les analyses, 24% déclaraient avoir fumé durant au moins un trimestre de leur grossesse. Une minorité d'entre elles fumait exclusivement durant leur 1^{er} trimestre de grossesse (n=132, 7,9%), et la majorité (n=275, 16,5%), déclarait fumer durant le 2^e et/ou 3^e trimestre(s) de leur grossesse. Pour les femmes dont les autres facteurs n'étaient pas disponibles, les prévalences étaient assez comparables (8,5% pour le 1^{er} trimestre exclusivement et 16,5% pour le 2^e et/ou 3^e trimestre).

Comparées aux mères ne fumant pas pendant leur grossesse, les mères fumant soit au 1^{er} trimestre soit tout au long de la grossesse étaient plus jeunes au moment de l'accouchement (respectivement $28,5 \pm 5$ et $28,5 \pm 5,4$ ans *versus* $30,1 \pm 4,6$) et avaient un niveau d'études plus faible (40% des non fumeuses avaient un niveau d'études équivalent au lycée ou inférieur et ce pourcentage augmentait graduellement avec la durée de tabagisme pendant la grossesse atteignant 51,5% chez les fumeuses exclusivement au 1^{er} trimestre et 69,8% chez les fumeuses tout au long de la grossesse).

Les femmes fumant tout au long de la grossesse initiaient moins souvent un allaitement au sein que les non fumeuses (61,8% vs 76,9%). Parmi les femmes initiant un allaitement maternel, comparées aux mères non fumeuses, les femmes fumant au 1^{er} trimestre ou tout au long de la grossesse allaitaient moins longtemps (respectivement 3 et 2 mois en moyenne *versus* 4 mois).

Les enfants de femmes fumeuses tout au long de la grossesse avaient un plus petit poids de naissance que les enfants de non fumeuses (3157 ± 486 *versus* 3311 ± 507 g) et une taille de naissance plus faible ($49,0 \pm 2,4$ *versus* $49,7 \pm 2,3$ cm). Ces paramètres anthropométriques ne différaient pas significativement entre les fumeuses exclusivement au 1^{er} trimestre et les non fumeuses (tableau 3).

Tableau 3 : Caractéristiques des enfants (n = 1666) et de leurs parents selon le statut tabagique maternel pendant la grossesse, EDEN, France, 2003-2012

	Non Fumeuses (n = 1259, 75,57 %)			Exclusivement le 1 ^{er} trimestre (n = 132, 7,92 %)			Fin de grossesse (n = 275, 16,51 %)		
	%	Moyenne (ÉT) ou Médiane (min-max)	P ^a	%	Moyenne (ÉT) ou Médiane (min-max)	P ^a	%	Moyenne (ÉT) ou Médiane (min-max)	P ^b
Caractéristiques maternelles									
Age		30,1 (4,6)	<0,001		28,5 (5)			28,5 (5,4)	<0,001
Taille, cm		163,6 (6,1)	0,85		163,7 (6)			163,3 (6,4)	0,54
Poids, kg		62,5 (12,7)	0,08		60,5 (12,1)			61,6 (12,7)	0,33
IMC ^c , kg/m ²		22,2 (14,9 - 60,4)	0,02		21,4 (15,6 - 38)			21,8 (14 - 43)	0,28
En surpoids ou obèse	25,9		0,09	23,5		0,29	26,6		0,08
Centre de recrutement (Poitiers)	45,9		0,29	50,8		53,1		53,1	0,03
Niveau d'études									
Niveau Lycée ou moins	38,8		<0,001	51,5		69,8		69,8	<0,001
BAC+2	23,8			20,4		15,6		15,6	
>BAC+2	37,4			28		14,6		14,6	
Primipare	43,8		0,004	56,8		43,3		43,3	0,86
Nombre de cigarettes par jour									
1 ^{er} Trimestre					6,0 (1-35)			8,0 (1-40)	
2 ^{ème} Trimestre								5,0 (1-20)	
3 ^{ème} Trimestre								5,0 (1-25)	
Caractéristiques paternelles									
IMC ^c , kg/m ²		24,8 (15,8 - 40,8)	0,12		24,5 (18 - 35,2)			24,3 (17,6 - 35,5)	0,07
Statut tabagique		29,7	<0,001		62,1			75,1	<0,001
Caractéristiques de l'enfant									
Sexe Masculin	52,3			49,2		0,51	51,6		0,85
Age gestationnel, semaines ^e		40 (28 - 42)	0,92		40 (34 - 42)			39 (30 - 42)	0,36
Taille à la naissance, cm		49,71 (23,1)	0,19		49,43 (2,02)			49,03 (2,39)	<0,001
Poids à la naissance, g		3311 (507)	0,92		3316 (455)			3157 (486)	<0,001
Allaitement Oui/Non	76,9		0,14	71,2		0,14	61,8		<0,001
Allaitement ≥3mois	50,2		0,04	40,9		0,04	26,9		<0,001
Durée d'allaitement chez les enfants allaités, mois		4 (0,2 - 13)	<0,001		3 (0,2 - 13)			2 (0,2 - 13)	<0,001
Prématurité	5,8		0,09	2,3		0,09	4,4		0,35

Abbreviations: IMC, Indice de Masse Corporelle; ÉT, Écart-type

^aComparison entre les caractéristiques des mères fumeur exclusivement le 1^{er} trimestre, leurs enfants et leurs partenaires et celles des non-fumeuses en utilisant des tests du X² ou des tests de Student

^bComparison entre les caractéristiques des mères fumeur en fin de grossesse, leurs enfants et leurs partenaires et celles des non-fumeuses en utilisant des tests du X² ou des tests de Student

^cMédiane(min-max) présentés en présence de distribution asymétrique

3.4.2 Caractéristiques anthropométriques des enfants

Nous présentons dans cette partie les caractéristiques de croissance de la population d'étude (cf. méthode décrite p 74).

3.4.2.1 Croissance pondérale

3.4.2.1.1 Paramètres du modèle de croissance pondérale

Le Tableau 4 décrit la distribution des paramètres individuels caractérisant les trajectoires de croissance des enfants inclus dans les analyses.

Tableau 4 : Distribution des paramètres de croissance du poids des enfants (0-5 ans)

Paramètres des trajectoires pondérales	Sexe								P
	Garçons				Filles				
	Moyenne	Min	Max	écart-type	Moyenne	Min	Max	écart-type	
A	1,10	0,01	1,56	0,17	1,06	-0,27	1,46	0,18	<0,001
B	-5,16	-5,73	-4,34	0,20	-5,16	-5,87	-4,20	0,21	0,77
C	1,64	0,49	2,51	0,23	1,52	0,63	2,73	0,27	<0,001
D	-4,94	-6,52	-3,50	0,34	-5,00	-6,82	-3,96	0,35	<0,001

Les paramètres A, C et D étaient significativement plus faibles chez les filles comparés aux garçons.

Ceci illustre un poids plus faible chez les filles dans les premiers jours de vie, et un gain pondéral de la naissance à 2 ans plus faible. La vitesse de croissance pondérale dans la phase linéaire de croissance (paramètre B) n'était pas significativement différente selon le sexe.

3.4.2.1.2 Prédictions individuelles de poids

Le Tableau 5 résume les distributions des prédictions de poids et de vitesses instantanées de croissance du poids dans la population étudiée.

Tableau 5 : Prédictions des poids et des vitesses de croissance pondérale instantanées, EDEN, 0-5 ans

Prédictions	Sexe								
	Garçons				Filles				
	Moyenne	Min	Max	écart-type	Moyenne	Min	Max	écart-type	
Poids, kg									
1 mois	4,26	1,85	6,14	0,51	3,97	1,61	5,40	0,49	
3 mois	6,07	2,87	8,57	0,64	5,58	2,74	7,35	0,62	
6 mois	7,87	4,14	10,70	0,76	7,23	4,32	9,43	0,76	
9 mois	9,07	5,27	12,68	0,85	8,39	5,78	11,42	0,86	
1 an	9,97	6,27	14,22	0,93	9,27	6,60	13,16	0,95	
2 ans	12,55	9,39	18,10	1,20	11,87	8,40	17,63	1,24	
3 ans	14,76	11,04	20,79	1,47	14,09	10,16	21,83	1,53	
4 ans	16,91	12,67	24,62	1,79	16,26	11,73	27,29	1,89	
5 ans	19,05	14,30	29,40	2,15	18,41	13,18	32,76	2,28	
Vitesses de croissance, kg/mois									
1 mois	1,09	0,49	1,70	0,19	0,95	0,44	1,48	0,17	
3 mois	0,75	0,39	1,14	0,11	0,68	0,41	1,08	0,10	
6 mois	0,48	0,26	0,78	0,08	0,45	0,24	0,80	0,08	
9 mois	0,34	0,17	0,60	0,07	0,33	0,18	0,64	0,06	
1 an	0,27	0,14	0,49	0,05	0,27	0,16	0,52	0,05	
2 ans	0,19	0,12	0,40	0,04	0,19	0,12	0,46	0,04	
3 ans	0,18	0,11	0,40	0,04	0,18	0,11	0,46	0,04	
4 ans	0,18	0,10	0,40	0,04	0,18	0,10	0,46	0,04	
5 ans	0,18	0,10	0,40	0,04	0,18	0,10	0,46	0,04	

3.4.2.2 Croissance staturale

3.4.2.2.1 Paramètres du modèle de croissance staturale

Le Tableau 6 résume les distributions des paramètres individuels décrivant les trajectoires de croissance staturale dans la population étudiée.

Tableau 6 : Distribution des paramètres de croissance staturale, EDEN, 0-5 ans

Paramètres des trajectoires staturales	Sexe								P
	Garçons				Filles				
	Moyenne	Min	Max	écart-type	Moyenne	Min	Max	écart-type	
A	3,92	3,65	4,02	0,04	3,90	3,59	4,00	0,04	<0,001
B	-3,89	-4,22	-3,46	0,10	-3,90	-4,23	-3,39	0,12	0,38
C	3,14	2,73	3,70	0,13	3,11	2,40	3,60	0,15	<0,001
D	-5,37	-6,40	-4,51	0,26	-5,45	-6,48	-4,46	0,28	<0,001

Comme pour le poids, les paramètres A, C et D étaient significativement plus faibles chez les filles comparés aux garçons. Ceci illustre une taille à la naissance plus faible et un gain statural de la naissance à 2 ans plus faible. La vitesse de croissance staturale dans la phase linéaire de croissance (paramètre B) n'était pas significativement différente selon le sexe.

3.4.2.2 Prédications de la taille

Le tableau 7 résume les distributions des prédictions de taille et de vitesses instantanées de croissance staturale dans la population étudiée.

Tableau 7 : Prédications de taille et de vitesses instantanées de croissance staturale, EDEN, 0-5 ans

Prédications	Sexe							
	Garçons				Filles			
	Moyenne	Min	Max	écart-type	Moyenne	Min	Max	écart-type
Taille, cm								
1 mois	54,2	42,6	59,2	1,99	52,9	40,8	58,6	2,04
3 mois	60,4	49,7	66,7	2,06	58,7	47,6	65,3	2,07
6 mois	67,4	56,2	74,6	2,19	65,4	54,4	72,3	2,2
9 mois	72,6	61,2	79,6	2,24	70,5	60,3	77,2	2,29
1 an	76,6	65,7	83,7	2,29	74,5	65,4	82	2,4
2 ans	87,6	77,9	96,6	2,74	85,8	78,1	95,5	2,9
3 ans	95,9	84,9	105,9	3,18	94,2	85,3	105,1	3,33
4 ans	103,6	91,8	114,4	3,63	101,9	91,8	115,1	3,77
5 ans	111,1	98,7	124,7	4,15	109,5	98,2	124,9	4,3
Vitesses de croissance, cm/mois								
1 mois	3,49	2,12	5,69	0,52	3,21	1,76	5,04	0,47
3 mois	2,74	1,9	3,66	0,28	2,58	1,65	3,64	0,27
6 mois	1,99	1,52	2,66	0,16	1,93	1,43	2,7	0,16
9 mois	1,51	1	2,14	0,15	1,5	0,98	2,13	0,15
1 an	1,21	0,77	1,76	0,13	1,22	0,79	1,73	0,13
2 ans	0,75	0,54	1,04	0,07	0,76	0,56	1,04	0,07
3 ans	0,65	0,48	0,96	0,06	0,66	0,51	1,03	0,06
4 ans	0,63	0,47	0,96	0,06	0,63	0,48	1,03	0,07
5 ans	0,63	0,47	0,96	0,06	0,63	0,46	1,03	0,07

3.4.2.3 Croissance de l'IMC

3.4.2.3.1 Prédictions individuelles d'IMC

Le Tableau 8 résume les distributions des prédictions d'IMC et de vitesses instantanées de croissance de l'IMC dans la population étudiée.

Tableau 8 : Prédictions d'IMC et de vitesses instantanées de croissance de l'IMC, EDEN, 0-5 ans

Prédictions	Sexe							
	Garçons				Filles			
	Moyenne	Min	Max	écart-type	Moyenne	Min	Max	écart-type
IMC, kg/m²								
1 mois	14,5	8,9	18,8	1,1	14,1	9,0	18,0	1,1
3 mois	16,6	11,2	21,0	1,2	16,2	12,0	21,0	1,3
6 mois	17,3	13,1	21,6	1,2	16,9	13,0	22,0	1,3
9 mois	17,2	14,0	21,6	1,2	16,9	14,0	23,0	1,3
1 an	17,0	14,0	21,7	1,2	16,7	13,0	23,0	1,3
2 ans	16,3	13,1	20,9	1,1	16,1	13,0	23,0	1,2
3 ans	16,0	12,7	21,4	1,1	15,9	13,0	22,0	1,2
4 ans	15,7	12,4	22,3	1,1	15,6	13,0	23,0	1,2
5 ans	15,4	12,1	22,6	1,2	15,3	12,0	24,0	1,3
Vitesses de croissance, kg/m²/mois								
1 mois	1,87	-0,12	3,87	0,58	1,69	0,40	3,70	0,54
3 mois	1,08	-0,25	2,61	0,39	1,00	0,10	2,30	0,36
6 mois	0,03	-0,35	0,55	0,13	0,06	-0,30	0,60	0,12
9 mois	-0,07	-0,29	0,28	0,09	-0,05	-0,40	0,30	0,09
1 an	-0,08	-0,26	0,19	0,06	-0,07	-0,30	0,20	0,07
2 ans	-0,03	-0,14	0,18	0,03	-0,03	-0,10	0,20	0,03
3 ans	-0,02	-0,11	0,11	0,03	-0,02	-0,10	0,10	0,03
4 ans	-0,03	-0,10	0,08	0,02	-0,02	-0,10	0,10	0,02
5 ans	-0,03	-0,09	0,06	0,02	-0,03	-0,10	0,10	0,02

3.4.2.3.2 Pic d'IMC

Sur l'ensemble de l'échantillon, l'IMC au pic d'IMC était de 17,5 et 17,1 en moyenne respectivement pour les garçons et les filles. L'âge au pic d'IMC était de 235 jours (7 mois ½) et de 275 jours (9 mois) respectivement pour les garçons et les filles. Les âges au pic d'IMC ainsi que les IMC atteints à ce pic sur la sous-population d'enfants

conservés dans les analyses portant sur l'âge et l'IMC au pic d'IMC (i.e. dont l'âge au pic d'IMC excédait 1 jour et n'excédait pas 2 ans) sont également présentés.

Tableau 9 : Age et IMC au Pic d'IMC

Caractéristiques au Pic d'IMC	Sexe									
	Garçons			Filles						
	Moyenne	Min	Max	écart-type	Effectif	Moyenne	Min	Max	écart-type	Effectif
Age, jours										
Ensemble des Enfants	235	1	1679	146	865	275	85	1991	225	798
Enfants dont les informations sur l'âge et l'IMC au pic d'IMC ont été conservées dans les analyses correspondantes	223	72	722	103	851	237	85	723	107	766
IMC, kg/m²										
Ensemble des Enfants	17,5	14,2	21,7	1,2	865	17,1	14	23,1	1,3	798
Enfants dont les informations sur l'âge et l'IMC au pic d'IMC ont été conservées dans les analyses correspondantes	17,5	14,2	21,7	1,2	851	17,1	14	23,1	1,3	766

3.4.3 Association entre le tabagisme maternel pendant la grossesse et la croissance des enfants de la naissance à 5 ans

La méthode en « 1 étape » permet de tester les associations globales entre le tabagisme maternel et les 4 paramètres du modèle de Jentsch-Bayley respectivement pour les croissances pondérale et staturale (tableau 10). La méthode en « 2 étapes » évalue les associations entre le tabagisme maternel et les prédictions de taille, poids, IMC et vitesses correspondantes à des âges spécifiques (tableau 11).

3.4.3.1 Croissance staturale

La Figure 14A1 illustre les trajectoires de croissance staturale moyennes par sous-groupe de tabagisme pour un jeu donné de caractéristiques socio-démographiques et la Figure 14A2 illustre les différences de trajectoires moyennes de taille entre les enfants de mères non fumeuses et ceux de mères fumant exclusivement au 1^{er} trimestre ou tout au long de la grossesse. Les trajectoires de croissance staturale étaient significativement différentes chez les enfants exposés au tabagisme maternel durant le 1^{er} trimestre exclusivement et tout au long de la grossesse ($P < 0,001$) comparées aux trajectoires des enfants de non fumeuses (tableau 10).

Comparées aux enfants de mères non fumeuses pendant la grossesse, les trajectoires de croissance de la taille des enfants exposés au tabagisme maternel durant le 1^{er} trimestre exclusivement étaient caractérisées par un paramètre D significativement plus élevé ($\beta = 0,074$, $IC_{95\%}$ [0,010 ; 0,137]) (tableau 10). Si l'on regarde les résultats de la modélisation en 2 étapes, on peut constater que ceci témoigne d'une croissance staturale plus rapide durant une courte période ne dépassant pas 3 mois environ (tableau 11). La vitesse de croissance de ces enfants était également plus rapide à 5 ans sans que leur taille moyenne ne soit plus élevée comparée à celle des enfants de non fumeuses.

Le tabagisme maternel tout au long de la grossesse était associé à un paramètre A significativement plus faible ($\beta = -0,015$, $IC_{95\%}$ [-0,021 ; -0,008]), ce qui est cohérent avec les résultats obtenus à partir des données observées qui montraient que les enfants de mères fumant tout au long de la grossesse mesuraient en moyenne 0,63 cm de moins à la naissance que les enfants de non fumeuses. Comparés aux enfants de femmes non fumeuses, les enfants exposés tout au long de la grossesse avaient une vitesse de

croissance staturale significativement plus rapide jusqu'à 6 mois (à 6 mois, $\beta=0,022$ cm/mois, $IC95\%$ [0,000 ; 0,044], tableau 11, modèle en 2 étapes) ce qui est cohérent avec le fait que leur paramètre D est plus élevé bien que non significativement ($\beta=0,038$, $IC95\%$ [-0,011 ; 0,087], Tableau 10, modèle en 1 étape).

3.4.3.2 Croissance pondérale

La Figure 14B1 illustre les trajectoires de croissance pondérale moyennes par sous-groupe de tabagisme et la Figure 14B2 illustre les différences de trajectoires moyennes de poids entre les enfants de mères non fumeuses et ceux de mères fumant exclusivement au 1^{er} trimestre ou tout au long de la grossesse.

Le tabagisme maternel durant la grossesse était associé à une modification significative ($P=0,03$) du profil de croissance pondérale des enfants de la naissance à 5 ans (tableau 10).

Pour les enfants de mères fumeuses tout au long de la grossesse, le poids de naissance était significativement plus faible que celui des enfants de non fumeuses ($\beta=-120$ grammes, $IC95\%$ [-187 ; -54]). Leur vitesse de croissance était plus rapide dans les premiers mois de vie, comme l'atteste un paramètre D plus élevé, bien que la différence soit non significative ($\beta=0,022$, $IC95\%$ [-0,037;0,081]). Le calcul des vitesses de croissance montrait que celles-ci étaient significativement plus élevées que celle des enfants de non fumeuses entre 1 et 2 ans ($\beta=0,009$ kg/mois, $IC95\%$ [0,002 ; 0,016] à 1 an et $\beta=0,006$ kg/mois, $IC95\%$ [0,001 ; 0,011] à 2 ans).

Cette vitesse de croissance accrue persistait de 2 à 5 ans ce qui est cohérent avec l'association positive (à la limite de la significativité) observée entre le tabagisme maternel tout au long de la grossesse et le paramètre B ($\beta=0,035$, $IC95\%$ [-0,001 ; 0,071]). Par conséquent, les enfants de femmes fumeuses tout au long de la grossesse avaient un poids moyen significativement plus élevé que celui des enfants de femmes non fumeuses à 4 et 5 ans ($\beta=249$ g, $IC95\%$ [1 ; 498] et $\beta=307$ g, $IC95\%$ [8 ; 606] respectivement).

3.4.3.3 Croissance de l'IMC

La Figure 14C1 illustre les trajectoires moyennes d'évolution de l'IMC par sous-groupe de tabagisme et la Figure 14C2 illustre les différences de trajectoires moyennes

d'IMC entre les enfants de mères non fumeuses et ceux de mères fumant exclusivement au 1^{er} trimestre ou tout au long de la grossesse.

Malgré une trajectoire de croissance staturale légèrement différente, les trajectoires moyennes d'IMC des enfants de femmes non fumeuses et de femmes fumant exclusivement au 1^{er} trimestre étaient assez similaires. Aucune différence significative n'a été mise en évidence entre les IMC prédits des enfants des femmes fumant au 1^{er} trimestre exclusivement et ceux des femmes non fumeuses pendant leur grossesse (tableau 11).

En ce qui concerne, les enfants de femmes fumant tout au long de la grossesse, leur trajectoire moyenne d'IMC dépassait celle des enfants de femmes non fumeuses dès les 1^{ers} mois de vie. Le tabagisme maternel tout au long de la grossesse n'était pas significativement associé à l'âge au pic d'IMC, mais il était associé à une augmentation – bien que non significative – de l'IMC au moment du pic ($\beta=0,12$, *IC95%* [-0,05 ; 0,29]).

Bien que les différences soient non significatives, les vitesses de croissance de l'IMC des enfants de femmes fumeuses tout au long de la grossesse étaient supérieures à celle des enfants de non fumeuses sur toute la durée du suivi excepté à 4 et 5 ans où les vitesses de croissance étaient comparables. Ceci illustre une tendance à une vitesse de croissance plus élevée de la naissance au pic d'IMC et une décroissance de l'IMC plus lente à partir du pic d'IMC.

A deux ans, les enfants de femmes fumant tout au long de la grossesse avaient un IMC supérieur de 0,23 kg/m² (*IC95%* [0,07; 0,39]) à celui des enfants de non fumeuses et cette différence atteignait 0,24 kg/m² (*IC95%* [0,07; 0,41]) à 5 ans.

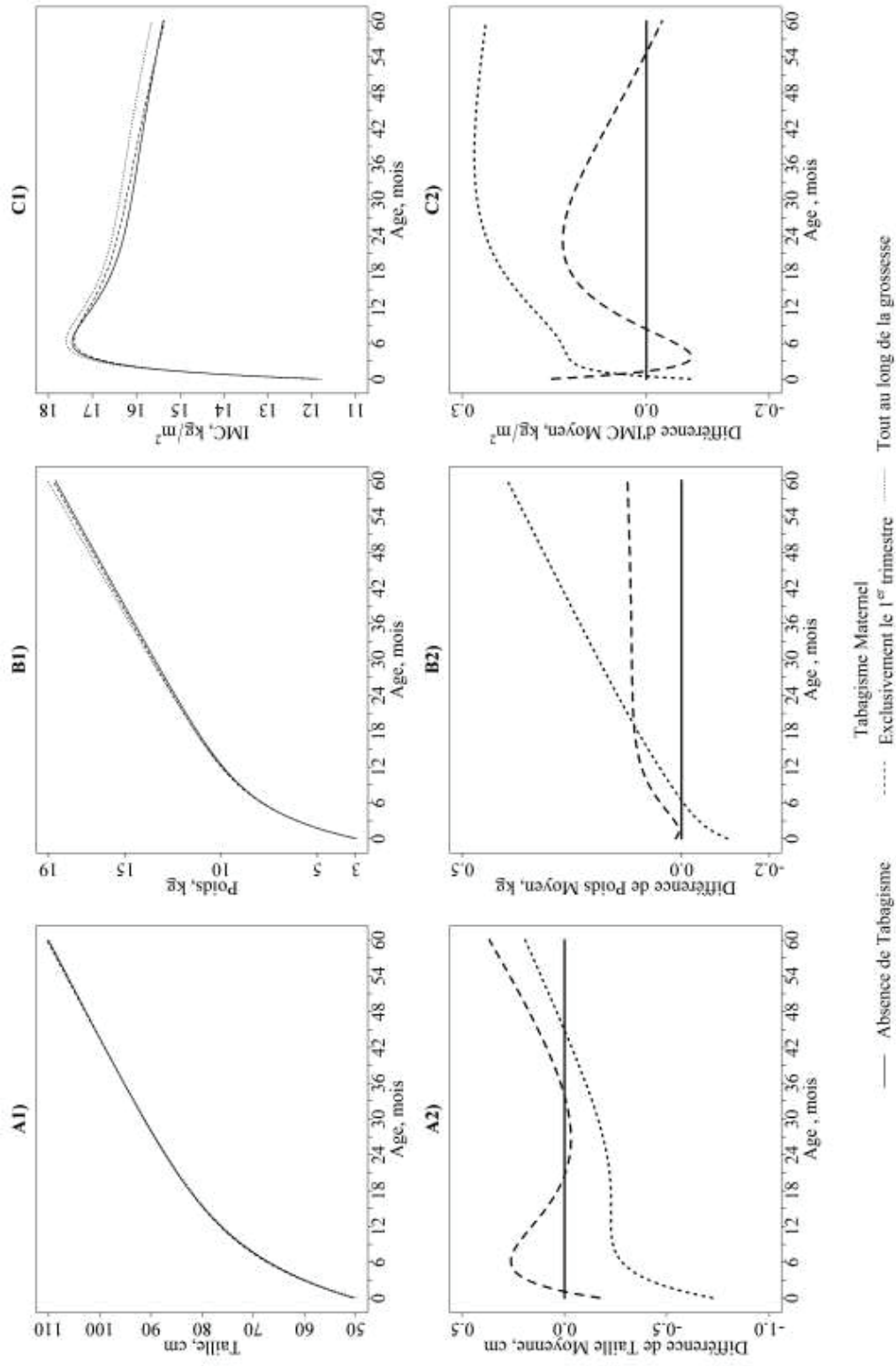


Figure 14 Trajectoires de croissance moyennes de taille, poids et IMC (A1 ; B1 ; C1) selon le statut tabagique de la mère et différences de taille, poids et IMC (A2 ; B2 ; C2) moyens entre les enfants exposés durant le 1^{er} trimestre exclusivement ou tout au long de la grossesse et les enfants non exposés

Tableau 10 : Différences moyennes des paramètres des modèles de croissances staturale et pondérale de la naissance à 5 ans pour chaque catégorie de tabagisme maternel et la catégorie de référence (absence de tabagisme pendant la grossesse), méthode en 1 étape^a, étude EDEN, France, 2003-2012

Variable	A		B		C		D		P global
	Taille/Poids Naissance/ t_0		Vitesse de croissance -Enfance-		Poussée de croissance -0-2 ans-		Degré de courbure -Premiers mois-		
	β_A	IC 95%	β_B	IC 95%	β_C	IC 95%	β_D	IC 95%	
Taille									
Exclusivement le 1 ^{er} trimestre	-0,003	[-0,012;0,005]	0,028	[-0,001;0,057]	-0,022	[-0,057;0,013]	0,074*	[0,010;0,137]	<0,001
Tout au long de la grossesse	-0,015*	[-0,021;-0,008]	0,022	[-0,001;0,046]	0,004	[-0,024;0,031]	0,038	[-0,011;0,087]	
Poids									
Exclusivement le 1 ^{er} trimestre	0,008	[-0,025;0,041]	-0,004	[-0,05;0,042]	0,016	[-0,036;0,068]	-0,023	[-0,099;0,054]	0,03
Tout au long de la grossesse	-0,037*	[-0,062;-0,012]	0,035	[-0,001;0,071]	0,008	[-0,032;0,048]	0,022	[-0,037;0,081]	

Abréviations: IMC, Indice de Masse Corporelle, IC, Intervalle de Confiance.

^aAjustées sur le niveau d'études, l'IMC et l'âge à l'accouchement de la mère; la durée d'allaitement; le centre de recrutement; le sexe de l'enfant.

* $P < 0,05$.

Tableau 11 : Différences moyennes des caractéristiques anthropométriques prédites de la naissance à 5 ans pour chaque catégorie de tabagisme maternel et la catégorie de référence (absence de tabagisme pendant la grossesse), méthode en 2 étapes^a, étude EDEN, France, 2003-2012

Variables selon le statut tabagique	Naissance		3 mois		6 mois		1 an		2 ans		4 ans		5 ans	
	β	IC 95%	β	IC 95%	β	IC 95%	β	IC 95%	β	IC 95%	β	IC 95%	β	IC 95%
Taille, cm														
Exclusivement le 1 ^{er} trimestre	-0,24	[-0,63 ; 0,15]	0,16	[-0,22 ; 0,53]	0,23	[-0,16 ; 0,62]	0,18	[-0,24 ; 0,60]	0,06	[-0,45 ; 0,57]	0,20	[-0,46 ; 0,87]	0,35	[-0,40 ; 1,11]
Tout au long de la grossesse	-0,63*	[-0,92 ; -0,33]	-0,40*	[-0,68 ; -0,12]	-0,30*	[-0,59 ; 0]	-0,23	[-0,55 ; 0,08]	-0,19	[-0,57 ; 0,19]	-0,06	[-0,56 ; 0,44]	0,02	[-0,55 ; 0,58]
Poids, g														
Exclusivement le 1 ^{er} trimestre	28	[-61 ; 117]	15	[-98 ; 129]	39	[-98 ; 176]	75	[-95 ; 245]	101	[-119 ; 321]	125	[-206 ; 455]	137	[-261 ; 535]
Tout au long de la grossesse	-120*	[-187 ; -54]	-53	[-138 ; 33]	-22	[-125 ; 81]	33	[-95 ; 161]	121	[-44 ; 287]	249*	[1 ; 498]	307*	[8 ; 606]
IMC, kg/m²														
Exclusivement le 1 ^{er} trimestre	0,25	[-0,01 ; 0,52]	-0,05	[-0,27 ; 0,17]	-0,04	[-0,27 ; 0,19]	0,03	[-0,19 ; 0,26]	0,1	[-0,12 ; 0,31]	0,04	[-0,17 ; 0,25]	0,00	[-0,22 ; 0,22]
Tout au long de la grossesse	-0,16	[-0,37 ; 0,04]	0,08	[-0,09 ; 0,24]	0,11	[-0,07 ; 0,28]	0,16	[-0,01 ; 0,33]	0,23*	[0,07 ; 0,39]	0,25*	[0,09 ; 0,41]	0,24*	[0,07 ; 0,41]
Vitesse de croissance staturale, cm/mois														
Exclusivement le 1 ^{er} trimestre	0,049*	[0,002 ; 0,096]	0,007	[-0,023 ; 0,036]	0,007	[-0,023 ; 0,036]	-0,02	[-0,039 ; 0,009]	-0,003	[-0,015 ; 0,009]	0,011	[0,000 ; 0,023]	0,013*	[0,001 ; 0,025]
Tout au long de la grossesse	0,049*	[0,013 ; 0,084]	0,022*	[0,000 ; 0,044]	0,005	[-0,013 ; 0,022]	0,005	[-0,013 ; 0,022]	0,004	[-0,005 ; 0,013]	0,006	[-0,002 ; 0,015]	0,006	[-0,003 ; 0,015]
Vitesse de croissance pondérale, kg/mois														
Exclusivement le 1 ^{er} trimestre	0,007	[-0,012 ; 0,026]	0,008	[-0,006 ; 0,022]	0,008	[-0,006 ; 0,022]	0,004	[-0,005 ; 0,013]	0,001	[-0,005 ; 0,008]	0,001	[-0,006 ; 0,008]	0,001	[-0,006 ; 0,008]
Tout au long de la grossesse	0,011	[-0,003 ; 0,026]	0,010	[-0,001 ; 0,020]	0,010	[-0,001 ; 0,020]	0,009*	[0,002 ; 0,016]	0,006*	[0,001 ; 0,011]	0,005	[0,000 ; 0,01]	0,005	[0,000 ; 0,01]
Vitesse de croissance de l'IMC, kg/m²/mois														
Exclusivement le 1 ^{er} trimestre	-0,068*	[-0,135 ; -0,001]	0,011	[-0,012 ; 0,035]	0,011	[-0,012 ; 0,035]	0,011	[-0,001 ; 0,023]	0,001	[-0,005 ; 0,006]	-0,003	[-0,007 ; 0,001]	-0,003	[-0,007 ; 0,000]
Tout au long de la grossesse	0,023	[-0,028 ; 0,073]	0,009	[-0,009 ; 0,026]	0,009	[-0,009 ; 0,026]	0,009	[0,000 ; 0,017]	0,003	[-0,001 ; 0,008]	0,000	[-0,003 ; 0,003]	-0,001	[-0,003 ; 0,002]
Age au pic d'IMC^b, jours														
Exclusivement le 1 ^{er} trimestre	7,31	[-11,93 ; 26,55]												
Tout au long de la grossesse	4,97	[-9,46 ; 19,39]												
IMC au pic d'IMC^b, kg/m²														
Exclusivement le 1 ^{er} trimestre	-0,04	[-0,27 ; 0,19]												
Tout au long de la grossesse	0,12	[-0,05 ; 0,29]												

Abbreviations: IMC, Indice de Masse Corporelle; IC, Intervalle de Confiance.

^aAjusté sur le niveau d'études, l'IMC et l'âge à l'accouchement de la mère; la durée d'allaitement; le centre de recrutement; le sexe de l'enfant.

^bAnalyses restreintes aux enfants ayant des données valides sur l'âge et l'IMC au pic d'IMC (pic d'IMC identifiable et identifié avant 732 jours). Le nombre d'enfants inclus dans les analyses était 1617.

* $p < 0,05$.

3.4.3.4 Relation dose-effet

Nos résultats suggèrent l'existence d'une relation dose-effet dans l'association entre le tabagisme maternel tout au long de la grossesse et la croissance. Les paramètres A du poids et de la taille étaient négativement associés avec le nombre de cigarettes fumées par jour. Des associations positives à la limite de la significativité sont mises en évidence entre le paramètre D de la taille et entre le paramètre B du poids et le nombre de cigarettes fumées par jour (tableau 12).

Tableau 12 : Différences moyennes des paramètres des modèles de croissances staturale et pondérale de la naissance à 5 ans pour une augmentation de 5 cigarettes de la consommation journalière de tabac aux 2^e et 3^e trimestres, Méthode en 1 étape^a, étude EDEN, France, 2003-2012

Variable	A		B		C		D	
	Taille à la naissance	Extrapolation du poids de naissance	Vitesse de croissance	-Enfance-	Poussée de croissance	-0-2 ans-	Degré de courbure	-Premiers mois-
	β_A	IC 95%	β_B	IC 95%	β_C	IC 95%	β_D	IC 95%
Taille	-0,009*	[-0,013;-0,005]	0,009	[-0,006;0,025]	-0,001	[-0,019;0,017]	0,027	[-0,005;0,059]
Poids	-0,026*	[-0,042;-0,011]	0,018	[-0,005;0,042]	0,008	[-0,017;0,034]	0,014	[-0,025;0,052]

Abréviations: IMC, Indice de Masse Corporelle, IC, Intervalle de Confiance.

^aAjustées sur le niveau d'études, l'IMC et l'âge à l'accouchement de la mère; la durée d'allaitement; le centre de recrutement; le sexe de l'enfant.

* $P < 0,05$.

Une augmentation de 5 cigarettes de la consommation journalière tout au long de la grossesse était associée à une diminution de la taille à la naissance de 0,39 cm (*IC95%* [-0,58 ; -0,20]); une diminution de la taille à 6 mois de 0,21 cm (*IC95%* [-0,40; -0,02]); une diminution du poids de naissance de 80 grammes (*IC95%* [-120 ; -40]); une augmentation, à la limite de la significativité, du poids moyen prédit à 4 ans de 155 grammes (*IC95%* [-1; 312]) et du poids moyen prédit à 5 ans de 188 grammes (*IC95%* [0 ; 376]) (tableau 13). Une augmentation de 5 cigarettes de la consommation journalière était associée à une augmentation significative de l'IMC à partir de 1 an (0,15 kg/m² [0,04 ; 0,26]) qui atteignait 0,19 kg/m² à 5 ans (*IC95%* [0,09; 0,30]).

Tableau 13 : Différences moyennes des caractéristiques anthropométriques prédites de la naissance à 5 ans pour une augmentation de 5 cigarettes de la consommation journalière de tabac aux 2^e et 3^e trimestres, méthode en 2 étapes ^a, étude EDEN, France, 2003-2012

Variables selon le statut tabagique	Naissance		3 mois		6 mois		1 an		2 ans		4 ans		5 ans	
	β	IC 95%	β	IC 95%	β	IC 95%	β	IC 95%	β	IC 95%	β	IC 95%	β	IC 95%
Taille, cm	-0,39*	[-0,58 ; -0,20]	-0,26*	[-0,44 ; -0,08]	-0,21*	[-0,40 ; -0,02]	-0,19	[-0,39 ; 0,01]	-0,21	[-0,46 ; 0,03]	-0,19	[-0,51 ; 0,13]	-0,16	[-0,52 ; 0,19]
Poids, g	-82*	[-125 ; -39]	-28	[-83 ; 27]	-2	[-67 ; 64]	37	[-45 ; 118]	86	[-19 ; 191]	155	[-1 ; 312]	188	[0 ; 376]
IMC, kg/m ²	-0,12	[-0,25 ; 0,01]	0,06	[-0,04 ; 0,17]	0,10	[-0,01 ; 0,21]	0,15*	[0,04 ; 0,26]	0,19*	[0,09 ; 0,30]	0,20*	[0,10 ; 0,30]	0,19*	[0,09 ; 0,30]
Vitesse de croissance staturale, cm/mois			0,028*	[0,005 ; 0,050]	0,01	[-0,004 ; 0,024]	-0,001	[-0,013 ; 0,010]	-0,001	[-0,007 ; 0,005]	0,002	[-0,004 ; 0,007]	0,002	[-0,003 ; 0,008]
Vitesse de croissance pondérale, kg/mois			0,010*	[0,001 ; 0,019]	0,008*	[0,001 ; 0,015]	0,005*	[0,001 ; 0,010]	0,003*	[0,000 ; 0,006]	0,003	[-0,001 ; 0,006]	0,003	[-0,001 ; 0,006]
Vitesse de croissance de l'IMC, kg/m ² /mois			0,024	[-0,008 ; 0,057]	0,010	[-0,001 ; 0,021]	0,006*	[0,001 ; 0,012]	0,002	[-0,001 ; 0,005]	0,000	[-0,002 ; 0,002]	-0,001	[-0,002 ; 0,001]
Age au pic d'IMC ^b , jours					6,79	[-2,57 ; 16,14]								
IMC au pic d'IMC ^c , kg/m ²					0,12*	[0,00 ; 0,23]								

Abbreviations: IMC, Indice de Masse Corporelle; IC, Intervalle de Confiance.

^aAjusté sur le niveau d'études, l'IMC et l'âge à l'accouchement de la mère; la durée d'allaitement; le centre de recrutement; le sexe de l'enfant.

^bAnalyses restreintes aux enfants ayant des données valides sur l'âge et l'IMC au pic d'IMC (pic d'IMC identifiable et identifié avant 732 jours). Le nombre d'enfants inclus dans les analyses était 1490.

*P<0,05.

3.4.3.5 Analyses de sensibilité

1. Exclusion des enfants prématurés

Les analyses de sensibilité, réalisées en excluant les 88 enfants prématurés (nés avant 37 SA) n'atténuent pas les mesures d'association entre le tabagisme maternel pendant la grossesse et la croissance mises en évidence pour l'ensemble des enfants (cf. Tableau 11). Les associations positives observées entre le tabagisme maternel tout au long de la grossesse et l'IMC entre 2 et 5 ans étaient renforcées par l'exclusion des enfants prématurés. Chez les enfants nés à terme, l'existence d'une association significative émergeait plus tôt dans l'enfance. L'IMC des enfants de femmes fumeuses tout au long de la grossesse dépassait celui des enfants de non fumeuses de 0,19 kg/m² à 1 an (IC95% [0,03; 0,36]) (tableau 14).

Tableau 14 : Différences moyennes des caractéristiques anthropométriques pour chaque catégorie de tabagisme maternel et la catégorie de référence (absence de tabagisme pendant la grossesse), méthode en 2 étapes^a, enfants nés à terme, étude EDEN, France, 2003-2012

Variables selon le statut tabagique	Naissance		3 mois		6 mois		1 an		2 ans		4 ans		5 ans	
	β	IC 95%	β	IC 95%	β	IC 95%	β	IC 95%	β	IC 95%	β	IC 95%	β	IC 95%
Taille, cm														
Exclusivement le 1 ^{er} trimestre	-0,43*	[-0,78 ; -0,08]	-0,05	[-0,39 ; 0,29]	0,06	[-0,31 ; 0,43]	0,06	[-0,36 ; 0,47]	-0,03	[-0,54 ; 0,48]	0,08	[-0,59 ; 0,75]	0,2	[-0,56 ; 0,96]
Tout au long de la grossesse	-0,69*	[-0,95 ; -0,42]	-0,48*	[-0,74 ; -0,23]	-0,35*	[-0,64 ; -0,07]	-0,26	[-0,57 ; 0,06]	-0,20	[-0,58 ; 0,19]	-0,07	[-0,58 ; 0,43]	-0,01	[-0,58 ; 0,57]
Poids, g														
Exclusivement le 1 ^{er} trimestre	-17	[-95 ; 60]	-31	[-139 ; 76]	-8	[-142 ; 127]	30	[-141 ; 200]	53	[-168 ; 274]	58	[-274 ; 391]	59	[-342 ; 460]
Tout au long de la grossesse	-140*	[-199 ; -81]	-58	[-139 ; 24]	-17	[-119 ; 85]	48	[-81 ; 177]	140	[-27 ; 308]	266*	[14 ; 518]	322*	[19 ; 626]
IMC, kg/m²														
Exclusivement le 1 ^{er} trimestre	0,15	[-0,08 ; 0,39]	-0,08	[-0,30 ; 0,14]	-0,06	[-0,29 ; 0,17]	0,01	[-0,21 ; 0,23]	0,07	[-0,15 ; 0,28]	0,01	[-0,20 ; 0,23]	-0,03	[-0,25 ; 0,20]
Tout au long de la grossesse	-0,20*	[-0,38 ; -0,02]	0,11	[-0,06 ; 0,28]	0,15	[-0,03 ; 0,32]	0,19*	[0,03 ; 0,36]	0,25*	[0,09 ; 0,41]	0,27*	[0,11 ; 0,43]	0,26*	[0,09 ; 0,43]
Vitesse de croissance staturale, cm/mois														
Exclusivement le 1 ^{er} trimestre	0,061*	[0,015 ; 0,106]	0,017	[-0,010 ; 0,045]	-0,009	[-0,032 ; 0,015]	-0,003	[-0,015 ; 0,009]	0,009	[-0,002 ; 0,021]	0,009	[-0,002 ; 0,021]	0,011	[-0,001 ; 0,023]
Tout au long de la grossesse	0,059*	[0,025 ; 0,094]	0,029*	[0,008 ; 0,05]	0,008	[-0,006 ; 0,022]	0,008	[-0,005 ; 0,014]	0,005	[-0,005 ; 0,014]	0,006	[-0,003 ; 0,014]	0,005	[-0,004 ; 0,015]
Vitesse de croissance pondérale, kg/mois														
Exclusivement le 1 ^{er} trimestre	0,006	[-0,012 ; 0,025]	0,006	[-0,006 ; 0,022]	0,008	[-0,006 ; 0,022]	0,004	[-0,005 ; 0,013]	0,001	[-0,006 ; 0,007]	0,000	[-0,007 ; 0,007]	0,000	[-0,007 ; 0,007]
Tout au long de la grossesse	0,015*	[0,001 ; 0,029]	0,012*	[0,002 ; 0,023]	0,010*	[0,003 ; 0,016]	0,006*	[0,001 ; 0,011]	0,005	[0 ; 0,01]	0,005	[0 ; 0,01]	0,005	[-0,001 ; 0,01]
Vitesse de croissance de l'IMC, kg/m²/mois														
Exclusivement le 1 ^{er} trimestre	-0,059	[-0,127 ; 0,008]	0,011	[-0,011 ; 0,034]	0,010	[-0,001 ; 0,022]	0,001	[-0,005 ; 0,007]	0,001	[-0,005 ; 0,007]	-0,003	[-0,007 ; 0,001]	-0,003	[-0,006 ; 0,000]
Tout au long de la grossesse	0,030	[-0,021 ; 0,081]	0,009	[-0,008 ; 0,026]	0,008	[-0,001 ; 0,016]	0,003	[-0,002 ; 0,007]	0,000	[-0,002 ; 0,007]	0,000	[-0,003 ; 0,003]	-0,001	[-0,003 ; 0,002]
Age au pic d'IMC^b, jours														
Exclusivement le 1 ^{er} trimestre					6,49	[-12,82 ; 25,79]								
Tout au long de la grossesse					4,22	[-10,37 ; 18,81]								
IMC au pic d'IMC^b, kg/m²														
Exclusivement le 1 ^{er} trimestre					-0,06	[-0,30 ; 0,17]								
Tout au long de la grossesse					0,15	[-0,02 ; 0,33]								

Abbreviations: IMC, Indice de Masse Corporelle; IC, Intervalle de Confiance.

^aAjusté sur le niveau d'études, l'IMC et l'âge à l'accouchement de la mère; la durée d'allaitement; le centre de recrutement; le sexe de l'enfant.

^bAnalyses restreintes aux enfants ayant des données valides sur l'âge et l'IMC au pic d'IMC (pic d'IMC identifiable et identifié avant 732 jours). Le nombre d'enfants inclus dans les analyses était 1532.

*P<0,05.

2. Prise en compte de la confusion résiduelle liée aux modes de vie

- Exposition au tabagisme durant la période postnatale

Les poids, taille et IMC prédits aux différents âges d'intérêt des enfants non exposés au tabagisme durant la période postnatale (et non exposés en prénatal) ne différaient pas significativement de ceux des enfants exposés au tabagisme maternel et/ou paternel durant la période postnatale exclusivement (tableau 15).

Tableau 15 : Différences moyennes des caractéristiques anthropométriques de 3 mois à 5 ans entre les enfants exposés au tabagisme durant la période postnatale exclusivement versus les enfants non exposés, Méthode en 2 étapes ^a, étude EDEN, France, 2003-2012

Variables selon le statut tabagique	3 mois		6 mois		1 an		2 ans		4 ans		5 ans	
	β	IC 95%	β	IC 95%	β	IC 95%	β	IC 95%	β	IC 95%	β	IC 95%
Taille, cm												
Exposition Postnatale au Tabagisme	-0,07	[-0,35 ; 0,21]	-0,13	[-0,42 ; 0,16]	-0,18	[-0,49 ; 0,13]	-0,21	[-0,58 ; 0,17]	-0,25	[-0,73 ; 0,24]	-0,27	[-0,82 ; 0,28]
Poids, g												
Exposition Postnatale au Tabagisme	-8	[-92 ; 77]	-23	[-124 ; 79]	-45	[-170 ; 80]	-56	[-216 ; 104]	-39	[-274 ; 197]	-26	[-308 ; 256]
IMC, kg/m ²												
Exposition Postnatale au Tabagisme	0,03	[-0,14 ; 0,19]	0,02	[-0,15 ; 0,19]	0,00	[-0,16 ; 0,17]	0,00	[-0,15 ; 0,16]	0,04	[-0,11 ; 0,18]	0,05	[-0,10 ; 0,20]

Abbréviations: IMC, Indice de Masse Corporelle; IC, Intervalle de Confiance.

^aAjusté sur le niveau d'études, l'IMC et l'âge à l'accouchement de la mère; la durée d'allaitement; le centre de recrutement; le sexe de l'enfant.

* $P < 0,05$.

- Ajustement supplémentaire sur les profils de pratiques alimentaires

Les associations entre le tabagisme de la mère pendant la grossesse et les caractéristiques de croissance des enfants étaient légèrement modifiées par l'ajustement supplémentaire sur les profils de pratiques alimentaires dans le sous-échantillon de 971 enfants pour lesquels cette donnée était disponible. En ce qui concerne l'IMC, après cet ajustement, les enfants de mères fumant tout au long de la grossesse avaient un IMC moyen supérieur de 0,25 ($P=0,02$); 0,24 ($P=0,03$); 0,21($P=0,07$) et 0,17 kg/m² ($P=0,15$) à 2, 3, 4 et 5 ans respectivement.

3. Ajustement supplémentaire sur la taille des parents

L'ajustement sur la taille des parents atténuait légèrement les associations entre les vitesses de croissance pondérale après 1 an et le tabagisme tout au long de la grossesse mises en évidence dans l'analyse principale (cf. Tableau 11). De ce fait, les différences de poids observées entre les enfants exposés tout au long de la grossesse et les enfants non exposés au tabagisme maternel étaient légèrement atténuées ainsi que les différences d'IMC. Les différences d'IMC entre ces deux groupes n'étaient significatives ou à la limite de la significativité qu'après 4 ans. Bien que des différences significatives soient observées plus tardivement que dans l'échantillon complet, ce qui peut être expliqué par une diminution du nombre de sujets inclus dans les analyses, l'ampleur des différences observées sur l'ensemble de la période d'étude était proche (Tableau 16).

Tableau 16 : Différences moyennes des caractéristiques anthropométriques pour chaque catégorie de tabagisme maternel et la catégorie de référence (absence de tabagisme pendant la grossesse), méthode en 2 étapes^a, chez les enfants dont la taille des parents est connue, étude EDEN, France, 2003-2012

Variables selon le statut tabagique	Naissance		3 mois		6 mois		1 an		2 ans		4 ans		5 ans	
	β	IC 95%	β	IC 95%	β	IC 95%	β	IC 95%	β	IC 95%	β	IC 95%	β	IC 95%
Taille, cm														
Exclusivement le 1 ^{er} trimestre	-0,27	[-0,7 ; 0,15]	0,11	[-0,28 ; 0,49]	0,18	[-0,22 ; 0,58]	0,13	[-0,28 ; 0,55]	0,02	[-0,47 ; 0,52]	0,14	[-0,50 ; 0,79]	0,27	[-0,47 ; 1,01]
Tout au long de la grossesse	-0,57*	[-0,9 ; -0,23]	-0,39*	[-0,70 ; -0,08]	-0,29	[-0,61 ; 0,03]	-0,19	[-0,52 ; 0,15]	-0,12	[-0,52 ; 0,28]	-0,01	[-0,53 ; 0,51]	0,05	[-0,54 ; 0,64]
Poids, g														
Exclusivement le 1 ^{er} trimestre	20	[-77 ; 117]	-15	[-138 ; 109]	12	[-136 ; 160]	63	[-120 ; 245]	104	[-128 ; 336]	133	[-206 ; 472]	146	[-259 ; 552]
Tout au long de la grossesse	-108*	[-186 ; -31]	-52	[-151 ; 47]	-16	[-135 ; 102]	38	[-108 ; 184]	104	[-81 ; 290]	183	[-88 ; 454]	217	[-107 ; 542]
IMC, kg/m ²														
Exclusivement le 1 ^{er} trimestre	0,23	[-0,07 ; 0,52]	-0,11	[-0,36 ; 0,13]	-0,08	[-0,33 ; 0,17]	0,03	[-0,22 ; 0,28]	0,11	[-0,12 ; 0,34]	0,06	[-0,16 ; 0,29]	0,02	[-0,21 ; 0,26]
Tout au long de la grossesse	-0,13	[-0,36 ; 0,11]	0,08	[-0,12 ; 0,28]	0,12	[-0,08 ; 0,32]	0,15	[-0,05 ; 0,35]	0,18	[-0,01 ; 0,37]	0,18	[0,00 ; 0,36]	0,17	[-0,02 ; 0,35]
Vitesse de croissance staturale, cm/mois														
Exclusivement le 1 ^{er} trimestre	0,046	[-0,004 ; 0,096]	0,006	[-0,024 ; 0,037]	0,006	[-0,024 ; 0,037]	-0,013	[-0,039 ; 0,012]	-0,003	[-0,016 ; 0,010]	0,010	[-0,002 ; 0,022]	0,011	[-0,001 ; 0,024]
Tout au long de la grossesse	0,044*	[0,004 ; 0,084]	0,027*	[0,002 ; 0,051]	0,027*	[0,002 ; 0,051]	0,010	[-0,011 ; 0,030]	0,004	[-0,006 ; 0,014]	0,005	[-0,005 ; 0,015]	0,005	[-0,005 ; 0,016]
Vitesse de croissance pondérale, kg/mois														
Exclusivement le 1 ^{er} trimestre	0,006	[-0,015 ; 0,027]	0,010	[-0,005 ; 0,025]	0,010	[-0,005 ; 0,025]	0,006	[-0,004 ; 0,016]	0,002	[-0,005 ; 0,008]	0,001	[-0,006 ; 0,008]	0,001	[-0,006 ; 0,008]
Tout au long de la grossesse	0,013	[-0,003 ; 0,030]	0,011	[-0,002 ; 0,023]	0,011	[-0,002 ; 0,023]	0,008	[0,000 ; 0,015]	0,004	[-0,001 ; 0,009]	0,003	[-0,003 ; 0,008]	0,003	[-0,003 ; 0,008]
Vitesse de croissance de l'IMC, kg/m ² /mois														
Exclusivement le 1 ^{er} trimestre	-0,071	[-0,144 ; 0,002]	0,019	[-0,006 ; 0,045]	0,019	[-0,006 ; 0,045]	0,015*	[0,002 ; 0,027]	0,001	[-0,005 ; 0,007]	-0,003	[-0,007 ; 0,001]	-0,003	[-0,007 ; 0,000]
Tout au long de la grossesse	0,030	[-0,029 ; 0,088]	0,008	[-0,012 ; 0,028]	0,008	[-0,012 ; 0,028]	0,004	[-0,006 ; 0,014]	0,001	[-0,004 ; 0,006]	-0,001	[-0,004 ; 0,003]	-0,001	[-0,004 ; 0,002]
Age au pic d'IMC ^b , jours														
Exclusivement le 1 ^{er} trimestre					13,11	[-8,28 ; 34,50]								
Tout au long de la grossesse					4,29	[-12,58 ; 21,16]								
IMC au pic d'IMC ^b , kg/m ²														
Exclusivement le 1 ^{er} trimestre					-0,06	[-0,32 ; 0,19]								
Tout au long de la grossesse					0,10	[-0,10 ; 0,30]								

Abbreviations: IMC, Indice de Masse Corporelle; IC, Intervalle de Confiance.

^aAjusté sur le niveau d'études, l'IMC et l'âge à l'accouchement de la mère; la durée d'allaitement; le centre de recrutement; le sexe de l'enfant.

^bAnalyses restreintes aux enfants ayant des données valides sur l'âge et l'IMC au pic d'IMC (pic d'IMC identifiable et identifié avant 732 jours). Le nombre d'enfants inclus dans les analyses était 1296.

* $P < 0,05$.

4. Comparaison à d'autres méthodes de modélisation de l'IMC

Polynômes fractionnaires

Pour les filles, le modèle minimisant le BIC était un polynôme fractionnaire de degré 7 ayant pour combinaison de puissances $-2/-1/-0,5/0/0,5/2/3$ et pour les garçons, un polynôme de degré 6 avec pour combinaison de puissances $-2/-1/-0,5/1/2/3$ (tableau 17).

Tableau 17 : Modélisation de l'IMC, Polynômes Fractionnaires

Nb de degrés	Filles		Garçons	
	Puissances d'âge sélectionnées	BIC	Puissances d'âge sélectionnées	BIC
8	$-2/-1/-0,5/0/0,5/1/2/3$	31971	$-2/-1/-0,5/0/0,5/1/2/3$	34976
7	$-2/-1/-0,5/0/0,5/2/3$	31936	$-2/-1/-0,5/0/0,5/2/3$	34898
6	$-2/-1/-0,5/1/2/3$	31965	$-2/-1/0,5/1/2/3$	34897
5	$-1/0,5/1/2/3$	31974	$-1/0,5/1/2/3$	34932
4	$0,5/1/2/3$	32315	$0,5/1/2/3$	31515
3	$0/1/2$	33243	$0/0,5/2$	35992
2	$-0,5/0,5$	35126	$-0,5/0,5$	37802

Néanmoins, après visualisation des trajectoires obtenues avec les différents polynômes sélectionnés, les polynômes fractionnaires de degrés 2 ont été retenus, les polynômes de degré supérieur aboutissant à des trajectoires d'IMC que nous avons jugées sur-paramétrées, surtout aux âges extrêmes. Ces polynômes de degrés supérieurs semblaient en effet ne pas s'adapter aux trajectoires de croissance individuelles en dehors de leur fenêtre d'observations. La figure 15 illustre les trajectoires obtenues pour 4 garçons avec le polynôme de degré 6 sélectionné et le polynôme de degré 2 retenu et celles obtenues par notre méthode indirecte (à partir des modélisations de la croissance pondérale et staturale séparément chez les filles et les garçons sans inclusion de covariables).

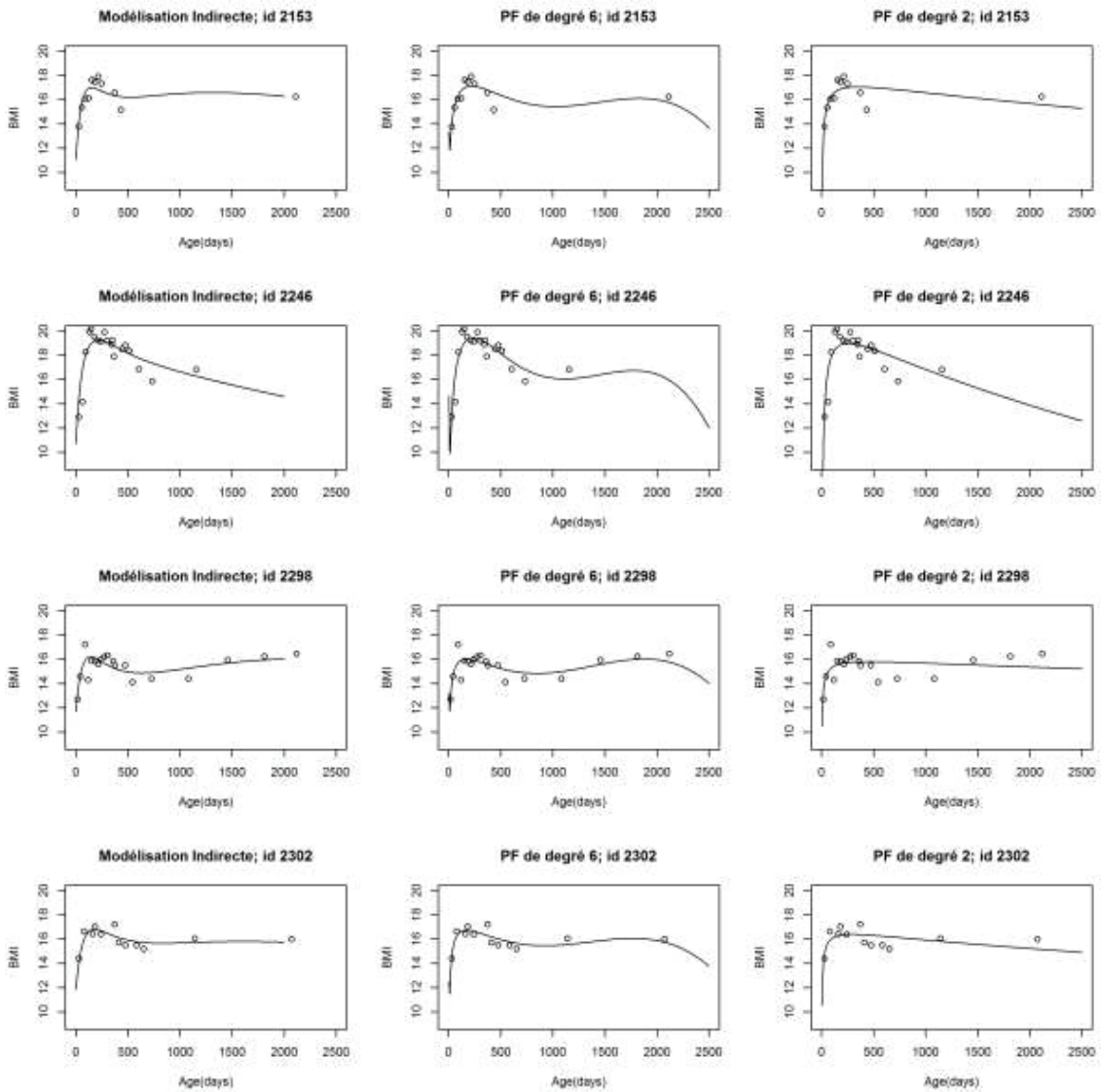


Figure 15 : Trajectoires d'IMC obtenues pour 4 garçons avec les polynômes de degré 2 et 6 et la méthode de modélisation indirecte

Considérant les polynômes de degré 2 (méthode habituelle des polynômes fractionnaires), les 8 modèles possibles considérant la répétition de deux puissances identiques ont été testés. Les modèles précédemment identifiés (puissances -0,5 et 0,5) venaient juste après les modèles incluant une répétition de puissance 0 (log) de l'âge en terme de minimisation du BIC que ce soit pour les filles ou les garçons (tableau 18).

Tableau 18 : Modélisation de l'IMC, Polynômes fractionnaires de degré 2

Nb de degrés	Filles		Garçons	
	Puissances d'âge sélectionnées	BIC	Puissances d'âge sélectionnées	BIC
2	-0,5/0,5	35126	-0,5/0,5	37802
2	0/0	34764	0/0	37419

Le modèle final retenu était donc :

$$E(IMC) = b_0 + b_1 * \log(Age) + b_2 * \log(Age) * \log(Age)$$

Et les trajectoires des 4 garçons représentées précédemment sont représentées dans la figure 16.

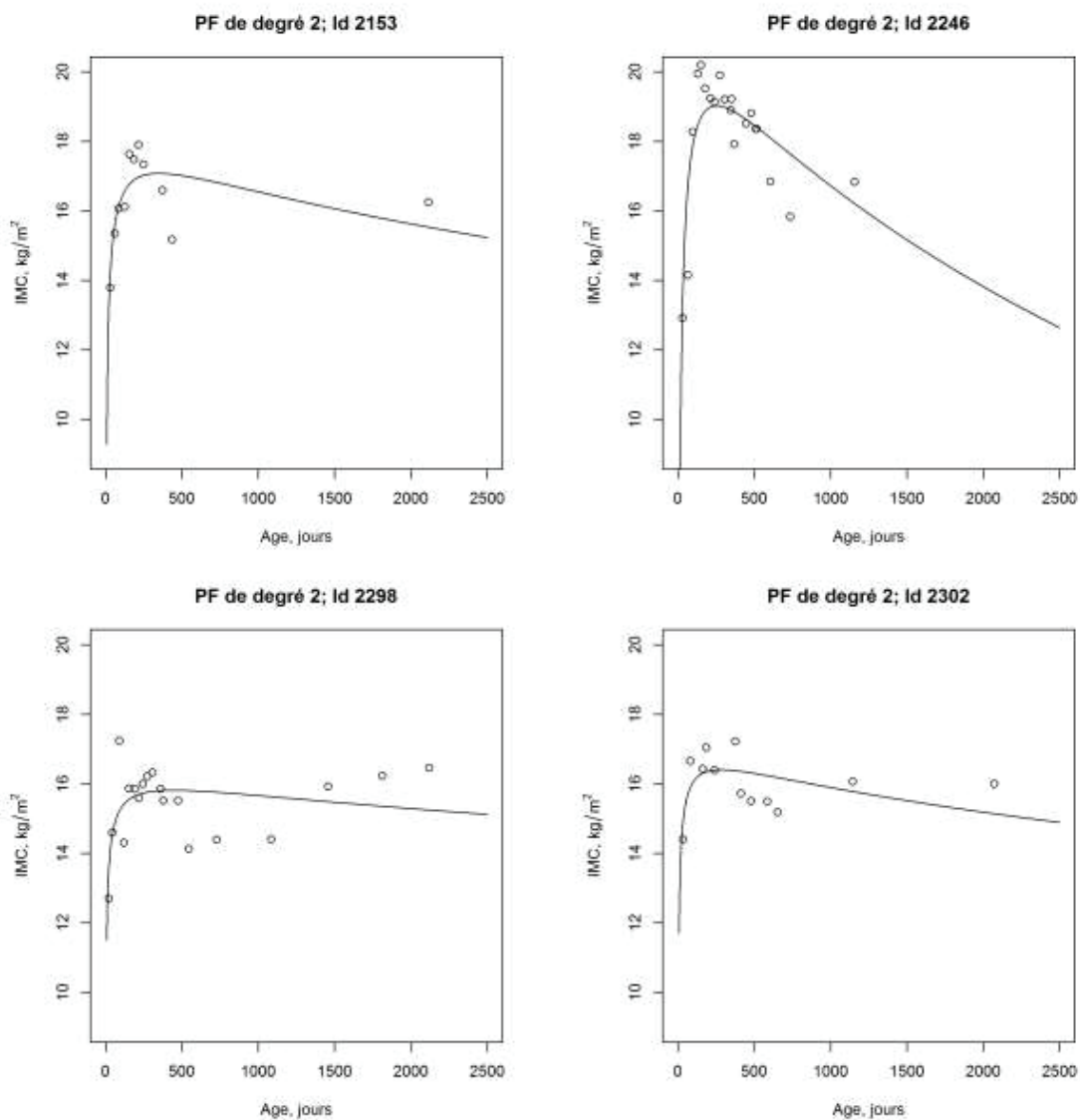


Figure 16 : Trajectoires d'IMC obtenues pour 4 garçons avec le polynôme de degré 2 retenu, méthode de Royston et Altman

Cette figure montre que ce polynôme de degré 2 modélise les trajectoires de croissance de manière imparfaite mais le choix de deux puissances n'autorise qu'un point d'inflexion et de ce fait permet d'obtenir des profils de trajectoires d'IMC conformes aux profils habituels.

Splines

Le modèle minimisant l'AIC et le BIC était le modèle incluant 2 nœuds, au 10^e jour et au 410^e jour (13 mois ½) que ce soit pour les filles ou les garçons.

Pour chaque nombre de nœuds, le tableau 19 décrit l'emplacement des nœuds minimisant l'AIC et le BIC pour la modélisation de l'IMC des filles. On peut noter que, pour un nombre de nœuds supérieur à 2, les nœuds supplémentaires qui permettaient l'amélioration de l'ajustement étaient placés en fin de distribution.

Tableau 19 : Choix de l'emplacement des nœuds, Splines, Modélisation de l'IMC chez les filles

Nb de nœuds	AIC	BIC	Emplacement des nœuds (Age en jours)
2	14284	24116	10/410
3	15039	26475	10/210/410
4	15565	27790	10/210/410/2410
5	15572	27806	10/210/410/2310/2410
6	15585	27842	10/210/410/2110/2210/2310

Les AIC et BIC des modèles testés avec les nœuds placés aux percentiles de la distribution ne permettaient pas d'atteindre des AIC/BIC plus faibles que les modèles précédemment sélectionnés.

La figure 17 illustre la trajectoire moyenne et la Figure 18 les trajectoires des 4 individus illustrées précédemment ajustées par les splines chez les filles et les garçons.

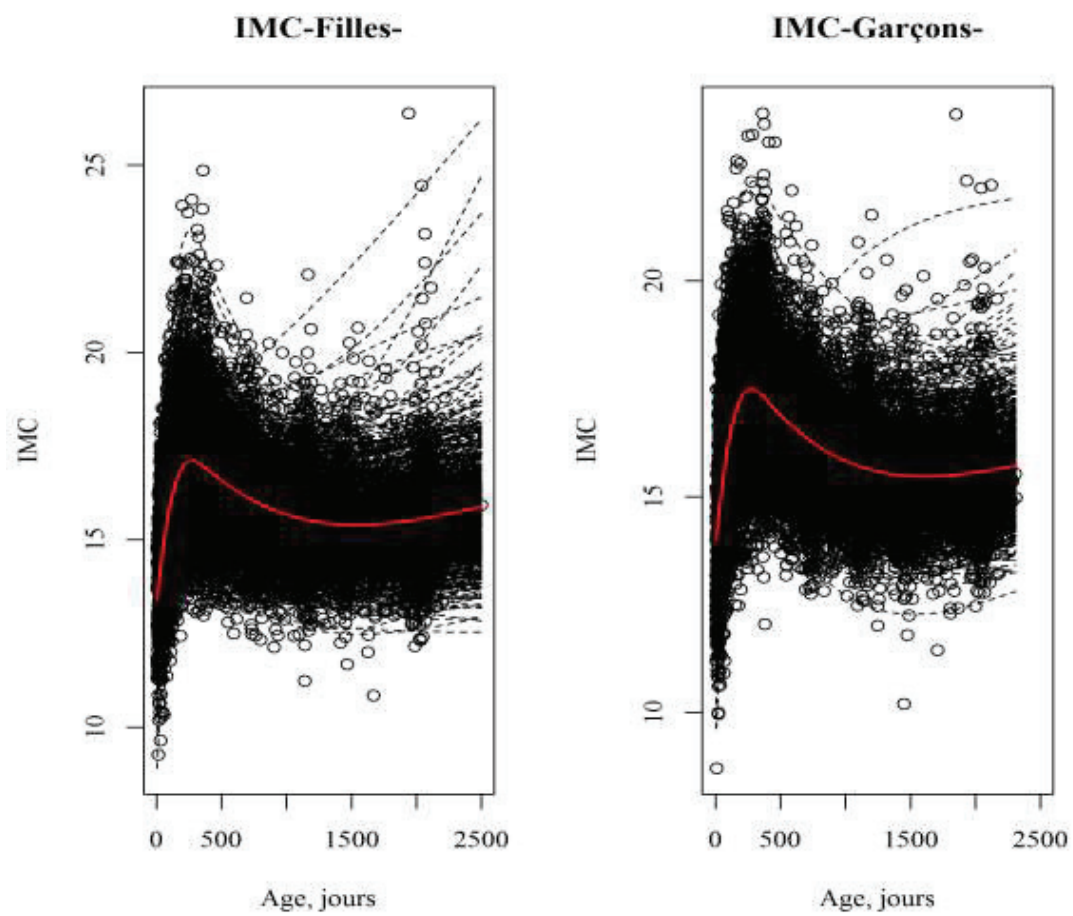


Figure 17 : Modélisation de l'IMC des filles et des garçons, splines

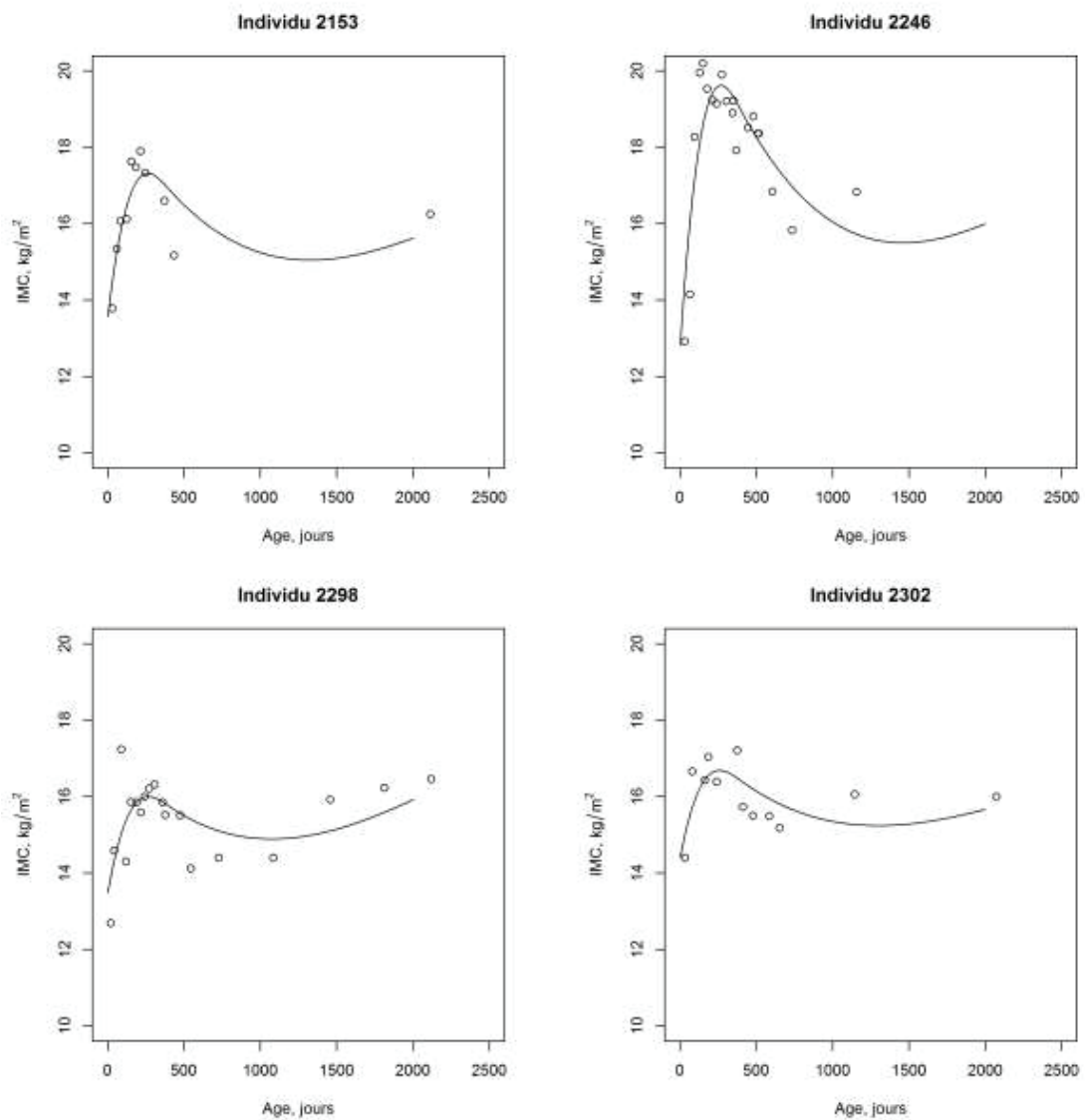


Figure 18 : Trajectoires d'IMC obtenues pour 4 garçons avec le modèle splines retenu

L'inspection des trajectoires illustrées dans la Figure 18 montre que ce modèle par splines semble mieux montrer le début de remontée de l'IMC que ne le fait la modélisation indirecte.

La distribution des résidus d'IMC obtenue avec les différentes approches de modélisation (figure 19) montre que la qualité d'ajustement des polynômes fractionnaires ne semblait pas supérieure à celle de la méthode indirecte.

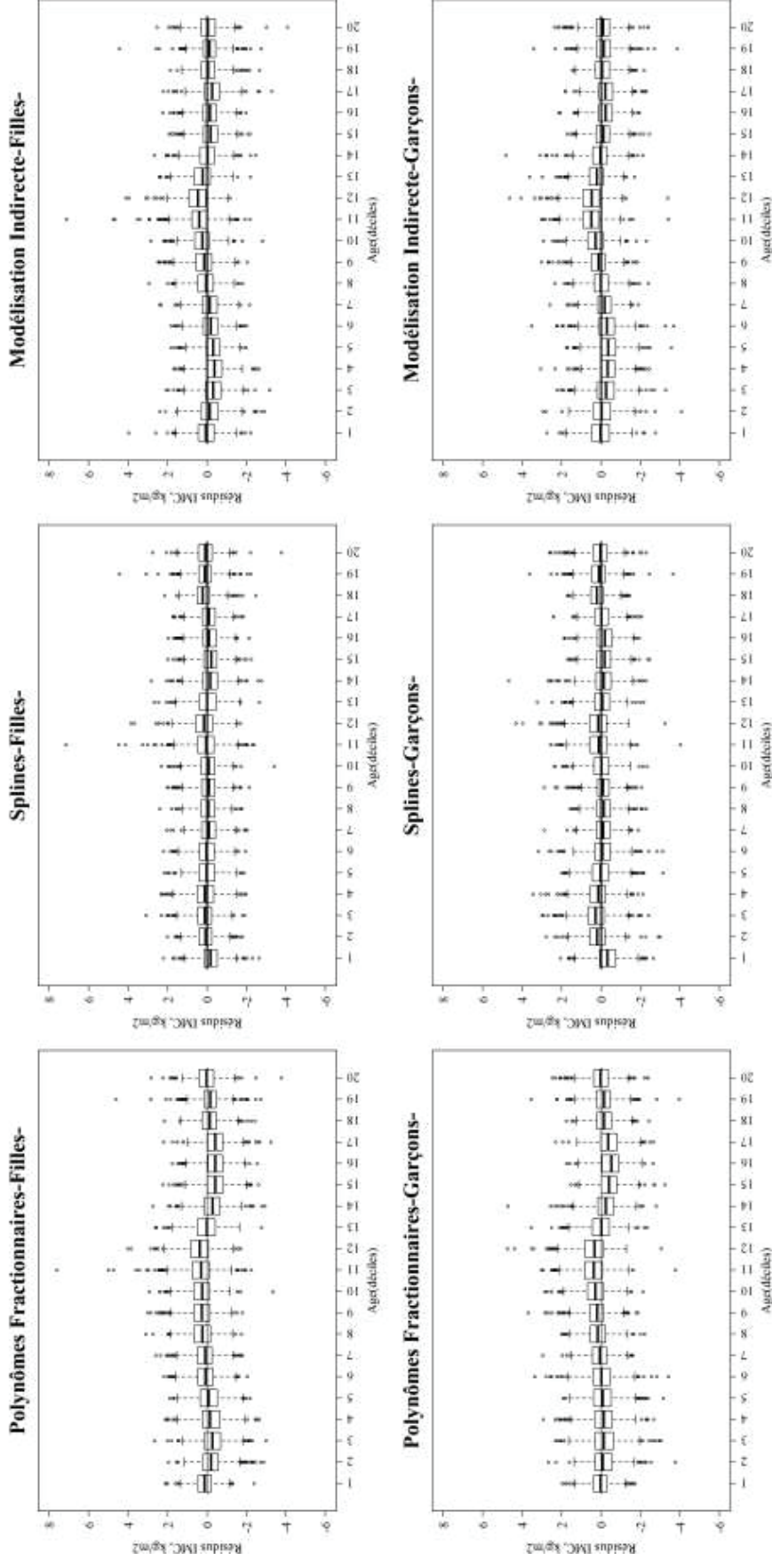


Figure 19 : Distributions des résidus d'IMC, polynômes fractionnaires, splines, méthode indirecte

L'absence de tendance (sur-estimation, sous-estimation) des données observées pour les splines laisse penser que leur qualité d'ajustement est supérieure à celle de la méthode indirecte. La moindre qualité d'ajustement de la méthode indirecte résulte principalement d'une modélisation imparfaite de la taille (figure 20). La distribution des résidus de taille montrait que le modèle sous-estimait les données dans les premiers mois de la fenêtre d'étude et par la suite les surestimait légèrement. En revanche, les données de poids étaient particulièrement bien prédites par le modèle de Jenss-Bayley.

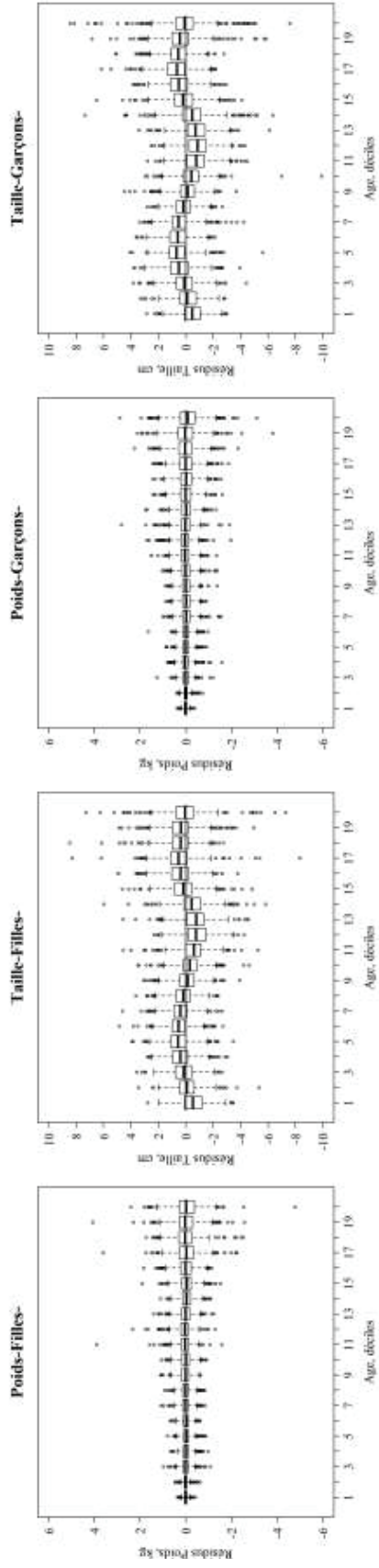


Figure 20 : Distribution des résidus de poids et de taille, méthode indirecte

Malgré une qualité d'ajustement non optimale pour la taille, les trajectoires individuelles de croissance du poids, de la taille et de l'IMC ne semblaient pas montrer de défaut majeur des modèles (figure 21 ; figure 22 ; figure 23).

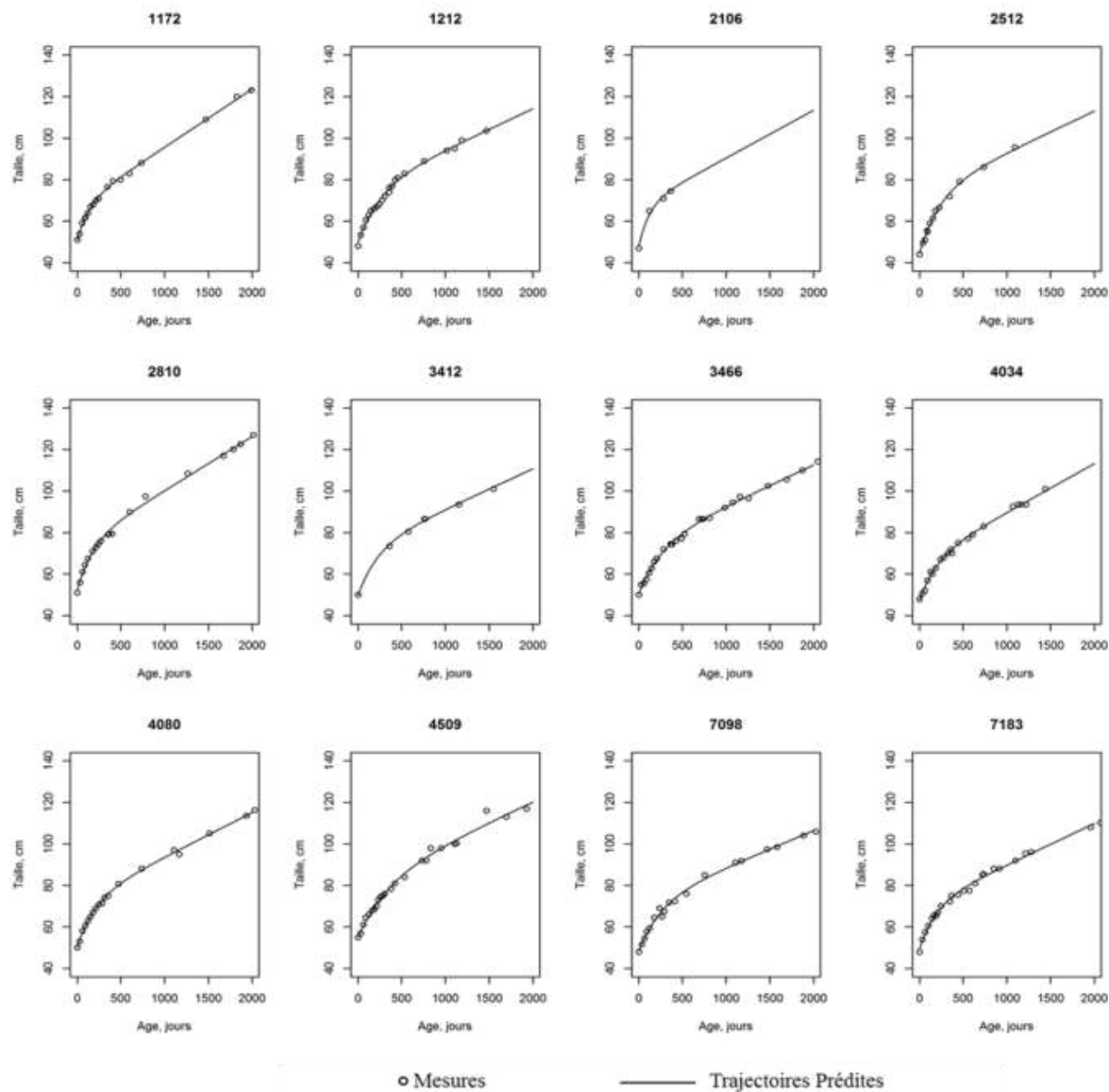


Figure 21 : Mesures de taille et trajectoires prédites de taille pour 12 individus sélectionnés aléatoirement

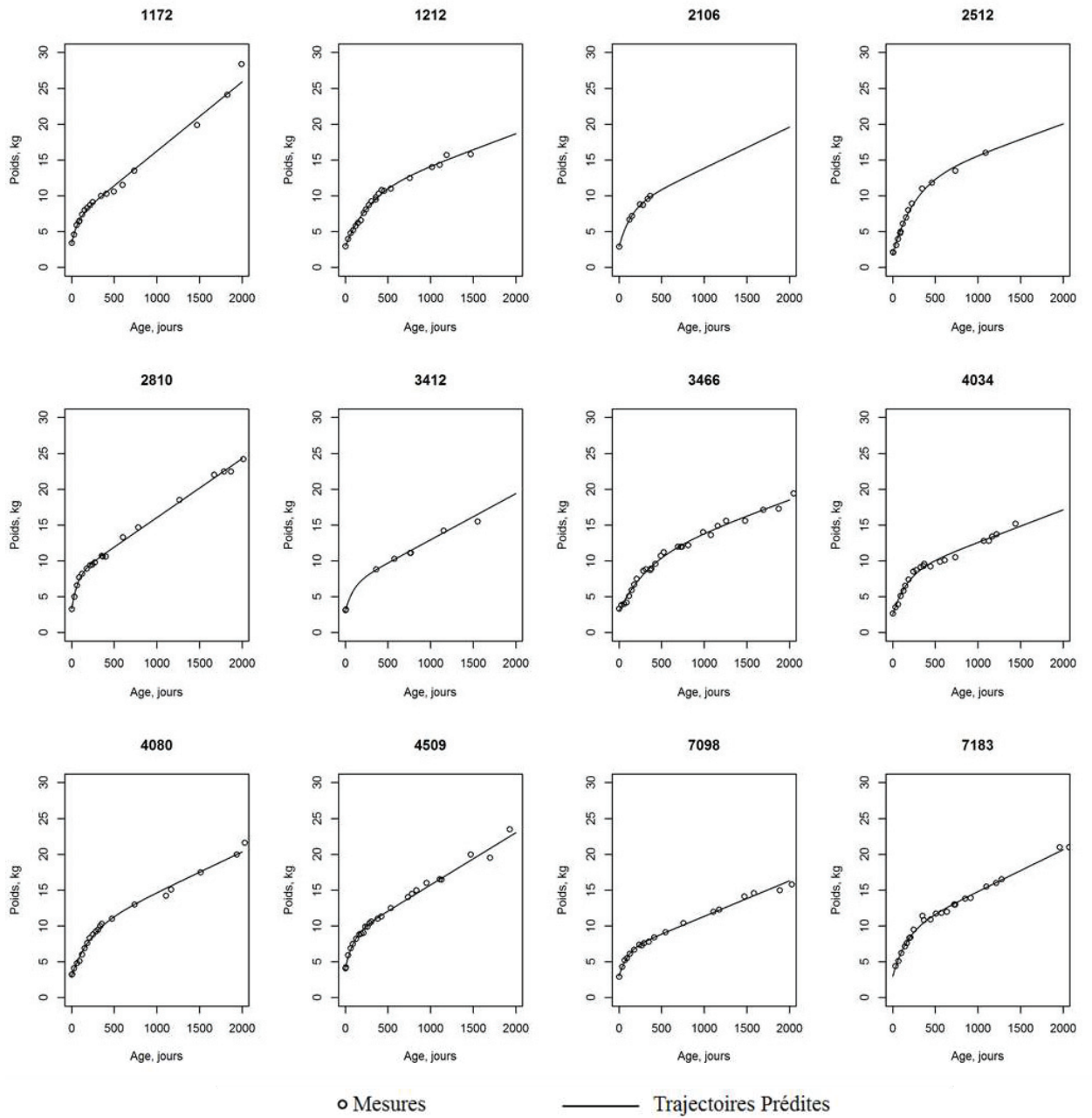


Figure 22 : Mesures de poids et trajectoires prédites de poids pour 12 individus sélectionnés aléatoirement

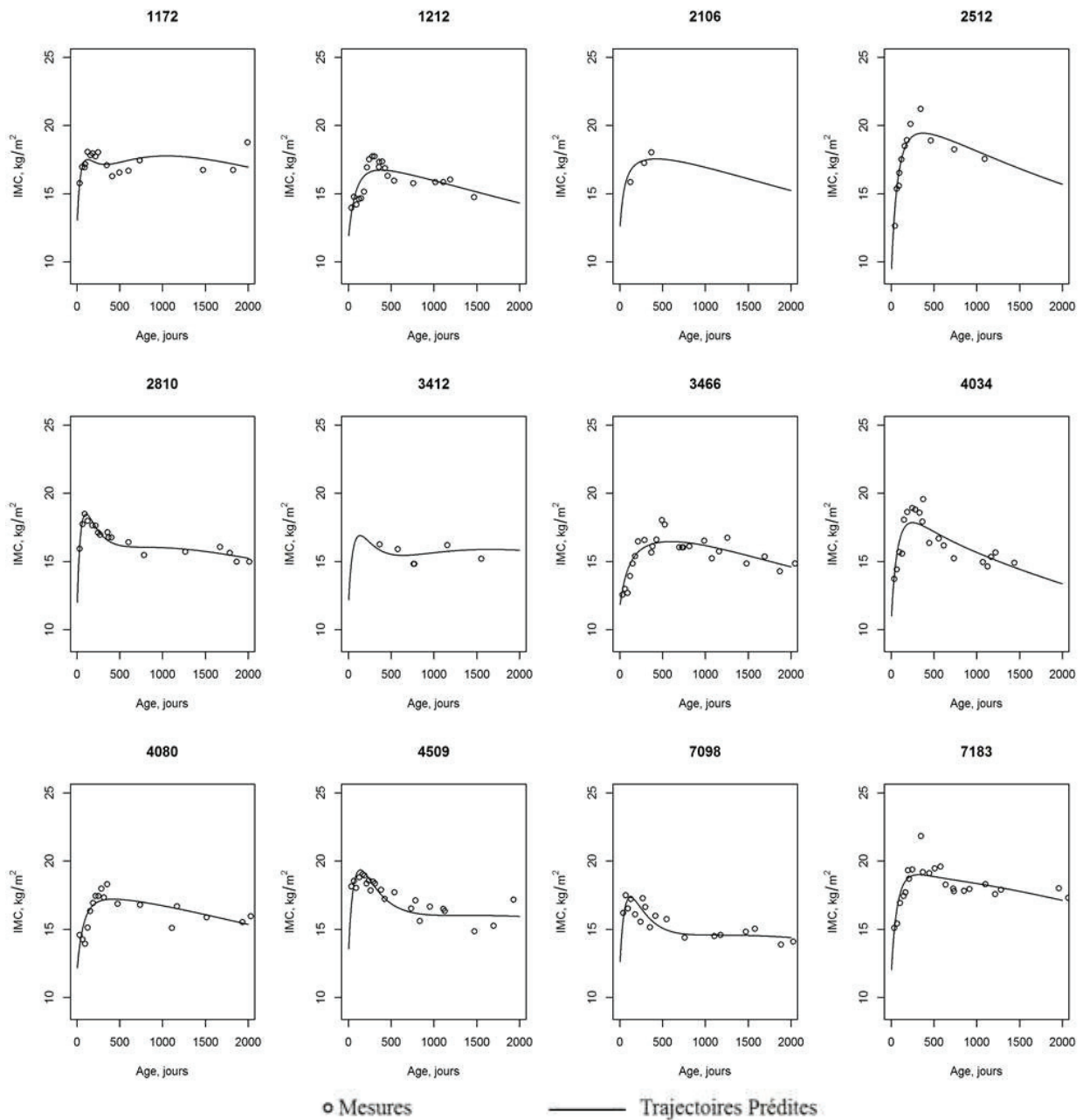


Figure 23 : Mesures d'IMC et trajectoires prédites d'IMC pour 12 individus sélectionnés aléatoirement

Malgré les différences de qualité d'ajustement qui pouvaient être constatées entre les différentes approches de modélisation de l'IMC, les mesures d'associations entre le tabagisme maternel pendant la grossesse et les IMC prédits entre 3 mois et 5 ans étaient très semblables (Tableau 20).

Tableau 20 : Différences moyennes des IMC prédits de 3 mois à 5 ans pour chaque catégorie de tabagisme maternel et la catégorie de référence (absence de tabagisme maternel), Méthode en 2 étapes (polynômes fractionnaires, splines, modélisation indirecte)^a, étude EDEN, France, 2003-2012

Variables selon le statut tabagique et la méthode de modélisation	3 mois			6 mois			1 an			2 ans			4 ans			5 ans			
	β	IC 95%	ÉT	β	IC 95%	ÉT	β	IC 95%	ÉT	β	IC 95%	ÉT	β	IC 95%	ÉT	β	IC 95%	ÉT	
IMC, kg/m ²																			
Exclusivement le 1 ^{er} trimestre																			
Modélisation Indirecte	-0,05	[-0,27 ; 0,17]	0,11	-0,04	[-0,27 ; 0,19]	0,12	0,03	[-0,19 ; 0,26]	0,11	0,10	[-0,12 ; 0,31]	0,11	0,04	[-0,17 ; 0,25]	0,11	0,00	[-0,22 ; 0,22]	0,11	
Polynômes Fractionnaires	-0,03	[-0,23 ; 0,17]	0,10	0,01	[-0,2 ; 0,22]	0,11	0,04	[-0,17 ; 0,25]	0,11	0,05	[-0,15 ; 0,24]	0,10	0,05	[-0,14 ; 0,24]	0,10	0,04	[-0,15 ; 0,24]	0,10	
Splines	-0,04	[-0,24 ; 0,15]	0,10	0,03	[-0,2 ; 0,25]	0,11	0,07	[-0,16 ; 0,29]	0,12	0,06	[-0,12 ; 0,25]	0,10	0,07	[-0,11 ; 0,26]	0,10	0,09	[-0,11 ; 0,28]	0,10	
Tout au long de la grossesse																			
Modélisation Indirecte	0,08	[-0,09 ; 0,24]	0,09	0,11	[-0,07 ; 0,28]	0,09	0,16	[-0,01 ; 0,33]	0,09	0,23*	[0,07 ; 0,39]	0,08	0,25*	[0,09 ; 0,41]	0,08	0,24*	[0,07 ; 0,41]	0,08	
Polynômes Fractionnaires	0,08	[-0,07 ; 0,23]	0,08	0,11	[-0,05 ; 0,27]	0,08	0,14	[-0,02 ; 0,30]	0,08	0,18*	[0,03 ; 0,32]	0,07	0,21*	[0,07 ; 0,35]	0,07	0,22*	[0,07 ; 0,37]	0,07	
Splines	0,07	[-0,08 ; 0,21]	0,07	0,14	[-0,02 ; 0,31]	0,09	0,19*	[0,02 ; 0,35]	0,09	0,19*	[0,05 ; 0,33]	0,07	0,22*	[0,08 ; 0,36]	0,07	0,24*	[0,10 ; 0,39]	0,07	

Abbreviations: IMC, Indice de Masse Corporelle; IC, Intervalle de Confiance; ÉT, Écart-type

^aAjusté sur le niveau d'études, l'IMC et l'âge à l'accouchement de la mère; la durée d'allaitement; le centre de recrutement; le sexe de l'enfant.

*P < 0,05

4. Comparaison des mesures d'association obtenues à partir des données cliniques et des données prédites

Les différences mises en évidence entre les 3 groupes de tabagisme maternel à partir des données des examens cliniques étaient très semblables à celles mises en évidence à partir des données prédites par les modèles utilisés dans notre approche sur le sous-échantillon d'enfants ayant des données cliniques, à 1, 3 et 5 ans (tableau 21).

Tableau 21 : Différences moyennes de taille, poids et IMC obtenues à partir des examens cliniques (1, 3 et 5 ans) et des données prédites par la modélisation dans chaque catégorie de tabagisme maternel et la catégorie de référence (absence de tabagisme)^a, méthode en « 2 étapes »

Variables selon le statut tabagique et type de données	1 an		3 ans		5 ans	
	β (n)	IC 95%	β (n)	IC 95%	β (n)	IC 95%
Taille, cm						
Exclusivement le 1 ^{er} trimestre						
Données issues des examens cliniques	0,04 (119)	[-0,43 ; 0,51]	0,08 (99)	[-0,66 ; 0,81]	-0,15 (82)	[-1,19 ; 0,89]
Données prédites par le modèle (sous-échantillon d'enfants ayant des données cliniques)	0,08 (119)	[-0,35 ; 0,52]	0,06 (99)	[-0,62 ; 0,74]	-0,15 (82)	[-1,11 ; 0,81]
Tout au long de la grossesse						
Données issues des examens cliniques	-0,14 (228)	[-0,50 ; 0,22]	-0,17 (159)	[-0,78 ; 0,44]	-0,25 (142)	[-1,09 ; 0,58]
Données prédites par le modèle (sous-échantillon d'enfants ayant des données cliniques)	-0,24 (228)	[-0,55 ; 0,08]	-0,17 (159)	[-0,74 ; 0,39]	-0,26 (142)	[-1,03 ; 0,51]
Poids, g						
Exclusivement le 1 ^{er} trimestre						
Données issues des examens cliniques	5 (118)	[-194 ; 204]	-62 (99)	[-396 ; 272]	-21 (82)	[-612 ; 571]
Données prédites par le modèle (sous-échantillon d'enfants ayant des données cliniques)	29 (118)	[-151 ; 208]	72 (99)	[-242 ; 387]	-20 (82)	[-537 ; 497]
Tout au long de la grossesse						
Données issues des examens cliniques	79 (228)	[-74 ; 232]	125 (160)	[-152 ; 402]	358 (141)	[-117 ; 833]
Données prédites par le modèle (sous-échantillon d'enfants ayant des données cliniques)	71 (228)	[-67 ; 210]	151 (160)	[-110 ; 413]	324 (141)	[-90 ; 739]
IMC, kg/m²						
Exclusivement le 1 ^{er} trimestre						
Données issues des examens cliniques	-0,01 (118)	[-0,29 ; 0,27]	-0,09 (98)	[-0,33 ; 0,15]	0,01 (82)	[-0,29 ; 0,30]
Données prédites par le modèle (sous-échantillon d'enfants ayant des données cliniques)	0,01 (118)	[-0,23 ; 0,24]	0,05 (98)	[-0,19 ; 0,28]	0,01 (82)	[-0,27 ; 0,29]
Tout au long de la grossesse						
Données issues des examens cliniques	0,20 (228)	[-0,01 ; 0,42]	0,18 (158)	[-0,02 ; 0,37]	0,36* (141)	[0,12 ; 0,60]
Données prédites par le modèle (sous-échantillon d'enfants ayant des données cliniques)	0,23* (228)	[0,05 ; 0,41]	0,21* (158)	[0,02 ; 0,41]	0,33* (141)	[0,11 ; 0,55]

Abréviations: IMC, Indice de Masse Corporelle, IC, Intervalle de Confiance.

^aAjustées sur le niveaux d'études, l'IMC et l'âge à l'accouchement de la mère; la durée d'allaitement; le centre de recrutement; le sexe de l'enfant.

*P<0,05.

3.5 Discussion

Les enfants de l'étude EDEN exposés au tabagisme de leur mère tout au long de la grossesse avaient un IMC significativement plus élevé que celui des enfants non exposés à partir de 2 ans. Bien que l'association ne soit pas significative, les trajectoires d'IMC de ces deux groupes se distinguaient dès les premiers mois. Ces différences étaient expliquées par un rattrapage pondéral plus important que le rattrapage statural et renforcées par un gain de poids plus rapide à partir de 1 an.

D'un point de vue méthodologique, la méthode en « 1 étape » permettait d'examiner les trajectoires de manière longitudinale et les associations globales entre l'exposition et les trajectoires de croissances staturale et pondérale. De plus, elle permettait d'étudier les associations entre l'exposition et les paramètres du modèle de Jenss-Bayley qui représentent des caractéristiques spécifiques aux différentes périodes de la croissance. La méthode en « 2 étapes » fournissait des tests statistiques à chaque âge d'intérêt et particulièrement avec l'IMC, permettant de quantifier l'ampleur des différences entre les différents sous-groupes de tabagisme.

Les performances des polynômes fractionnaires et des splines ont montré des résultats comparables sur des données simulées (Binder et al., 2013). Sur notre échantillon, les splines (mais pas les polynômes fractionnaires) semblaient mieux prédire les données d'IMC que notre méthode de modélisation indirecte. Cependant, les qualités d'ajustement de celle-ci étaient bonnes et d'autres avantages lui confèrent une plus-value. Tout d'abord, les paramètres des modèles de poids et de taille ont une interprétation biologique. D'autre part, les modèles utilisés sont moins sensibles aux valeurs extrêmes contrairement aux modèles non-structuraux (Chirwa et al., 2014). De plus, la méthode de modélisation indirecte ne nécessite pas, pour la prédiction de l'IMC, de disposer à la fois de mesure de poids et de taille aux mêmes jours, contrairement aux méthodes de modélisation directes, ce qui permet d'augmenter le nombre de mesures disponibles (30515 mesures de poids et 28381 mesures de taille *versus* 26693 pour l'IMC). Bien qu'il y ait, comme dans toute cohorte de ce type, une certaine attrition à 5 ans, la population disponible pour la conduite des analyses à partir des données observées, représentait 91%, 72% et 65% de l'échantillon initial respectivement à 1, 3 et 5 ans. Comparées aux analyses faites à partir des données observées, les analyses

réalisées à partir des données prédites par les modèles permettent de prendre en compte les données manquantes et de limiter un biais de sélection potentiel. Le biais que pourrait toutefois induire la prédiction des trajectoires de croissance en dehors des intervalles d'âge pour lesquels des mesures sont disponibles (enfants n'ayant pas été suivi jusqu'à 5 ans) a été exploré en répétant les analyses à partir des enfants 1211 enfants ayant des mesures après 4 ans. Les conclusions sur ce jeu de données étaient comparables à celles faites à partir de l'échantillon complet. Enfin, les données observées, contrairement aux données prédites par les modèles, ne permettent d'évaluer les associations que pour les âges de l'étude et, de ce fait, ne permettent pas de caractériser finement l'évolution des différences entre les sous-groupes de tabagisme au cours du temps. Elles ne permettent pas non plus d'évaluer les associations avec des caractéristiques clés de la croissance comme les vitesses de croissance instantanées et l'âge et l'IMC au pic d'IMC.

Une des faiblesses réside dans le fait que la méthode de modélisation indirecte ne permet pas d'évaluer les associations globales entre l'exposition d'intérêt et les trajectoires d'IMC comme cela est possible pour les trajectoires de poids et de taille (même si des problèmes de convergence peuvent être rencontrés avec l'augmentation du nombre de covariables).

Une des limites potentielles susceptibles de biaiser les associations entre le tabagisme de la mère exclusivement pendant le 1^{er} trimestre et la croissance des enfants, est que le statut tabagique en début de grossesse a été utilisé comme un proxy du statut tabagique au 1^{er} trimestre. On peut donc penser que le sous-groupe de mères fumeuses exclusivement au 1^{er} trimestre est composé d'un fort pourcentage de femmes arrêtant de fumer précocement. Les enfants de ces mères pourraient donc n'avoir été exposés que très transitoirement ce qui pourrait expliquer l'absence d'association entre le tabagisme de la mère au 1^{er} trimestre et les caractéristiques anthropométriques des enfants à la naissance. De plus, le fait que le statut tabagique ait été recueilli à l'aide d'auto-questionnaires pourrait conduire à un biais de classement. En effet, les femmes pourraient avoir tendance à sous-déclarer ou sous-estimer leur tabagisme ou leur consommation de tabac respectivement dû à un biais de désirabilité. Cependant, on ne s'attend pas à ce que les différences des caractéristiques anthropométriques mises en évidence entre les enfants exposés et les enfants non-exposés soient sur-estimées par

une sous-déclaration car le questionnaire était posé avant le recueil des données de santé de l'enfant.

Dans des revues récentes sur les déterminants du surpoids ou de l'obésité, le tabagisme de la mère pendant la grossesse a été associé à l'obésité de l'adulte (Brisbois et al., 2012a) et au surpoids et à l'obésité de l'enfant (Ino, 2010; Oken et al., 2008). Ces revues sont en faveur d'un effet à long terme du tabagisme sur la santé. Notre étude s'ajoute aux études précédemment conduites, suggérant l'émergence de ces associations dans la toute petite enfance. Les quelques études ayant caractérisé la croissance précoce expliquent la supériorité de l'IMC des enfants de femmes fumeuses comparé à celui des enfants de femmes non fumeuses par un déficit de poids à la naissance suivi par un rattrapage pondérale rapide avant 1 an conduisant à un poids similaire ou plus élevé plus tard dans l'enfance (Chen et al., 2006; Durmus et al., 2011; Timmermans et al., 2013). Alors que ces enfants ont tendance à rattraper leur déficit de poids, leur déficit en taille semble persister dans l'enfance (Chen et al., 2006; Durmus et al., 2011; Howe et al., 2012). Nos résultats sur les vitesses de croissance ainsi que sur les paramètres C et D montrent des associations plus significatives du tabagisme maternel avec la croissance staturale qu'avec la croissance pondérale dans les 6 premiers mois de vie, indiquant qu'il existe également un phénomène de rattrapage statural clair bien que plus lent. Si les effets sont moins significatifs sur le poids, ils vont cependant dans le même sens et la trajectoire moyenne d'IMC des enfants exposés tout au long de la grossesse suggère néanmoins une augmentation de poids plus rapide que celui de la taille dès les premiers mois, avec un IMC plus élevé au pic d'IMC et une différence maintenue durant toute l'enfance.

Dans notre population, le tabagisme maternel tout au long de la grossesse était associé à un poids et une taille à la naissance plus faibles. Ces associations n'étaient pas retrouvées chez les enfants exposés exclusivement durant le 1^{er} trimestre. Ces résultats sont en accord avec les associations mises en évidence entre le tabagisme maternel et une restriction de la croissance fœtale (Iñiguez et al., 2013; Jaddoe et al., 2007a). De plus, le tabagisme maternel exclusivement durant le 1^{er} trimestre n'était associé à l'IMC de l'enfant à aucun des âges d'intérêt. Ces résultats vont dans le sens de ceux mis en évidence par une étude ne mettant pas en évidence d'association entre le tabagisme maternel au 1^{er} trimestre et l'IMC des enfants à 4 ans (Durmus et al., 2011), mais

contrastent avec une autre mettant en évidence un risque accru de surpoids entre 5 et 7 ans (Mendez et al., 2008).

Les associations entre le tabagisme et la croissance pourraient être attribuées à des effets directs ou indirects. Les résultats de la littérature suggèrent que le tabagisme maternel pourrait avoir un effet direct via l'augmentation des concentrations sanguines maternelles et fœtales en carboxyhémoglobine, responsable d'hypoxie fœtale et d'un risque de retard de croissance (DiFranza et al., 2004). Les associations avec la croissance postnatale pourraient être expliquées par une dérégulation des centres hypothalamiques en charge de la régulation de l'apport alimentaire, la croissance et le nombre d'adipocytes. Le tabac pourrait également avoir un effet potentialisateur sur les structures hypothalamiques conduisant à un signal et un métabolisme de l'insuline altérés (Toschke et al., 2003).

Plusieurs études se sont attachées à comparer la croissance de frères et sœurs dont le statut tabagique de la mère pendant la grossesse différait, afin de s'affranchir le plus possible d'une confusion liée à des modes de vie délétères partagés. Les associations avec le tabagisme maternel et paternel ont également été comparées pour étudier le rôle de l'environnement. Les résultats de ces études sont hétérogènes. Des associations plus importantes entre le tabagisme maternel et la croissance, comparées à celles observées avec le tabagisme paternel (Riedel et al., 2014), sont en faveur d'un effet intra-utérin. Des analyses entre frères et sœurs montrant des associations atténuées entre le tabagisme maternel et le risque d'obésité chez de jeunes adultes (Iliadou et al., 2010) et des études montrant des associations similaires (Laura D Howe et al., 2012; Kwok et al., 2010; Leary et al., 2006; vonKries et al., 2008) ou plus fortes (Florath et al., 2014) entre le tabagisme paternel comparé au tabagisme maternel et le surpoids ou l'IMC de l'enfant suggèrent, quant à elles, l'existence d'un phénomène de confusion résiduelle liée à des facteurs familiaux non mesurés. L'ajustement de nos analyses sur les profils de pratiques alimentaires ne modifiait pas les associations mises en évidence. Cependant, ces profils portent sur des données alimentaires dans la 1^{re} année de vie, une période au cours de laquelle l'alimentation est principalement lactée et l'ajustement sur les profils de pratiques alimentaires pourrait apparaître comme comparable à l'effet d'un ajustement sur l'allaitement. Nous n'avons pas ajusté sur des caractéristiques reflétant le mode de vie plus tard dans l'enfance telles que l'alimentation après la première année ou l'activité physique. En effet, comme les différences observées sur

l'IMC apparaissent très tôt, l'effet de facteurs familiaux partagés semble peu probable pour expliquer les différences. L'augmentation rapide de l'IMC durant les 1^{res} semaines de vie pourrait suggérer que l'arrêt de l'exposition aux produits tels que la nicotine juste après la naissance pourrait contribuer aux associations observées, comme cela a été suggéré (Chen and Morris, 2007). En effet, l'arrêt du tabagisme entraîne généralement un gain de poids lié à un apport alimentaire accru et une dépense énergétique diminuée.

3.6 Conclusion

Nos résultats sont en faveur d'une association entre le tabagisme tout au long de la grossesse et des trajectoires de croissance de l'IMC plus élevées dans les 1^{ers} mois de vie et plus fortement après 2 ans. Ces conclusions soutiennent le concept stipulant que la programmation de l'enfant liée à une exposition fœtale à des conditions délétères a des effets persistant à long terme ou dans l'ensemble de la vie. Les deuxième et troisième trimestre(s) de la grossesse pourraient être des périodes particulièrement critiques dans l'induction d'une susceptibilité à l'obésité. Ce travail montre que notre approche basée sur la modélisation de la taille et du poids offre la possibilité d'estimer les caractéristiques clés des trajectoires de croissance et leurs associations avec des déterminants précoces.

**Chapitre 4 : Tabagisme maternel
pendant la grossesse et croissance de
l'enfant : une approche bayésienne**

4.1 Introduction

Dans le chapitre 1, nous avons décrit deux approches permettant d'estimer les paramètres dans les modèles non-linéaires à effets mixtes : l'approche fréquentiste, basée sur le principe du maximum de vraisemblance et l'approche bayésienne.

Les modèles de croissance staturo-pondérale modélisent communément la croissance de la taille et du poids de manière indépendante. Or, il existe une forte relation de dépendance entre le poids et la taille qu'il peut être utile de caractériser ou de prendre en compte pour l'amélioration des performances des modèles. Par ailleurs, l'IMC, qui est une résultante de ces 2 paramètres, pourrait gagner à être déduit d'un modèle conjoint. Enfin, ces modèles sont majoritairement implémentés à l'aide de méthodes fréquentistes. Or, des approches bayésiennes appliquées aux modèles non linéaires à effets aléatoires (NLMEMs) ont été proposées, particulièrement pour l'étude de la dynamique du virus de l'immunodéficience humaine (VIH), et se sont montrées puissantes pour l'analyse de la complexité de ce type de données. En effet, la spécification de distributions a priori sur les paramètres des modèles permet de bénéficier non seulement des connaissances apportées par les études antérieures, mais également de résoudre des problèmes d'identifiabilité numérique (De la Cruz-Mesía and Marshall, 2006). L'approche bayésienne est également adaptée à l'ajustement de modèles complexes comme les modèles hiérarchiques et multi-niveaux. De plus, un des intérêts majeurs inhérents au raisonnement bayésien est la possibilité de disposer de la distribution a posteriori complète des paramètres. Ainsi, cette approche offre la possibilité de faire des prédictions en tenant compte de l'incertitude sur les paramètres.

4.2 Objectifs spécifiques

Le chapitre 3 a illustré de quelle manière l'algorithme SAEM peut être utilisé pour étudier les associations entre des expositions précoces (le tabagisme maternel pendant la grossesse dans l'exemple choisi) et la croissance de la taille et du poids de la naissance à 5 ans et en tirer des informations sur la croissance de l'IMC.

L'objectif de cette partie était d'illustrer l'utilisation d'une approche bayésienne sur les mêmes données. Nous nous sommes intéressés à cette approche pour ses facilités calculatoires, spécialement quand le nombre de facteurs à considérer en association avec la croissance est élevé. Nous avons postulé que ces avantages devaient nous

permettre, à l'aide d'un modèle bivarié, de modéliser conjointement le poids et la taille et d'évaluer simultanément les associations entre des déterminants d'intérêt et la croissance du poids et de la taille et finalement de faire une inférence sur l'IMC. De plus, nous avons fait l'hypothèse que nous gagnerions en efficacité pour estimer les paramètres caractérisant la croissance de la taille grâce aux informations issues de la modélisation du poids et réciproquement.

4.3 Méthodes

4.3.1 Modèle de croissance

Le modèle conjoint est composé de deux sous-modèles non-linéaires mixtes, un pour la taille et un pour le poids, dont les paramètres sont liés par une matrice de covariances. Il peut être présenté comme un modèle hiérarchique à 3 niveaux (Davidian and Giltinan, 1995).

1^{er} niveau : variabilité intra-individuelle

Nous commençons par écrire le modèle selon une expression fréquentiste :

$$y_{ijk} = f(\phi_{ik}, t_{ijk}) + e_{ijk} \quad (1)$$

Où $k = (P, T)$ définit les deux événements d'intérêt i.e. le poids et la taille respectivement. $y_{ik} = (y_{i1k}, \dots, y_{in_{ik}k})^T$ dénote un vecteur de dimension $n_{ik} \times 1$ des mesures observées de taille et de poids de l'individu i au temps t_{ij} , $i=1, \dots, N$, $j=1, \dots, n_{ik}$.

Où f est, pour la variable anthropométrique k , une fonction non-linéaire du m -vecteur des paramètres spécifiques de chaque individu ϕ_{ik} ; $e_{ik} = (e_{i1k}, \dots, e_{in_{ik}k})^T$ est un vecteur de dimension $n_{ik} \times 1$ d'erreurs résiduelles supposées indépendantes et identiquement distribuées selon une loi normale centrée en zéro et de variance telle que :

$$e_{ik} = (e_{ijk}) \sim N(0, \sigma_{e_k}^2 I_{n_{ik}}) \quad (2a)$$

Pour l'étude de simulation (application aux données de croissance simulées) (cf. paragraphes 4.3.3.1 et 4.4.1), il a été considéré que les variances résiduelles étaient homogènes au cours du temps, que ce soit pour le poids ou la taille.

$$\log(\sigma_{e_{ijk}}^2) = \delta_{0k}$$

Pour l'application aux données de croissance réelles (cf. paragraphes 4.3.3.2 et 4.4.2), une fonction générale de la variance permettant une variation de la variance avec l'âge a été considérée afin de prendre en compte l'hétéroscédasticité de la variance résiduelle comme cela a été décrit par Foulley et coll. et Duval et coll. (Duval et al., 2009;

Foulley et al., 1992). Plus précisément, nous avons sélectionné les meilleures fonctions permettant de décrire les variances résiduelles du poids et de la taille parmi deux fonctions candidates. Les équations des deux modèles d'erreur considérés étaient les suivantes :

$$\text{Modèle d'erreur 1 : } \log(\sigma_{e_{ijk}}^2) = \delta_{0k} + \delta_{1k}t_{ijk}$$

$$\text{Modèle d'erreur 2 : } \log(\sigma_{e_{ijk}}^2) = \delta_{0k} + \delta_{1k}t_{ijk} + \delta_{2k}t_{ijk}^2$$

Le choix du modèle d'erreur 1 pour la taille et du modèle d'erreur 2 pour le poids permettait de minimiser les DIC/WAIC.

Pour la taille, le modèle d'erreur suivant a donc été considéré :

$$\log(\sigma_{e_{ijT}}^2) = \delta_{0T} + \delta_{1T}t_{ijT}$$

Et pour le poids, le modèle d'erreur suivant a été considéré

$$\log(\sigma_{e_{ijP}}^2) = \delta_{0P} + \delta_{1P}t_{ijP} + \delta_{2P}t_{ijP}^2$$

Le choix des fonctions est expliqué dans les paragraphes 4.3.3.2 et 4.4.2.

Les formules (1) et (2a) équivalent, pour des modèles hiérarchiques bayésiens, à écrire la distribution des données conditionnellement aux paramètres individuels

$$y_{ik} | \phi_{ik} \sim N(f_{ik}, \sigma_{e_k}^2 I_{n_{ik}}) \quad (2b)$$

Où $y_{ik} = (y_{ijk})$ pour $j=1, \dots, n_{ik}$ est un vecteur de dimension $n_{ik} \times 1$ et de la même manière f_{ik} est définie comme $f_{ik} = (f(\phi_{ik}, t_{ijk}))$ pour $j=1, \dots, n_{ik}$ et $I_{n_{ik}}$ est la matrice identité.

Pour la modélisation du poids et de la taille, f est la modification de l'équation non-linéaire de Jeness-Bayley introduite précédemment, qui améliore l'interprétation des paramètres et tient compte de contraintes de signes sur les coefficients (Botton et al., 2014b; Jeness and Bayley, 1937). Comme déjà mentionné, cette fonction permet d'exprimer la taille (en cm) et le poids (en kg) attendus d'un individu i au temps t_{ij} de la manière suivante:

$$f(\phi_{ik}, t_{ijk}) = \exp(A_{ik}) + \exp(B_{ik}) t_{ijk} + \exp(C_{ik}) (1 - \exp(-\exp(D_{ik}) t_{ijk})) \quad (3)$$

Pour $k = (P, T)$ et $\phi_{ik} = (A_{ik}, B_{ik}, C_{ik}, D_{ik})^T$

La signification des paramètres est la même que décrite au chapitre 3, paragraphe 3.3.2.2 page 67).

2^e niveau : variabilité inter-individuelle

Les effets individuels sont traités comme aléatoires avec la distribution multinormale suivante :

$$\phi_i \sim N(\mu_i, \Omega) \quad (4)$$

Où $\phi_i = (\phi_{iT} \quad \phi_{iP})^T$ avec ϕ_{iT} et ϕ_{iP} représentant les vecteurs des paramètres de taille et de poids respectivement, de vecteur moyen $\mu_i = (\mu_{iT} \quad \mu_{iP})^T$ et de matrice de variances-covariances $\Omega = \begin{bmatrix} \Omega_{TT} & \Omega_{TP} \\ \Omega_{PT} & \Omega_{PP} \end{bmatrix}$ de dimension $2m \times 2m$.

Ici, la contribution individuelle $\phi_{ik} = (A_{ik}, B_{ik}, C_{ik}, D_{ik})^T$ dispose de $m=4$ paramètres pour chaque enfant i et chaque caractéristique k , et a pour moyenne

$$\mu_{ik} = (\mu_{ik}^A, \mu_{ik}^B, \mu_{ik}^C, \mu_{ik}^D)^T \quad \text{et pour matrice de variances-covariances}$$

$$\Omega_{kk} = \begin{pmatrix} \overline{\omega}_{A_k A_k} & \cdots & \overline{\omega}_{A_k D_k} \\ \vdots & \ddots & \vdots \\ \overline{\omega}_{D_k D_k} & \cdots & \overline{\omega}_{D_k D_k} \end{pmatrix} \text{ pour la caractéristique } k \text{ et comme matrice de covariances}$$

$$\Omega_{kk'} = \begin{pmatrix} \overline{\omega}_{A_k A_{k'}} & \cdots & \overline{\omega}_{A_k D_{k'}} \\ \vdots & \ddots & \vdots \\ \overline{\omega}_{D_k A_{k'}} & \cdots & \overline{\omega}_{D_k D_{k'}} \end{pmatrix} \text{ entre les caractéristiques } k \text{ et } k'.$$

Les moyennes μ_{ik} de la distribution (4) sont reliées à une matrice de variables explicatives X_i de dimension $m \times p$ (intercept et covariables qualitatives ou continues) via un modèle de régression classique impliquant un vecteur de paramètres β_k de dimension p .

$$\mu_{ik} = X_i \beta_k \quad (5)$$

Dans les modèles sans covariable, le vecteur d'effets fixes β_k est constitué des ordonnées à l'origine $(\beta_{0_{A_k}}, \beta_{0_{B_k}}, \beta_{0_{C_k}}, \beta_{0_{D_k}})^T$ correspondant à la moyenne des paramètres de croissance dans la population. Dans les modèles incluant un effet des covariables sur les paramètres de croissances (cf. application aux données réelles, paragraphes 4.3.3.2), β_k inclut à la fois un vecteur de dimension m d'intercepts $\beta_{0_{A_k}}, \beta_{0_{B_k}}, \beta_{0_{C_k}}, \beta_{0_{D_k}}$ et un vecteur de mp_{cov} coefficients de régression $\beta_{1_{A_k}} \cdots \beta_{p_{cov_{A_k}}}, \dots, \beta_{1_{D_k}} \cdots \beta_{p_{cov_{D_k}}}$ correspondant aux covariables p_{cov} d'intérêt.

3^e niveau : distributions a priori des hyperparamètres

Des distributions a priori communément utilisées pour les espérances et les variances des modèles hiérarchiques bayésiens ont été spécifiées (De la Cruz-Mesía and Marshall, 2006; Donnet et al., 2010). Les paramètres de ces distributions ont été choisis à partir des valeurs décrites par des études de modélisation de la croissance incluant celles réalisées sur les données de l'étude EDEN. Le Tableau 22 décrit les distributions a priori choisies dans ce travail (notations rjags).

Pour le sous-modèle de la croissance du poids, nous avons considéré une précision (inverse de la variance) de 16 pour l'ensemble des paramètres de croissance. Nous considérons que l'intercept du paramètre A du poids suit une loi normale de moyenne 1, ce qui correspond à une valeur moyenne a priori de 2,7 kg pour l'extrapolation du poids à la naissance et une probabilité de 95% que cette moyenne soit comprise entre 1,7 et 4,4 kg. Pour le paramètre B, nous avons considéré une moyenne de -5 pour le paramètre C, une moyenne de 1,5 et pour le paramètre D une moyenne de -5.

La même approche a été utilisée pour le sous-modèle de la taille. Cependant, pour chaque paramètre, les distributions a priori (moyennes des distributions normales) ont été adaptées pour représenter des valeurs de taille à la naissance et des trajectoires de croissance staturales plausibles.

En présence de covariables, nous avons spécifié, afin que les paramètres principaux des modèles (A_k, B_k, C_k, D_k) conservent des valeurs compatibles avec des trajectoires plausibles, des distributions a priori normales des coefficients de régression du vecteur β_k (i.e., excepté les intercepts $\beta_{0_{A_k}}, \beta_{0_{B_k}}, \beta_{0_{C_k}}, \beta_{0_{D_k}}$) centré en zéro et avec une précision de 16 (ce qui correspond à un écart-type de 0,25 soit une probabilité de 95% que le coefficient soit compris entre -0,5 et 0,5) pour $k = P, T$.

Pour la précision résiduelle et la matrice de précision, nous avons utilisé les distributions décrites par De la Cruz et coll., et utilisées par Donnet et coll., pour les modèles non linéaires à effets aléatoires (De la Cruz-Mesía and Marshall, 2006; Donnet et al., 2010). Pour l'étude de simulation, utilisant les propriétés de conjugaison, la précision résiduelle a été supposée suivre une distribution Gamma $\sigma_{e_k}^2 \sim \gamma(a; b)$. Ici, il a été considéré que $a = b = \varepsilon$ avec ε fixée à une valeur faible (e.g., ici $\varepsilon = 10^{-4}$) pour approximer une densité a priori vague de Jeffreys $\pi(\sigma_{e_k}^2) = \sigma_{e_k}^{-2}$. Pour l'application, les lois a priori ont été spécifiées sur les paramètres de la fonction décrivant la variance

résiduelle δ_{0T} et δ_{1T} pour la taille et δ_{0P} , δ_{1P} et δ_{2P} pour le poids. Des distributions a priori normales centrées en zéro et une précision de 16 ont été considérées pour l'ensemble de ces paramètres.

La matrice de précision $\Gamma = \Omega^{-1}$ des 8 paramètres individuels (A, B, C et D pour le poids et la taille) a été défini comme suivant une distribution de Wishart $W(R, \nu)$ qui est une distribution appropriée aux matrices définies positives. Ici, $R = \Gamma_0/\nu$ où $\Gamma_0 = \Omega_0^{-1}$ est la moyenne a priori de la distribution de Wishart et ν représente le nombre de degrés de liberté qui a été fixé égal au nombre de paramètres+1 (deux caractéristiques avec m composantes chacune plus un i.e. $\nu=2m+1$, comme préconisé par (Gelman and Hill, 2006). Ce nombre de paramètres doit être égal ou supérieur à la dimension de la matrice de précision et des valeurs faibles indiquent une loi a priori peu informative (Lunn et al., 2012). Spécifier une distribution de Wishart sur la précision Γ équivaut à dire que la matrice de variances-covariance Ω suit une loi de Wishart inverse $IW(S, \nu)$ de moyenne $E(\Omega) = S/(\nu - 2m - 1)$ avec $S = R^{-1}$ (Anderson, 1984). Dans rjags, la distribution de Wishart est spécifiée par $S = \nu\Omega_0$ plutôt que par R. Ainsi il a été considéré $\Gamma \sim W(S, 9)$ avec S choisi de manière empirique sur la base d'estimations antérieures de la matrice de variances-covariances entre les paramètres des modèles de croissance de la taille et du poids

La matrice de variances-covariances suivante a été considérée :

$$\Omega_0 = \begin{matrix} & \begin{matrix} A_P & B_P & C_P & D_P & A_T & B_T & C_T & D_T \end{matrix} \\ \begin{matrix} A_P \\ B_P \\ C_P \\ D_P \\ A_T \\ B_T \\ C_T \\ D_T \end{matrix} & \begin{pmatrix} 0,014 & 0,004 & 0,001 & 0,001 & 0,003 & 0,001 & -0,003 & 0,001 \\ 0,004 & 0,040 & -0,019 & 0,027 & 0,001 & 0,011 & 0,000 & 0,002 \\ 0,001 & -0,019 & 0,054 & -0,044 & 0,001 & -0,003 & 0,012 & -0,002 \\ 0,001 & 0,027 & -0,044 & 0,096 & 0,001 & 0,011 & -0,017 & 0,029 \\ 0,003 & 0,001 & 0,001 & 0,001 & 0,001 & 0,000 & 0,000 & -0,001 \\ 0,001 & 0,011 & -0,003 & 0,011 & 0,000 & 0,011 & -0,007 & 0,018 \\ -0,003 & 0,000 & 0,012 & -0,017 & 0,000 & -0,007 & 0,018 & -0,022 \\ 0,001 & 0,002 & -0,002 & 0,029 & -0,001 & 0,018 & -0,022 & 0,068 \end{pmatrix} \end{matrix}$$

Et ainsi S a été défini de la manière suivante :

$$S = \begin{pmatrix} 0,13 & 0,04 & 0,01 & 0,01 & 0,03 & 0,02 & -0,03 & 0,01 \\ 0,04 & 0,36 & -0,17 & 0,24 & 0,01 & 0,10 & 0,00 & 0,02 \\ 0,01 & -0,17 & 0,49 & -0,40 & 0,01 & -0,03 & 0,11 & -0,02 \\ 0,01 & 0,24 & -0,40 & 0,86 & 0,01 & 0,10 & -0,15 & 0,26 \\ 0,03 & 0,01 & 0,01 & 0,01 & 0,01 & 0,00 & 0,00 & -0,01 \\ 0,01 & 0,10 & -0,03 & 0,10 & 0,00 & 0,10 & -0,06 & 0,16 \\ -0,03 & 0,00 & 0,11 & -0,15 & 0,00 & -0,06 & 0,16 & -0,20 \\ 0,01 & 0,02 & -0,02 & 0,26 & -0,01 & 0,16 & -0,20 & 0,61 \end{pmatrix}$$

Tableau 22 : Distributions a priori des paramètres du modèle bayésien (notations rjags)

Paramètres	Poids ($k = P$)	Taille ($k = T$)
$\beta_{0_{A_k}}$	$dnorm(1; 16)$	$dnorm(4; 16)$
$\beta_{0_{B_k}}$	$dnorm(-5; 16)$	$dnorm(-4; 16)$
$\beta_{0_{C_k}}$	$dnorm(1,5; 16)$	$dnorm(3; 16)$
$\beta_{0_{D_k}}$	$dnorm(-5; 16)$	$dnorm(-5; 16)$
$\beta_{1_{(A-D)_k}}$ to $\beta_{p_{cov(A-D)_k}}$	$dnorm(0; 16)$	$dnorm(0; 16)$
Ω^{-1}		$dwish(S, 9)$
Variance résiduelle		
<i>Etude de simulation</i>		
$\sigma_{e_k}^{-2}$	$dgamma(0,0001; 0,0001)$	$dgamma(0,0001; 0,0001)$
<i>Application aux données réelles</i>		
δ_{0k}	$dnorm(0; 16)$	$dnorm(0; 16)$
δ_{1k}	$dnorm(0; 16)$	$dnorm(0; 16)$
δ_{2P}	$dnorm(0; 16)$	NA

4.3.2 Implémentation MCMC

4.3.2.1 Algorithme et logiciel

La distribution conditionnelle a posteriori de tous les paramètres des modèles a été calculée en utilisant des méthodes de Monte Carlo par chaînes de Markov avec le package rjags ;(version 3-15) dans R (version 3.1.0) (Plummer and Stukalov, 2014). Le package snow a permis d'effectuer l'estimation des différentes chaînes en parallèle. Dans ce package, les chaînes MCMC sont simulées avec un échantillonneur de Gibbs.

4.3.2.2 Evaluation de la convergence

Le diagnostic de convergence des chaînes s'est appuyé sur les critères précédemment décrits : l'inspection visuelle de la stationnarité et du mélange des 3 chaînes initiées avec des valeurs différentes en observant les valeurs simulées pour chaque paramètre d'intérêt en fonction du nombre d'itérations ; le diagnostic de Gelman

et Rubin ; le graphique des auto-corrélations et le graphique représentant l'évolution de la moyenne, du 2,5^e et du 97,5^e percentile des paramètres échantillonnés.

4.3.3 Application

4.3.3.1 Application sur données de croissance simulées

Une étude de simulation a été réalisée afin d'évaluer la capacité du modèle à estimer les caractéristiques clés de la croissance de manière adéquate. Des jeux de données de croissance staturale et pondérale ont été simulés. Au total, 100 jeux de données contenant les données anthropométriques fictives de 500 enfants avec 6 points de mesure allant de 5 à 1825 jours ont été générées soit 5 ; 30,44 (1 mois); 91,32 (3 mois); 182,64 (6 mois); 273,96 (9 mois); 365,25 (1 an) ; 730,5 (2 ans); 1095,75 (3 ans); 1461 (4 ans) et 1826,25 jours (5 ans). Le choix des valeurs moyennes des principaux paramètres définissant les trajectoires de croissance (A, B, C et D du poids et de la taille) dans la population (pour chaque jeu de donnée) a été guidé par les résultats de travaux précédents basés sur le même modèle et utilisant l'algorithme SAEM.

Les étapes suivantes ont permis la génération et l'analyse des jeux de données.

Etape 1:

Chaque paramètre individuel $A_{i_P}, B_{i_P}, C_{i_P}, D_{i_P}, A_{i_T}, B_{i_T}, C_{i_T}, D_{i_T}$ de l'individu i a été échantillonné dans une distribution multinormale avec pour valeurs moyennes, les valeurs décrites dans le Tableau 22. La matrice de variances-covariances des paramètres individuels du poids et de la taille a été choisie de manière empirique sur la base de travaux antérieurs.

Etape 2:

La taille et le poids ont ensuite été prédits pour chaque enfant du jeu de données simulé, aux différents temps de mesure (jours 5 à 1825) fixés, en appliquant l'équation de Jentsch-Bayley.

Etape 3:

En pratique, le poids et la taille sont sujets à une erreur de mesure. Pour chaque individu et chaque jeu de données, les données simulées au temps de mesure d'intérêt ont été obtenues en tirant une valeur dans une distribution normale avec comme moyenne la valeur prédite à l'étape 2 et un écart-type de 0,40 pour la taille et le poids.

Pour les modèles bayésiens, trois chaînes ont été lancées en parallèle en choisissant des valeurs initiales différentes pour les paramètres, une période de chauffe de 30000 itérations a été fixée, 10000 itérations supplémentaires ont été réalisées et les échantillons collectés par intervalle de 20 itérations.

Plusieurs éléments de validation, de nature fréquentiste, ont été calculés:

Le biais, représentant la différence entre l'estimation moyenne obtenue et la « vraie valeur » du paramètre (valeur moyenne utilisée à l'étape 1).

Le biais relatif, représentant le rapport entre cette différence et la « vraie valeur »

L'erreur moyenne quadratique, représentant la moyenne arithmétique des carrés des écarts entre les prévisions et les observations.

La probabilité de couverture, représentant la proportion des intervalles de crédibilité à 95% contenant la « vraie valeur ».

Les estimations des modèles bayésiens utilisées pour les calculs sont les moyennes des distributions a posteriori.

4.3.3.3 Application à l'analyse des associations entre le tabagisme maternel pendant la grossesse et la croissance postnatale

Le code rjags relatif à cette partie est disponible en Annexe 2.

4.3.3.3.1 Echantillon de données

Le même échantillon que celui du chapitre 3 a été utilisé pour ces analyses. Brièvement, nous avons analysé les données de 1666 enfants avec 30515 mesures de poids et 28381 mesures de taille de la naissance à 5 ans. Les variables d'ajustement étaient les mêmes que celles considérées précédemment. Pour des raisons calculatoires, l'âge de la mère a été centré sur la moyenne d'âge dans la population.

4.3.3.3.2 Aspects calculatoires

Les inférences sont réalisées à partir de 3 séquences (chaînes) indépendantes initiées en choisissant des valeurs initiales différentes. 40000 itérations ont été effectuées pour la phase de chauffe, ces itérations n'ont pas été incluses dans les inférences. Après avoir vérifié le mélange des chaînes, 20000 simulations ont été réalisées et les séquences ont été récupérées par intervalle de 20 itérations afin de limiter les autocorrélations.

4.3.3.3.3 Sélection et évaluation du modèle

4.3.3.3.3.1 Plus-value de la considération du statut tabagique

La plus-value de l'ajout des covariables caractérisant le statut tabagique dans les modèles de croissance a été évaluée en comparant les DICs (Deviance Information Criterion) et WAIC du modèle incluant les covariables et du modèle sans les covariables.

4.3.3.3.3.2 Qualité prédictive du modèle retenu

La qualité prédictive du modèle retenu a été évaluée tel que décrit au chapitre 1, p 38 en calculant la p-value prédictive a posteriori séparément pour le poids et pour la taille.

4.3.3.3.3 Sensibilité aux lois a priori

Afin de tester la sensibilité des résultats aux choix des distributions a priori, le modèle a été ré-évalué en spécifiant des lois très peu informatives. Ainsi, afin de fournir une information vague sur les intercepts des paramètres de croissance du poids et de la taille, les coefficients de régression, ainsi que les paramètres décrivant les fonctions des variances résiduelles du poids et de la taille, des lois normales centrées en 0 et de grande variance (1000000) ont été choisies (Tableau 23).

Tableau 23 : Distributions a priori des paramètres impliqués dans le modèle bayésien pour les analyses de sensibilité aux lois a priori

Paramètres	Poids ($k = P$)	Taille ($k = T$)
$\beta_{0_{A_k}}$	$dnorm(0; 10^{-6})$	$dnorm(0; 10^{-6})$
$\beta_{0_{B_k}}$	$dnorm(0; 10^{-6})$	$dnorm(0; 10^{-6})$
$\beta_{0_{C_k}}$	$dnorm(0; 10^{-6})$	$dnorm(0; 10^{-6})$
$\beta_{0_{D_k}}$	$dnorm(0; 10^{-6})$	$dnorm(0; 10^{-6})$
$\beta_{1_{(A-D)_k}}$ to $\beta_{p_{cov(A-D)_k}}$	$dnorm(0; 10^{-6})$	$dnorm(0; 10^{-6})$
Ω^{-1}		$dwish(S, 9)$
δ_{0k}	$dnorm(0; 10^{-6})$	$dnorm(0; 10^{-6})$
δ_{1k}	$dnorm(0; 10^{-6})$	$dnorm(0; 10^{-6})$
δ_{2W}	$dnorm(0; 10^{-6})$	NA

4.3.3.3.4 Différences de poids, taille et IMC selon le statut tabagique

Afin de quantifier les différences de taille, de poids et d'IMC entre les enfants exposés exclusivement au 1^{er} trimestre ou tout au long de la grossesse et la catégorie de référence, la taille, le poids et l'IMC aux 5^e, 15^e, 25^e et 30^e jours de vie et ensuite à chaque mois entre 30 jours et 5 ans ont été simulés pour des enfants fictifs dans chaque sous-groupe de tabagisme, en fixant les mêmes valeurs arbitraires pour les covariables que celles spécifiées dans le chapitre 3 (i.e. les garçons, allaités moins de 3 mois, nés à Poitiers, de mères ayant un IMC normal, un niveau d'études équivalent au lycée ou inférieur et ayant 30 ans au moment de l'accouchement).

La taille moyenne ainsi que le poids et l'IMC moyens ont été calculés pour chaque temps et chaque sous-groupe de tabagisme et les différences des valeurs moyennes

entre les enfants de mères fumeuses exclusivement au 1^{er} trimestre ou en tout au long de la grossesse et la catégorie de référence (enfants de mères non fumeuses) ont été calculées.

4.4 Résultats

4.4.1 Etude de simulation

Les estimations obtenues pour les 8 paramètres de croissance principaux (A, B, C et D pour le poids et la taille) étaient satisfaisantes avec des biais absolus et relatifs proches de 0 et des probabilités de couverture à 95% comprises entre 90 et 99%. La plupart des paramètres de variances-covariances était bien estimée par le modèle. Cependant, la capacité du modèle à prédire de manière adéquate la covariance entre le paramètre A du poids et D du poids ou de la taille était plus faible (biais relatif de -0,28) (tableau 24).

Tableau 24 : Valeurs réelles (fixées), biais et probabilité de couverture obtenus à partir du modèle bayésien conjoint de la taille et du poids sur 100 jeux de données simulés

Paramètre	Valeur réelle	Biais	Biais Relatif	Probabilité de couverture à 95%	RMSE
A _P	1,000	-0,00101	-0,00101	93	0,01005
B _P	-5,000	0,00074	0,00015	99	0,00743
C _P	1,500	-0,00056	-0,00037	97	0,00560
D _P	-5,000	-0,00031	-0,00006	96	0,00311
A _T	4,000	-0,00021	-0,00005	96	0,00206
B _T	-4,000	-0,00025	-0,00006	98	0,00246
C _T	3,000	0,00069	0,00023	95	0,00691
D _T	-5,000	-0,00098	-0,00020	96	0,00981
σ_{e_P}	0,400	0,00094	0,00234	91	0,00937
σ_{e_T}	0,400	0,00043	0,00107	96	0,00427
A _P A _P	0,014	0,00004	0,00281	96	0,00041
A _P B _P	0,004	-0,00004	0,00897	94	0,00040
A _P C _P	0,001	-0,00007	-0,06559	94	0,00073
A _P D _P	0,001	0,00032	-0,28490	97	0,00317
A _P A _T	0,003	0,00007	0,01989	93	0,00066
A _P B _T	0,001	-0,00017	0,15110	92	0,00168
A _P C _T	-0,003	0,00002	-0,00596	96	0,00020
A _P D _T	0,001	-0,00045	0,40938	94	0,00455
B _P B _P	0,040	0,00056	0,01395	98	0,00558
B _P C _P	-0,019	-0,00058	-0,03088	94	0,00583
B _P D _P	0,027	0,00037	0,01380	93	0,00368
B _P A _T	0,001	-0,00003	0,02553	96	0,00028
B _P B _T	0,011	0,00021	0,01914	91	0,00213
B _P C _T	0,000	-0,00003	NA	91	0,00025
B _P D _T	0,002	0,00018	0,08133	94	0,00181
C _P C _P	0,054	0,00157	0,02876	95	0,01566
C _P D _P	-0,044	-0,00088	-0,01976	92	0,00878
C _P A _T	0,001	0,00004	0,03741	91	0,00042
C _P B _T	-0,003	-0,00015	-0,04620	95	0,00154
C _P C _T	0,012	0,00040	0,03301	93	0,00403
C _P D _T	-0,002	-0,00025	-0,11475	95	0,00255
D _P D _P	0,096	0,00108	0,01126	95	0,01076
D _P A _T	0,001	-0,00004	0,03577	92	0,00040
D _P B _T	0,011	0,00041	0,03695	91	0,00411
D _P C _T	-0,017	-0,00078	-0,04672	93	0,00779
D _P D _T	0,029	0,00171	0,05936	95	0,01715
A _T A _T	0,001	0,00002	0,02037	94	0,00023
A _T B _T	0,000	-0,00004	NA	93	0,00036
A _T C _T	0,000	0,00002	NA	95	0,00016
A _T D _T	-0,001	-0,00008	-0,07391	91	0,00082
B _T B _T	0,011	0,00027	0,02464	90	0,00274
B _T C _T	-0,007	-0,00008	-0,01135	98	0,00076
B _T D _T	0,018	0,00057	0,03230	95	0,00574
C _T C _T	0,018	0,00023	0,01300	95	0,00231
C _T D _T	-0,022	-0,00038	-0,01689	96	0,00375
D _T D _T	0,068	0,00186	0,02748	94	0,01862

La figure 24 et la figure 25 illustrent, pour le poids et la taille, les données simulées et les trajectoires moyennes ajustées par le modèle ainsi que leur intervalle de crédibilité à 95%, pour 3 individus, chacun issu d'un jeu de données simulé différent et sélectionnés aléatoirement. Elles fournissent une validation visuelle de l'adéquation du modèle et de sa capacité à caractériser les trajectoires de croissance individuelles.

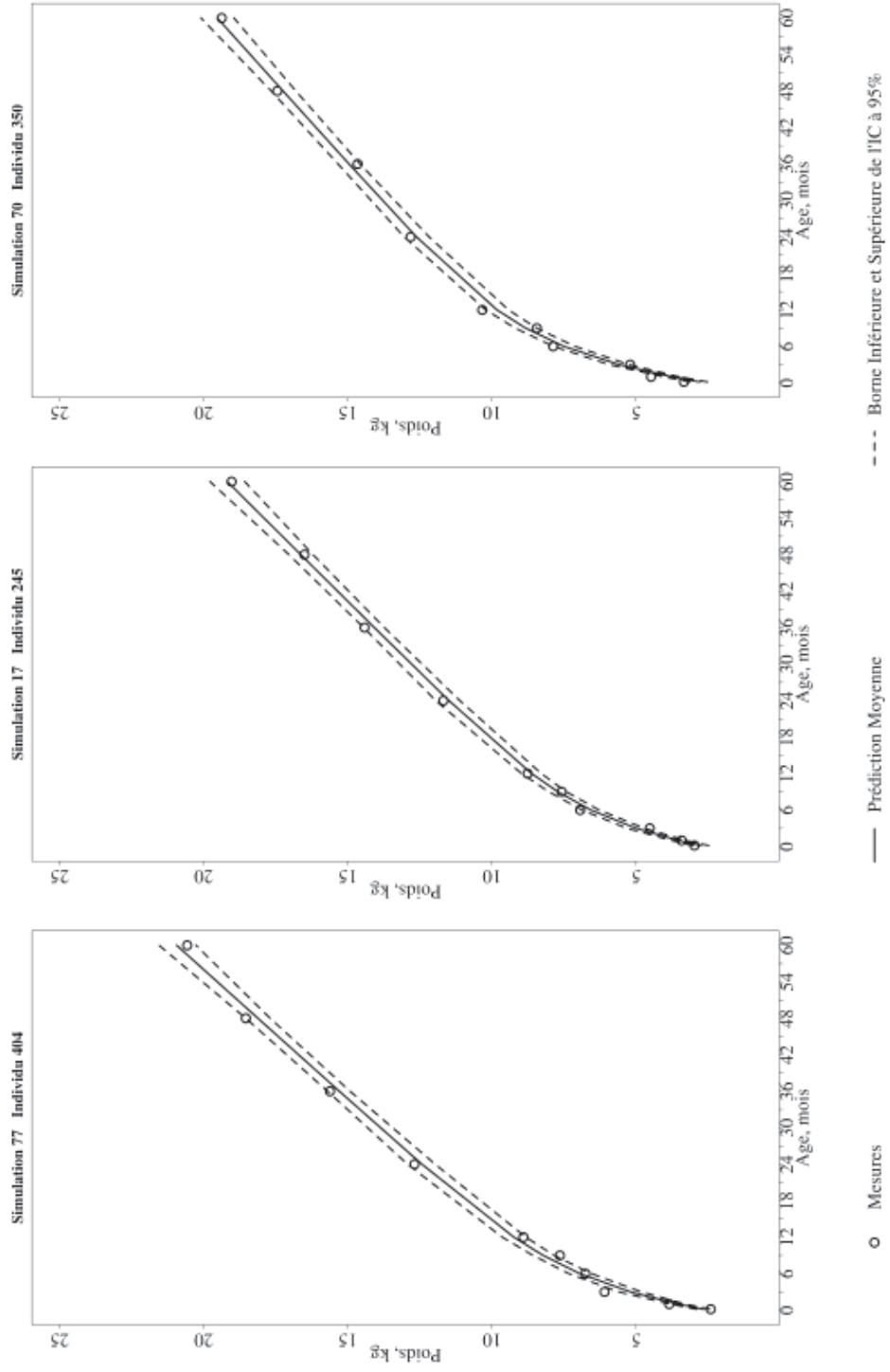


Figure 24 : Mesures de poids (données simulées), trajectoires moyennes ajustées par le modèle conjoint bayésien et intervalles de crédibilité à 95% obtenus pour 3 individus chacun issu d'un jeu de données simulé différent et sélectionnés aléatoirement.

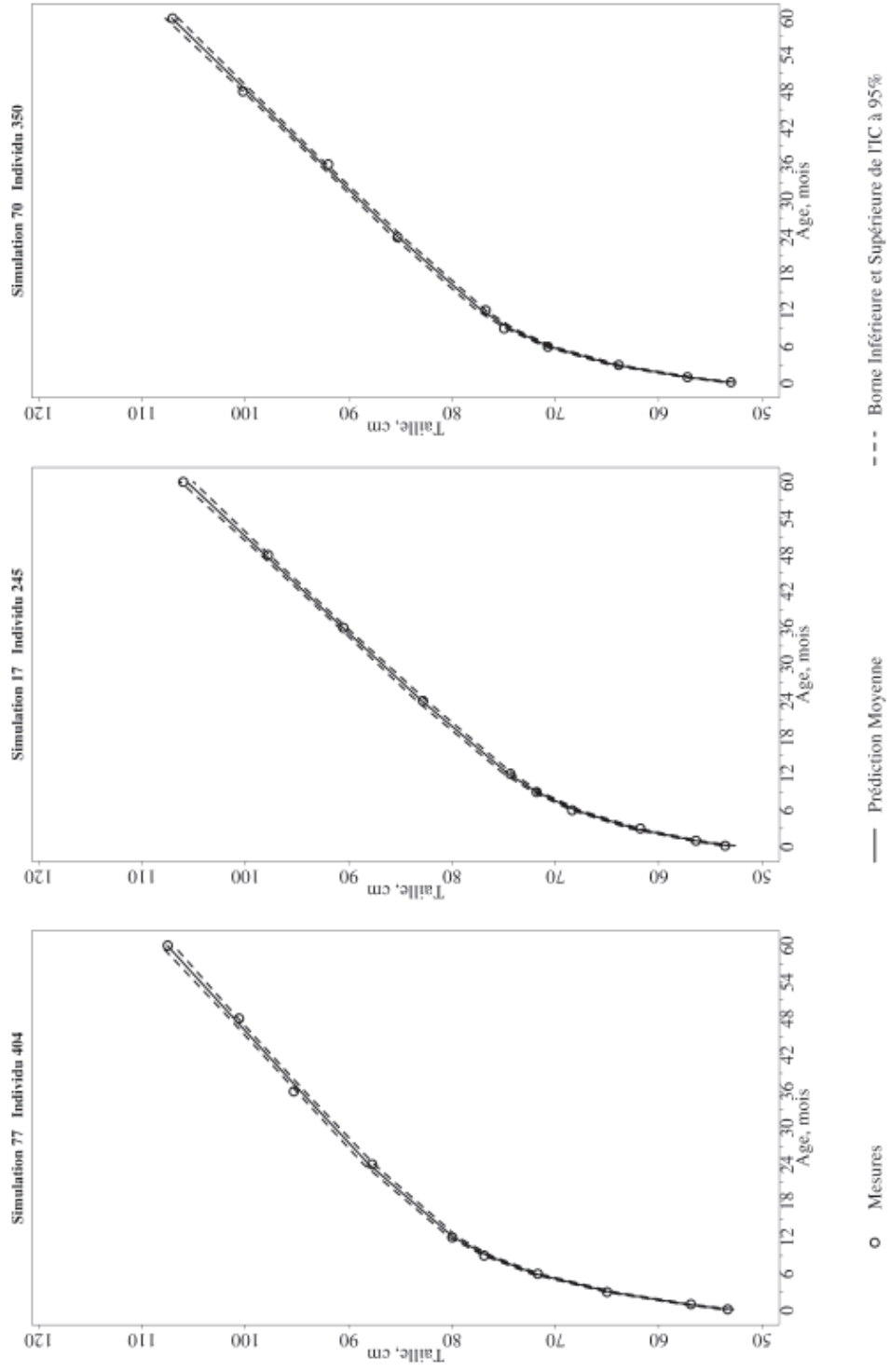


Figure 25 : Mesures de taille (données simulées), trajectoires moyennes ajustées par le modèle conjoint bayésien et intervalles de crédibilité à 95% obtenus pour 3 individus chacun issu d'un jeu de données simulé différent et sélectionnés aléatoirement.

4.4.2 Associations entre le tabagisme de la mère pendant la grossesse et la croissance de l'enfant

4.4.2.1 Evaluation de la convergence du modèle

Nous présentons dans cette partie l'évaluation de la convergence du modèle de croissance incluant les covariables de tabagisme. Excepté pour les valeurs du critère de Gelman, les critères de convergence sont illustrés à l'aide du coefficient de régression associé au tabagisme tout au long de la grossesse sur le coefficient B du poids. Ce choix est guidé par le fait que ce paramètre est un de ceux dont l'association avec l'exposition au tabagisme maternel était la plus forte.

Critère de Gelman

Les critères de Gelman étaient égaux à 1 pour l'ensemble des coefficients de régression associés au tabagisme (avec des bornes supérieures des intervalles de crédibilité comprises entre 1 et 1,01) témoignant de la convergence des chaînes. La Figure 26 illustre les valeurs du critère de Gelman, calculé à partir des 3 chaînes, au cours des itérations pour le coefficient de régression associé au tabagisme tout au long de la grossesse sur le coefficient B du poids.

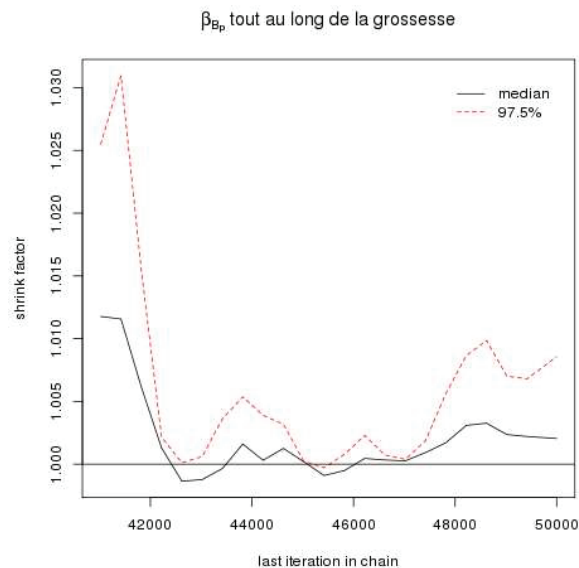


Figure 26 : Critère de Gelman pour le coefficient de régression associé au tabagisme maternel tout au long de la grossesse sur le paramètre B du modèle de croissance pondérale

Autocorrélations

Les autocorrélations semblaient satisfaisantes pour les trois chaînes et, de ce fait, les simulations des trois chaînes ont été conservées. La Figure 27 illustre les autocorrélations de la chaîne 1 pour le coefficient de régression associé au tabagisme tout au long de la grossesse sur le coefficient B du poids. Dans ce graphique, le « Lag » correspond à l'espacement (en terme de nombre d'itérations) entre les valeurs prédites utilisées pour le calcul de corrélation. Les chaînes ayant été échantillonnées toutes les 20 itérations, la corrélation de 1 (au Lag 0) correspond à la corrélation de la valeur prédite à la première itération t avec elle-même, puis sont représentées les corrélations entre les valeurs prédites à l'itération t et l'itération $t+20$ puis entre $t+20$ et $t+40$ et ainsi de suite.

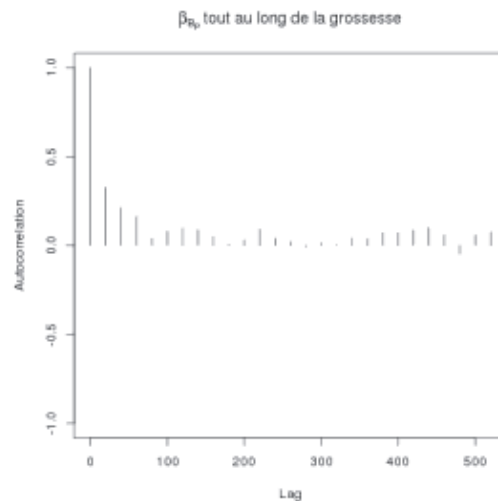


Figure 27 : Autocorrélation de l'une des 3 chaînes pour le coefficient de régression associé au tabagisme maternel tout au long de la grossesse sur le paramètre B du modèle de croissance pondérale

Mélange et stationnarité visuelle

Diagnostic visuel

Les trois chaînes utilisées étaient stationnaires et se confondaient sur la portion des itérations utilisées pour les inférences (après la phase de chauffe) témoignant de leur convergence. La Figure 28 illustre le mélange des 3 chaînes pour le coefficient de régression associé au tabagisme maternel tout au long de la grossesse sur le coefficient B du modèle de croissance du poids. Les figures illustrant le mélange des chaînes pour les intercepts des paramètres de croissance sont données en Annexe 3.

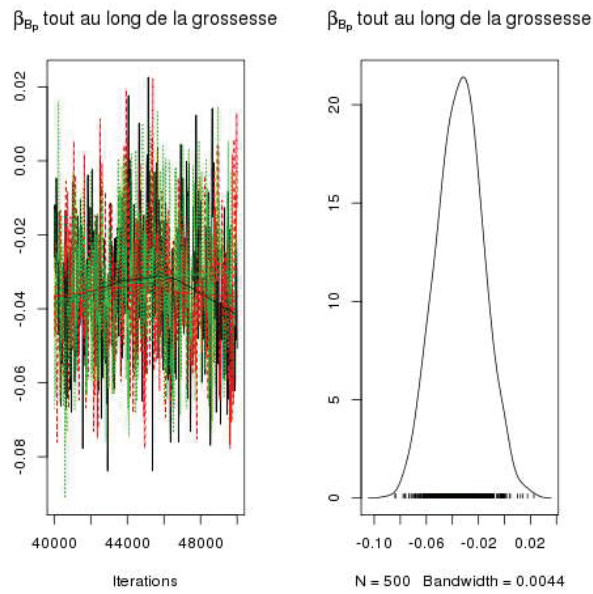


Figure 28 : Evolution des chaînes pour le coefficient de régression associé au tabagisme maternel tout au long de la grossesse sur le paramètre B du modèle de croissance pondérale

Evolution des quantiles

Le suivi de l'évolution des quantiles et de la moyenne de la distribution a posteriori des paramètres montrait également que la convergence était atteinte, les valeurs de ces mesures ne variant plus au cours des itérations. La Figure 29 illustre celui du coefficient de régression associé au tabagisme tout au long de la grossesse sur le paramètre B du modèle de croissance du poids pour la 1^{re} chaîne.

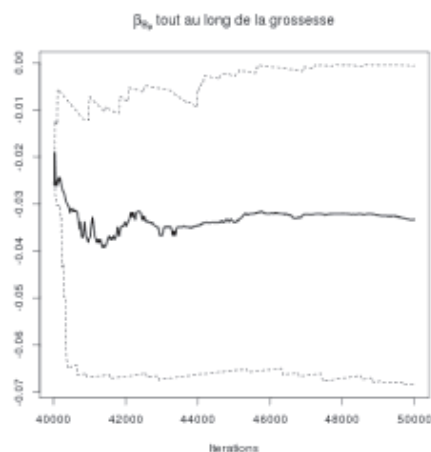


Figure 29 : Evolution des quantiles, en fonction du nombre d'itérations, de la distribution du coefficient de régression associé au tabagisme maternel tout au long de la grossesse et le paramètre B du modèle de croissance pondérale

(les lignes représentent de haut en bas, le 97,5^e percentile, la médiane et le 2,5^e percentile)

4.4.2.2 Sélection et évaluation du modèle

Le DIC du modèle n'incluant pas les covariables caractérisant le statut tabagique des mères pendant la grossesse était de 121013. Cette valeur était plus faible pour le modèle incluant ces covariables pour lequel le DIC était de 120997. L'inclusion de l'information sur le statut tabagique améliorerait donc la qualité du modèle pour l'évaluation de la croissance des enfants ce qui peut être interprété comme une association significative entre tabagisme maternel et croissance staturale et/ou pondérale. Les valeurs du WAIC de 121669 pour le modèle n'incluant pas et de 121646 pour le modèle incluant les covariables de tabagisme confirmaient ces résultats.

Considérant ce modèle, la p-value prédictive a posteriori était de 0,48 pour la taille et 0,60 pour le poids (c'est à dire que la probabilité que la statistique calculée soit supérieure avec les données répliquées à celle obtenue avec les données observées était respectivement de 0,48 et 0,60 pour le poids et la taille) indiquant une bonne capacité prédictive du modèle (figure 30).

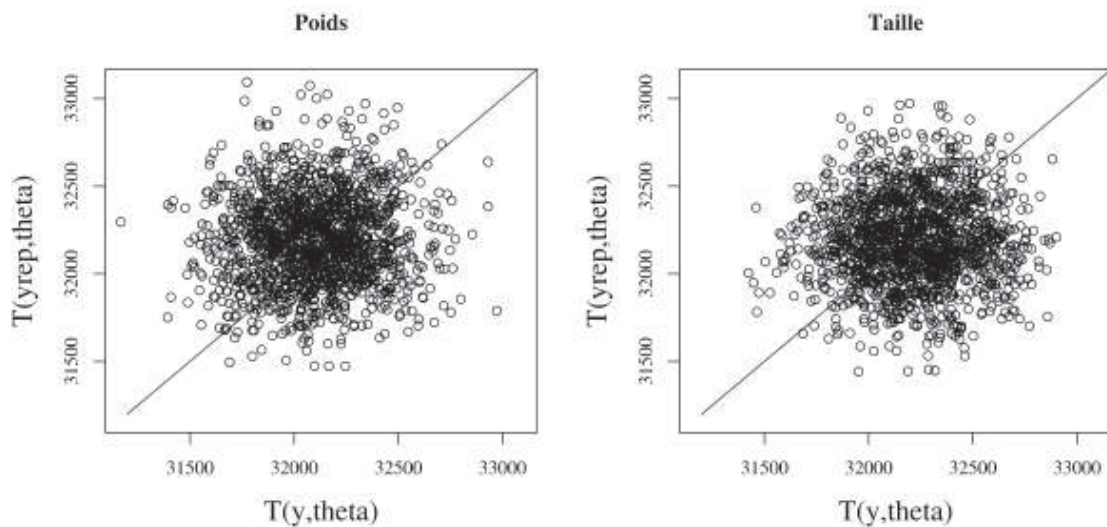


Figure 30 : Mesures de discrédance pour le poids et la taille pour les données répliquées en fonction des données observées

La figure 31 et la figure 32 décrivent, pour 6 enfants sélectionnés aléatoirement, les mesures de poids et de taille collectées dans le cadre de l'étude et leurs trajectoires de croissance moyennes prédites par le modèle (moyenne de la distribution a posteriori) associées à l'intervalle de crédibilité à 95%. Cela fournit une indication visuelle de la capacité du modèle à s'ajuster de manière adéquate aux données observées des croissances pondérale et staturale.

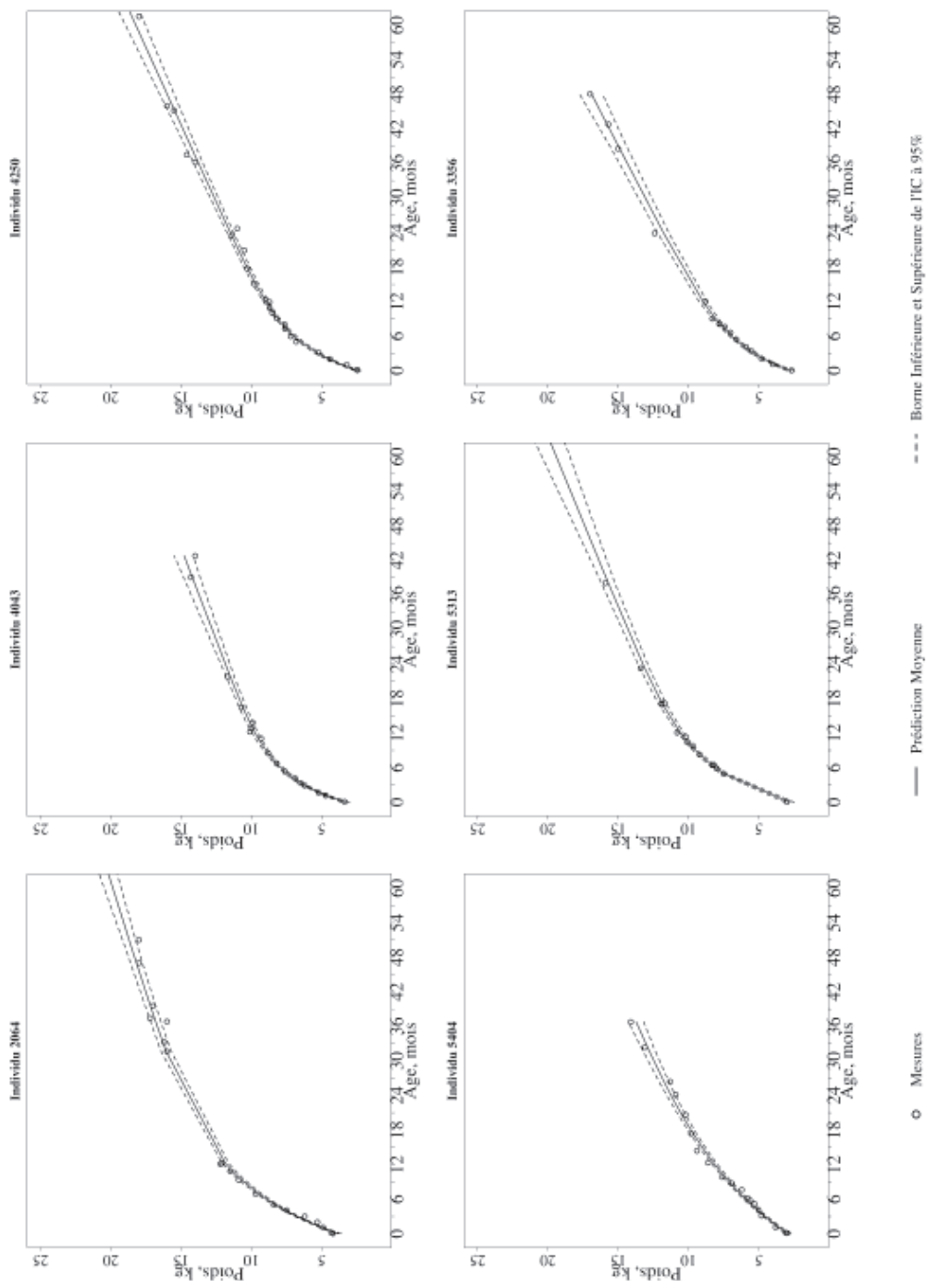


Figure 31 : Valeurs observées de poids, courbes moyennes ajustées par le modèle conjoint bayésien et intervalles de crédibilité à 95% de 6 enfants sélectionnés aléatoirement

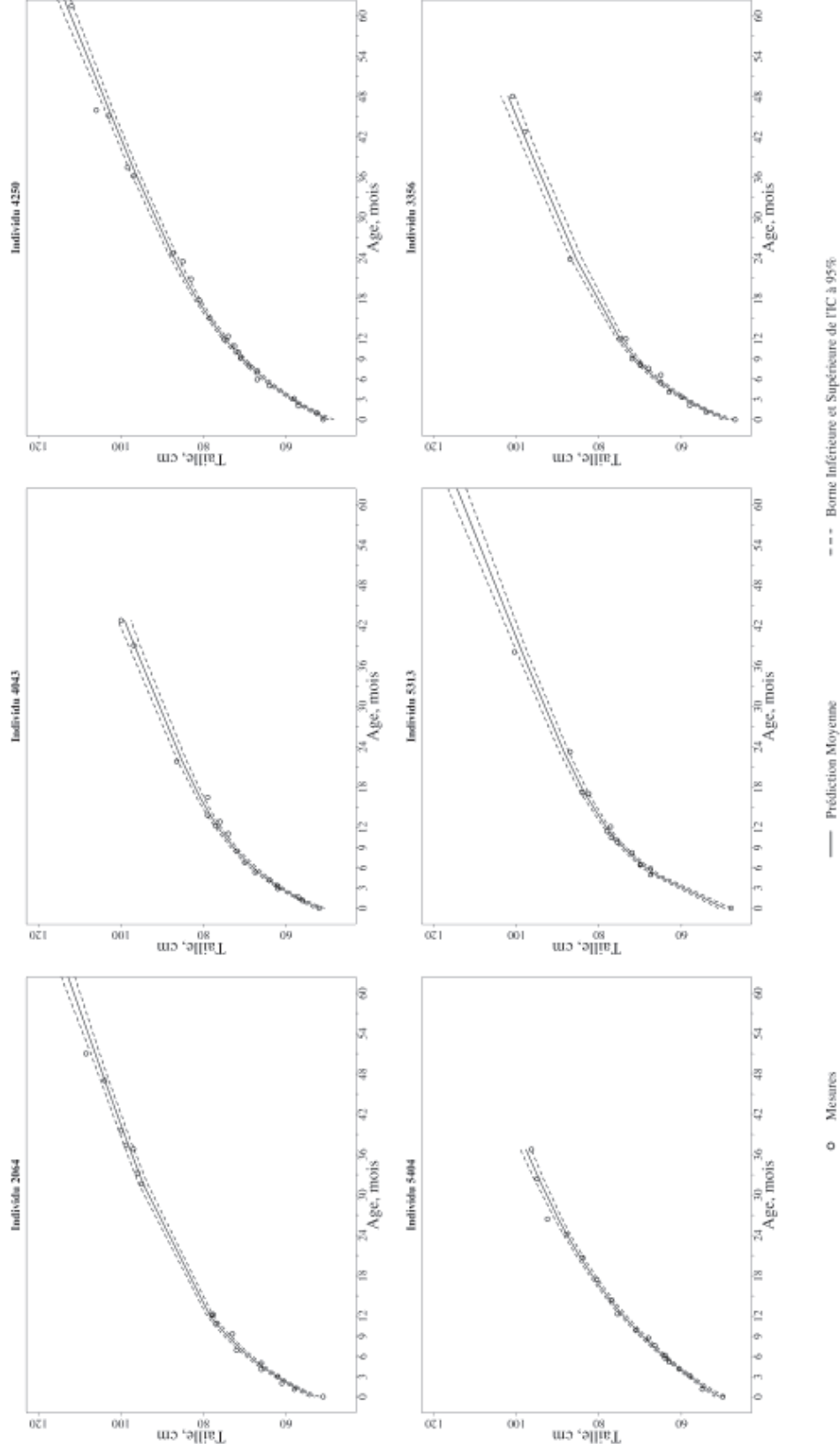


Figure 32 : Valeurs observées de taille, courbes moyennes ajustées par le modèle conjoint bayésien et intervalles de crédibilité à 95% de 6 enfants sélectionnés aléatoirement

4.4.2.3 Analyse des associations

Les 3 chaînes de Markov utilisées pour la modélisation ayant convergé vers la distribution a posteriori des paramètres que l'on souhaitait estimer, les distributions ont pu être caractérisées. Le tableau 25 illustre les associations entre le tabagisme de la mère au 1^{er} trimestre exclusivement et le tabagisme maternel tout au long de la grossesse et les principaux paramètres décrivant la croissance du poids et de la taille entre la naissance et 5 ans.

Tableau 25 : Coefficients de régressions associés au tabagisme et aux paramètres de croissance^a

Variable	A		B		C		D	
	β_A	IC 95%	Vitesse de croissance -Enfance-	IC 95%	Poussée de croissance -Premiers mois-	IC 95%	Degré de courbure -Premiers mois-	IC 95%
Taille								
Exclusivement le 1 ^{er} trimestre	-0,004	[-0,012 ; 0,005]	0,027	[-0,002 ; 0,057]	-0,021	[-0,057 ; 0,015]	0,072	[0,004 ; 0,136]
Tout au long de la grossesse	-0,015	[-0,021 ; -0,008]	0,021	[-0,002 ; 0,043]	0,005	[-0,023 ; 0,033]	0,039	[-0,012 ; 0,092]
Poids								
Exclusivement le 1 ^{er} trimestre	0,006	[-0,028 ; 0,040]	-0,001	[-0,050 ; 0,048]	0,014	[-0,042 ; 0,067]	-0,014	[-0,097 ; 0,067]
Tout au long de la grossesse	-0,037	[-0,061 ; -0,012]	0,034	[0,000 ; 0,068]	0,009	[-0,032 ; 0,052]	0,022	[-0,035 ; 0,080]

Abréviations: IMC, Indice de Masse Corporelle, IC, Intervalle de Crédibilité.

^aAjustées sur le niveau d'études, l'IMC et l'âge à l'accouchement de la mère; la durée d'allaitement; le centre de recrutement; le sexe de l'enfant.

* $P < 0,05$.

Les femmes fumant tout au long de la grossesse donnaient naissance à des enfants ayant un paramètre A_T plus faible c'est-à-dire une taille à la naissance plus faible ($\beta_{A_T} = -0,015$, *Intervalle de Crédibilité à 95%* [-0,021; -0,008]) et un paramètre A_P plus faible ($\beta_{A_P} = -0,037$, *Intervalle de Crédibilité à 95%* [-0,061; -0,012]). Le tabagisme de la mère tout au long de la grossesse était également associé à un paramètre B_P plus élevé indiquant une vitesse de croissance du poids plus élevée à partir de 2 ans ($\beta_{B_P} = 0,034$, *Intervalle de Crédibilité à 95%* [0,000 ; 0,068]). Le tabagisme de la mère exclusivement durant le 1^{er} trimestre était associé à un paramètre D_T plus élevé indiquant un degré de courbure de la trajectoire de croissance staturale plus important dans les premiers mois de vie correspondant à un rattrapage de la taille chez ces enfants ($\beta_{D_T} = 0,072$, *Intervalle de Crédibilité à 95%* [0,004 ; 0,136]).

Des associations plus modérées ont été observées avec plusieurs paramètres de croissance. En effet, les distributions a posteriori des coefficients de régression associés aux deux groupes de tabagisme maternel (exclusivement au 1^{er} trimestre ou tout au long de la grossesse) étaient décalées vers les valeurs positives pour le paramètre B_T . Le coefficient de régression pour le paramètre D_T de la taille était également décalé vers les valeurs positives pour le groupe de tabagisme maternel tout au long de la grossesse. La distribution a posteriori du coefficient de régression associé au paramètre C_T était décalée vers les valeurs négatives pour le tabagisme maternel durant le 1^{er} trimestre de grossesse exclusivement (figure 33). Ces résultats sont en faveur d'une vitesse de croissance staturale accrue après 2 ans chez les enfants exposés comparés aux enfants non exposés (pour B_T), un gain statural de la naissance à 2 ans accru chez les enfants exposés en début de grossesse (pour C_T) et un rattrapage de la taille chez les enfants des mères fumeuses tout au long de la grossesse (pour D_T).

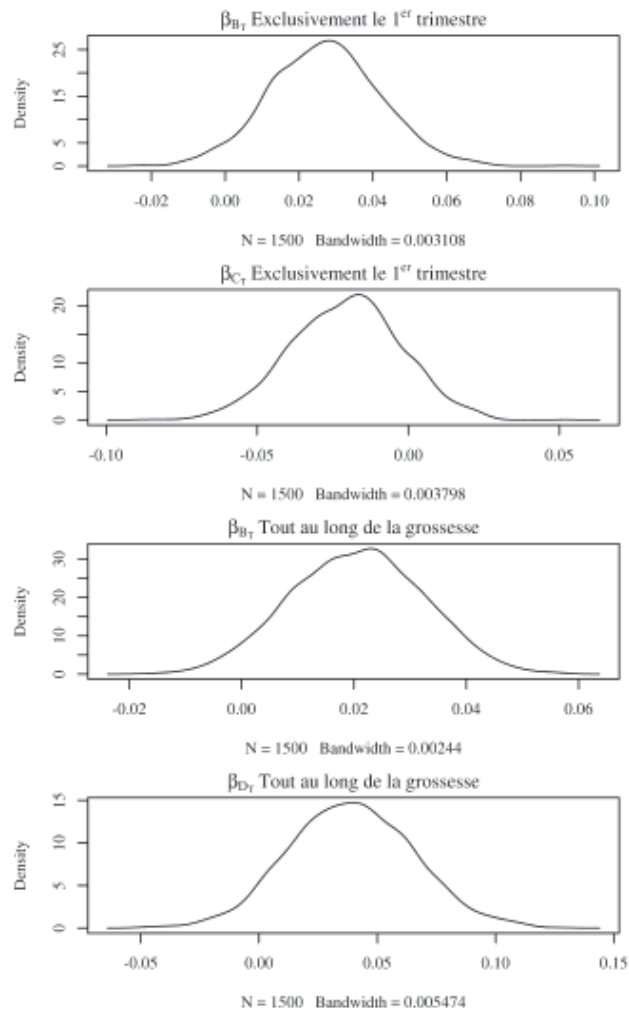


Figure 33 : Distribution a posteriori des coefficients de régression associés de manière modérée au tabagisme maternel

4.4.2.4 Sensibilité aux lois a priori

Les résultats obtenus en choisissant comme lois a priori sur les paramètres des lois peu informatives étaient très similaires à ceux obtenus avec le modèle initial (Tableau 26).

Tableau 26 : Sensibilité des résultats aux choix des lois a priori

Variable	A		B		C		D	
	Taille à la naissance		Vitesse de croissance		Poussée de croissance		Degré de courbure	
	Extrapolation du poids de naissance		-Enfance-		-Premiers mois-		-Premiers mois-	
	β_A	IC 95%	β_B	IC 95%	β_C	IC 95%	β_D	IC 95%
Taille								
Exclusivement le 1 ^{er} trimestre								
<i>Modèle Initial</i>	-0,004	-0,012; 0,005	0,027	-0,002; 0,057	-0,021	-0,057; 0,015	0,072	0,004; 0,136
<i>Sensibilité aux lois a priori</i>	-0,003	-0,012; 0,005	0,028	-0,001; 0,055	-0,021	-0,055; 0,014	0,072	0,007; 0,138
Tout au long de la grossesse								
<i>Modèle Initial</i>	-0,015	-0,021; -0,008	0,021	-0,002; 0,043	0,005	-0,023; 0,033	0,039	-0,012; 0,092
<i>Sensibilité aux lois a priori</i>	-0,015	-0,021; -0,008	0,021	-0,003; 0,045	0,005	-0,026; 0,031	0,040	-0,011; 0,092
Poids								
Exclusivement le 1 ^{er} trimestre								
<i>Modèle Initial</i>	0,006	-0,028; 0,040	-0,001	-0,050; 0,048	0,014	-0,042; 0,067	-0,014	-0,097; 0,067
<i>Sensibilité aux lois a priori</i>	0,006	-0,026; 0,039	0,002	-0,043; 0,046	0,012	-0,041; 0,062	-0,011	-0,088; 0,067
Tout au long de la grossesse								
<i>Modèle Initial</i>	-0,037	-0,061; -0,012	0,034	0; 0,068	0,009	-0,032; 0,052	0,022	-0,035; 0,080
<i>Sensibilité aux lois a priori</i>	-0,037	-0,063; -0,01	0,035	-0,001; 0,071	0,008	-0,033; 0,046	0,024	-0,035; 0,081

Abréviations: IC; Intervalle de Crédibilité

^aAjusté sur le niveau d'études; l'IMC et l'âge à l'accouchement de la mère; la durée d'allaitement; le centre de recrutement; le sexe de l'enfant

4.4.2.5 Différences de poids, taille et IMC selon le statut tabagique

Pour le jeu de covariables déterminé, l'IMC, la taille et le poids des enfants exposés au tabagisme durant le 1^{er} trimestre exclusivement ne différaient pas de ceux des enfants non exposés au tabagisme maternel à aucun des âges de prédiction (Tableau 27). Les enfants de femmes fumeuses tout au long de leur grossesse avaient une taille à 5 jours plus faible ($\beta=-0,71$ cm, *Intervalle de Crédibilité à 95%* [-1,03 ; -0,38]) et un poids à 5 jours plus faible ($\beta=-103$ grammes, *Intervalle de Crédibilité à 95%* [-176 ; -29]) comparés aux enfants de femmes non fumeuses. Leur IMC moyen n'était pas très différent ($\beta=-0,05$ kg/m², *Intervalle de Crédibilité à 95%* [-0,23 ; 0,12]). En moyenne, ces enfants rattrapaient les trajectoires de croissance pondérale et staturale des enfants de femmes non fumeuses respectivement entre 4 et 6 mois et vers 4 ans.

Les enfants de femmes fumeuses tout au long de la grossesse avaient un poids plus élevé à la fin de la période d'étude, comparé aux enfants de femmes ne fumant pas. A 4 et 5 ans, ils avaient respectivement, un poids plus élevé de 231 et 301 grammes ($\beta= 231$ et 301 grammes, *Intervalles de Crédibilité à 95%* [-36 ; 505] et [-24 ; 639]). Ces enfants avaient alors tendance à être plus petit et plus lourds comparés aux enfants de mères non fumeuses. Par conséquent, entre 570 et 600 jours (19-20 mois), l'IMC des enfants de mères fumeuses tout au long de la grossesse surpassait l'IMC des enfants de

mères non fumeuses et cette différence atteignait 0,21 kg/m² (Intervalle de Crédibilité à 95% [0,03 ; 0,40]) à 1825 jours (5 ans). Cette différence s'expliquait par une tendance de ces enfants à avoir un poids supérieur à celui des non exposés et une taille plus faible sur la plus grande partie de la période d'étude, leur déficit de taille n'étant plus notable à la fin de la période étudiée (Figure 34). Les prédictions par sous-groupes sont disponibles pour l'ensemble des âges considérés en Annexe 4.

Tableau 27 : Différences moyennes de poids, taille et d'IMC entre les enfants exposés au tabagisme maternel durant le 1^{er} trimestre exclusivement ou tout au long de la grossesse et les enfants non exposés de la naissance à 5 ans

Variable	Tabagisme Maternel			
	1 ^{er} Trimestre Exclusivement		Tout au long de la grossesse	
	Coefficient a posteriori	Intervalle de Crédibilité à 95%	Coefficient a posteriori	Intervalle de Crédibilité à 95%
Différence de poids moyen; g				
5 jours	19	-77;119	-103	-176;-29
30 jours (1 mois)	20	-81;124	-72	-150;6
90 jours (3 mois)	29	-100;160	-29	-129;69
180 jours (6 mois)	45	-107;201	2	-115;116
270 jours (9 mois)	58	-114;231	18	-112;146
365 jours (1 an)	69	-122;262	31	-112;175
730 jours (2 ans)	88	-155;328	90	-90;269
1095 jours (3 ans)	94	-194;384	158	-58;376
1470 jours (4 ans)	97	-255;449	231	-36;505
1825 jours (5 ans)	98	-323;520	301	-24;639
Différence de taille moyenne; cm				
5 jours	-0,15	-0,57;0,27	-0,71	-1,03;-0,38
30 jours (1 mois)	-0,02	-0,42;0,37	-0,60	-0,90;-0,29
90 jours (3 mois)	0,16	-0,24;0,55	-0,42	-0,73;-0,11
180 jours (6 mois)	0,23	-0,2;0,67	-0,29	-0,61;0,04
270 jours (9 mois)	0,20	-0,24;0,65	-0,24	-0,56;0,09
365 jours (1 an)	0,14	-0,32;0,60	-0,22	-0,56;0,12
730 jours (2 ans)	-0,01	-0,58;0,57	-0,20	-0,63;0,24
1095 jours (3 ans)	0,04	-0,60;0,69	-0,12	-0,61;0,38
1470 jours (4 ans)	0,19	-0,54;0,94	0,00	-0,56;0,58
1825 jours (5 ans)	0,37	-0,48;1,24	0,13	-0,52;0,82
Différence d'IMC moyen; kg/m²				
5 jours	0,14	-0,09;0,38	-0,05	-0,23;0,12
30 jours (1 mois)	0,08	-0,13;0,29	0,07	-0,09;0,24
90 jours (3 mois)	-0,01	-0,25;0,24	0,16	-0,04;0,36
180 jours (6 mois)	-0,02	-0,27;0,24	0,15	-0,04;0,36
270 jours (9 mois)	0,01	-0,24;0,27	0,15	-0,05;0,35
365 jours (1 an)	0,06	-0,20;0,31	0,15	-0,04;0,35
730 jours (2 ans)	0,12	-0,12;0,36	0,19	0,02;0,37
1095 jours (3 ans)	0,09	-0,13;0,31	0,21	0,05;0,38
1470 jours (4 ans)	0,03	-0,19;0,25	0,22	0,05;0,39
1825 jours (5 ans)	-0,02	-0,25;0,21	0,21	0,03;0,40

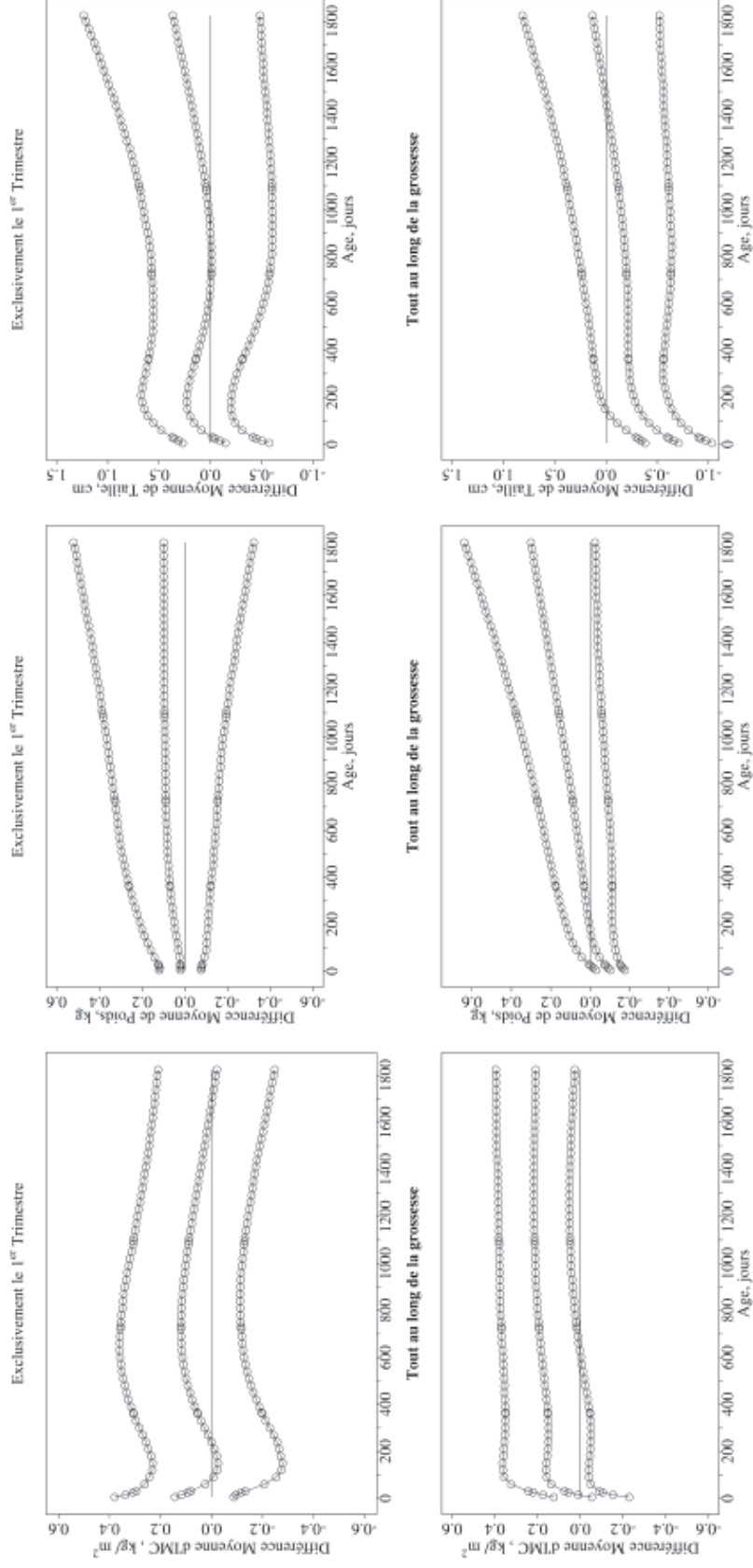


Figure 34 : Différences d'IMC, poids, taille moyens entre les enfants exposés tout au long de la grossesse ou durant le 1^{er} trimestre et la catégorie de référence (du haut vers le bas les lignes représentent la borne supérieure de l'intervalle de crédibilité à 95%, la moyenne et la borne inférieure de l'intervalle de crédibilité à 95%)

4.5 Discussion

Nous avons illustré dans cette partie de quelle manière les croissances du poids et de la taille de la naissance à 5 ans pouvaient être modélisées conjointement avec une approche bayésienne. L'étude de simulation montre que cette modélisation conjointe fournit des estimations non biaisées et une bonne probabilité de couverture ce qui reflète la capacité du modèle à s'ajuster correctement aux données de croissance. Cette approche a été utilisée comme une étape de validation du modèle dans le but d'évaluer les estimations faites par le modèle sur des données de croissance théoriques. Il faut toutefois souligner, que nous nous sommes placés dans des conditions optimales. En effet, cette étude a été réalisée en fixant un nombre de sujets et de mesures par sujet important. Les bonnes propriétés du modèle pourraient être différentes en se plaçant dans un contexte moins favorable et plus proche des conditions réelles de certaines études longitudinales moins riches en mesures des poids et taille (présence de données manquantes, de perdus de vue...).

Cette approche offre un moyen souple pour faire des prédictions sur les poids, taille et IMC dans l'enfance et étudier les déterminants potentiels de la croissance tels que des profils socio-démographiques ou des expositions environnementales. Dans notre exemple, le modèle conjoint a permis d'étudier les trajectoires de croissance et prédire les différences moyennes au cours du temps pour les différents sous-groupes de tabagisme maternel et de déterminer les intervalles de crédibilité des trajectoires de croissance du poids, de la taille mais également de l'IMC, pour un jeu de covariables donné. Il fournit des informations clés sur la dynamique des associations. Un des avantages de la modélisation conjointe, comme cela a été décrit précédemment, est qu'elle n'impose pas que les différents événements étudiés (ici la taille et le poids) soient mesurés aux mêmes temps et par conséquent que le même nombre de mesures soient disponibles pour ces événements (Fieuws and Verbeke, 2004a). De ce fait, l'ensemble de l'information disponible sur le poids et la taille peut être utilisé pour prédire les trajectoires d'IMC.

La modélisation de l'IMC directement aurait limité les analyses aux 26693 données d'IMC disponibles (jours où à la fois une mesure de poids et de taille est disponible) alors que les 30515 et 28381 mesures isolées de poids et de taille respectivement ont pu être utilisées dans cette analyse.

Une des principales forces de cette approche réside également dans le fait qu'elle permet, en 1 étape, basée sur la modélisation conjointe du poids et de la taille, de faire des inférences sur l'IMC et d'obtenir des intervalles de crédibilités sur les trajectoires d'IMC de chaque sous-groupe de tabagisme à statut fixé vis-à-vis des autres covariables, et sur les différences de trajectoires entre les sous-groupes d'intérêt en utilisant les distributions prédictives a posteriori des paramètres estimés. Cela constitue une différence importante entre l'approche fréquentiste et l'approche bayésienne. Dans la première approche, les paramètres de croissance de la taille et du poids peuvent être utilisés pour estimer l'IMC dans les différents sous-groupes d'intérêt et pour estimer les différences entre sous-groupes. Cependant, l'incertitude sur les paramètres est ignorée et de ce fait, les intervalles de confiance des différences ne peuvent pas être évalués en une étape comme avec l'approche bayésienne.

A notre connaissance, une seule étude récemment publiée a utilisé une approche similaire, en modélisant conjointement plusieurs caractéristiques de la croissance embryogéniques dans un contexte bayésien en utilisant une généralisation du modèle SITAR qui, comme l'équation de Jentsch-Bayley, modélise la croissance avec des paramètres spécifiques de chaque sujet ayant une signification biologique. Ce modèle a été utilisé pour étudier la croissance pré-natale mais pourrait être utilisé pour étudier la croissance postnatale.

Dans le travail réalisé ici, un modèle assez simple a été considéré en considérant des hypothèses standards. Cependant, les hypothèses sur les distributions des données du modèle auraient pu être modifiées ou complexifiées. La fonction de Jentsch-Bayley a été reparamétrisée. De ce fait, la distribution des paramètres au niveau individuel est supposée être lognormale plutôt que normale. La forme initiale aurait, par conséquent, pu être considérée :

$$f(\phi_{ik}, t_{ijk}) = A_{ik} + B_{ik}t_{ijk} + \exp(C_{ik}) [1 - \exp(D_{ik}t_{ijk})]$$

Avec des hypothèses de normalité sur les paramètres $\phi_{ik} = (A_{ik}, B_{ik}, C_{ik}, D_{ik})^T$.

Cependant, compte tenu des problèmes de convergence rencontrés en utilisant la forme initiale, malgré l'utilisation de distributions a priori normales tronquées en zéro pour les paramètres, la forme modifiée (3) a été conservée pour le modèle final.

Dans ce travail, nous n'avons pas supposé de corrélation ($\sigma_{e_{TP}} = 0$) entre les erreurs résiduelles de la taille e_{ijT} et du poids e_{ijP} mesurées à la même occasion de mesure j pour le même individu i . Une structure de corrélation aurait pu être spécifiée

entre les erreurs résiduelles du poids et de la taille. L'hypothèse d'indépendance des erreurs résiduelles faite dans ce travail semble toutefois raisonnable.

Dans ce travail, nous avons sélectionné les meilleures fonctions permettant de décrire les variances résiduelles du poids et de la taille parmi deux fonctions candidates ($\log(\sigma_{e_{ijk}}^2) = \delta_{0k} + \delta_{1k}t_{ijk}$ et $\log(\sigma_{e_{ijk}}^2) = \delta_{0k} + \delta_{1k}t_{ijk} + \delta_{2k}t_{ijk}^2$). Cependant, d'autres fonctions auraient pu être considérées et nous n'avons pas procédé à une comparaison formelle avec ces autres possibilités. Ainsi, nous ne pouvons pas exclure que d'autres fonctions amélioreraient le modèle choisi. Cette structure de variance hétérogène pourrait également être complexifiée en faisant par exemple varier les variances résiduelles en fonction de covariables susceptibles d'avoir un effet sur celle-ci.

L'hypothèse de normalité faite sur les erreurs résiduelles pourrait également être évaluée. Il a été proposé de considérer une distribution de Student pour les données ayant une distribution à longue queue ou ayant une distribution normale mais « contaminée » par des données extrêmes, ou plus généralement une distribution d'erreur généralisée pour toute sorte de phénomène d'aplatissement ou kurtosis (Zhang, 2013).

Dans ce travail, nous avons bénéficié des résultats d'études précédentes en spécifiant des distributions a priori sur les paramètres à estimer. En ce qui concerne la variance inter-individuelle, une distribution de Wishart a été spécifiée, mais cela implique que tous les éléments de la matrice de variance-covariance aient la même incertitude. Cette restriction peut être évitée en décomposant la matrice de variance-covariance Ω en $\Omega = S\tilde{\Omega}S$ où S est une matrice diagonale de paramètres d'échelle s_l pour $l=1,\dots,2m$ (matrice d'écart-types) et $\tilde{\Omega}$ une matrice de corrélation qui suit une distribution de Wishart inverse. Choisir $\tilde{\Omega} \sim IW(I, \nu)$ avec $\nu=2m+1$ équivaut à supposer une distribution uniforme sur chaque coefficients de corrélations (Gelman and Hill, 2006; John Barnard et al., 2000). Les lois a priori sur les paramètres d'échelle sont choisies selon les mêmes règles que pour les écart-types (e.g: uniforme, lognormale...). Ainsi, cette loi de distribution autorise une certaine flexibilité en spécifiant des informations a priori sur les différentes composantes de la variance en conservant les propriétés de la distribution inverse Wishart sur la distribution des coefficients de corrélation. Cependant, sur un sous-échantillon de 50 enfants issus de notre population d'étude, le modèle utilisant cette loi de distribution, ne convergeait pas. Ceci peut s'expliquer par le calibrage délicat des distributions a priori pour les paramètres

d'échelle. De plus, l'influence des distributions a priori devrait être négligeable compte tenu du large effectif de données de poids et de taille dont nous disposons.

Les facilités calculatoires apportées par l'approche bayésienne s'ajoutent aux avantages décrits ci-dessus et sont une des raisons principales qui nous a conduit à l'évaluer. Ici, les associations entre le tabagisme maternel et des dimensions multiples de la croissance ont été évaluées en prenant en compte plusieurs facteurs de confusion potentiels. Ce modèle pourrait être complexifié en ajoutant des covariables supplémentaires. Cependant, le modèle est déjà complexe et le temps de calcul non négligeable. L'utilisation de modèles plus complexes pourrait par conséquent être limitée par des aspects pratiques.

4.6 Perspectives

Les modèles multivariés ont été décrits pour analyser les associations entre plusieurs processus se déroulant en parallèle (Goldstein, 1986; MacCallum et al., 1997; Pickles and Croudace, 2010; Suvak et al., 2009). Des modèles utilisant des splines ont notamment été utilisés pour décrire de quelle manière les changements observés pour un événement précédent les changements affectant un autre événement, mais des problèmes de convergence étaient rapportés comme une difficulté majeure (Macdonald-Wallis et al., 2012). Nous pourrions étendre le modèle conjoint bayésien que nous avons utilisé dans ce travail pour collecter des informations sur la structure des associations entre le poids et la taille et de ce fait évaluer s'il y a des fenêtres où les corrélations entre le poids et la taille diffèrent et si les changements de poids observés précèdent les changements de taille et vice versa.

Chapitre 5 : Analyse de la modélisation conjointe entre croissances pondérale et staturale

5.1 Introduction et objectifs spécifiques

Dans le chapitre 3, deux méthodes ont été utilisées pour étudier l'effet du tabagisme maternel comme déterminant des caractéristiques de la croissance. Dans la méthode « en 1 étape », la variable a été directement incluse dans les modèles non linéaires à effets mixtes. Cette modélisation directe permet d'incorporer l'incertitude sur l'estimation des paramètres dans l'analyse des associations. Dans la méthode en « 2 étapes », la modélisation a été effectuée sans la variable « tabagisme maternel » et les caractéristiques d'intérêt ont été prédites afin d'étudier, dans un second temps, leurs associations avec la variable d'intérêt. Dans cette deuxième méthode, l'incertitude sur les prédictions individuelles, liée à la qualité de l'estimation, n'est pas considérée dans les analyses ; cela a théoriquement peu d'importance lorsque la variance intra-sujet est faible par rapport à la variance inter-sujets. Or, cette hypothèse n'a pas été formellement vérifiée et pourrait avoir des conséquences sur les conclusions faites sur les associations entre les caractéristiques anthropométriques prédites et les facteurs étudiés. En effet, si la variance inter-sujets est grande par rapport à la variance intra-sujets, il est probable que la différence inter-sujets soit expliquée par un facteur différenciant les sujets inclus dans les analyses. Si la variance inter-sujets n'est pas plus grande que la variance intra-sujet, les différences observées pourraient principalement être expliquées par ces fluctuations et donc liées aux hasard plutôt qu'être expliquées par un facteur discriminant, comme le tabagisme maternel.

Par ailleurs, un des bénéfices attendus ayant motivé la mise en place d'un modèle conjoint de la croissance pondérale et staturale était que, dans le cas où des données manquantes sont présentes pour l'une des variables, l'autre pourrait apporter une information permettant d'améliorer les prédictions. Ainsi, même en l'absence de donnée disponible pour une variable seulement, la trajectoire de croissance pourrait être prédite grâce aux données de la variable modélisée conjointement. Il a également été supposé que les prédictions des trajectoires de croissance de taille (poids) devraient être améliorées par la prise en compte à la fois de l'information apportée par les mesures de taille (poids) et par celle apportée par les mesures de poids (taille).

Le **premier objectif** de cette partie était donc d'utiliser le modèle bayésien modélisant conjointement la croissance pondérale et staturale des enfants au cours du temps pour analyser les variances des prédictions individuelles de poids et de taille

(variance intra-sujet) afin de les comparer aux variances inter-sujets aux différents temps de prédiction considérés.

Le **deuxième objectif** de cette partie était d'étudier l'impact de la présence de mesures de poids dans l'estimation des prédictions de taille dans un modèle conjoint pour pouvoir vérifier l'hypothèse selon laquelle la qualité des prédictions est améliorée.

5.2 Population et méthodes

5.2.1 Population d'étude

Ces analyses ont été restreintes à la sous-population des garçons pour des questions de temps de calcul et car il n'y avait a priori pas de raison d'observer de différence selon le sexe. La population source était donc constituée de l'ensemble des garçons pour lesquels au moins 2 données de poids et/ou 2 données de taille étaient disponibles. Au total, 942 garçons avaient au moins 2 mesures de poids (représentant un total de 16751 mesures) et 917 avaient au moins 2 mesures de taille (représentant un total de 15567 mesures).

5.2.2 Méthode

Comparaison des variances intra- et inter-sujets

Pour ce travail, aucune covariable n'a été considérée dans les modèles. Afin de réaliser cette comparaison, les résultats d'un modèle conjoint entre le poids et la taille (suivant la méthode décrite au chapitre 4, p.135) et ceux des modèles décrivant le poids et la taille séparément ont été exploités. Certains enfants avaient deux mesures de poids et ont donc été inclus dans la population d'étude mais n'avaient aucune mesure de taille (n=25). Dans la modélisation conjointe, les mesures de poids disponibles ont pu être utilisées pour prédire les croissances staturale même en l'absence de données de taille. Ainsi, les trajectoires de croissance staturale de l'ensemble des enfants ont pu être prédites. Dans le cas de la modélisation univariée de la taille, les trajectoires de croissance staturale de ces 25 enfants n'ont pas pu être ajustées. Quatre modélisations ont donc été effectuées pour ces analyses : une modélisation conjointe incluant l'ensemble des enfants, y compris les 25 enfants sans mesure de taille, appelée par la suite « Modélisation conjointe » ; une modélisation conjointe excluant les 25 enfants sans mesure de taille, appelée par la suite « Modélisation conjointe excluant les sujets

sans mesures de taille », afin de se ramener à la population disponible pour la modélisation univariée de la taille; une modélisation univariée du poids, appelée « modélisation du poids »; une modélisation univariée de la taille, appelée « modélisation de la taille ».

Pour ces modèles, trois chaînes ont été lancées en parallèle avec des valeurs initiales différentes; une phase de chauffe de 20000 itérations a été fixée; les échantillons utilisés pour les inférences ont été recueillis toutes les 20 itérations sur un intervalle de 20000 itérations.

Les poids, taille et IMC individuels et les vitesses correspondantes ont été prédits à différents âges : 3 mois et 1, 2, 3, 4 et 5 ans. L'espérance de la variance intra-sujet a été obtenue pour chaque caractéristique anthropométrique de chaque sujet et chaque temps de prédiction à partir de la distribution a posteriori de chaque prédiction individuelle. Nous disposons ainsi de la distribution de l'espérance des variances intra-sujet dans l'échantillon et nous en avons calculé la moyenne.

A chaque âge considéré, à partir des valeurs prédites pour chaque sujet, nous avons calculé la variance (inter-sujets) des variables poids, taille et IMC et des vitesses correspondantes.

Le rapport de l'estimation de la variance inter-sujets et de la moyenne des espérances des variances intra-sujet a été calculé à chaque temps et pour chaque caractéristique afin d'étudier la plausibilité de l'hypothèse considérant une variance intra-sujet faible par rapport à la variance inter-sujets.

Apport de l'information apportée par les mesures de poids sur les prédictions de croissance staturale

Afin de répondre à cet objectif, 20 % des données de taille ont été sélectionnées aléatoirement dans l'échantillon d'étude, soit 3113 mesures parmi les 15567 initialement disponibles et ont été codées comme données manquantes. Les mesures de poids (cf. paragraphe 5.2.1 de ce chapitre) ont été maintenues à l'identique.

Le modèle décrivant conjointement la croissance pondérale et staturale a été utilisé pour prédire les mesures de taille supprimées et estimer la variance de ces prédictions.

Une modélisation de la croissance staturale séparément (univariée), n'utilisant pas les mesures de poids, a été utilisée pour prédire les mêmes informations.

Les différences entre la valeur réelle des mesures supprimées et les prédictions moyennes obtenues avec le modèle conjoint d'une part et le modèle univarié d'autre part ont été comparées.

5.3 Résultats

Comparaison des variances intra- et inter-sujets

Le Tableau 28 présente l'estimation des moyennes des variances intra-sujet et les variances inter-sujets pour l'ensemble des mesures prédites de poids, taille et IMC. Pour le modèle conjoint ainsi que pour les modèles univariés du poids et de la taille, la moyenne des variances intra-sujet était plus faible que la variance inter-sujets pour l'ensemble des mesures prédites. Les rapports de la variance inter- sur la moyenne des variances intra-sujets étaient compris entre 1,6 et 19,9 (Tableau 28).

Pour le poids (cf. « Modélisation du poids », Tableau 28), la moyenne des variances intra-sujets, pour chaque âge de prédiction, était un peu plus faible en utilisant le modèle conjoint plutôt que le modèle univarié (e.g. respectivement 0,037 *versus* 0,041 à 3 mois et 0,57 *versus* 0,58 à 5 ans). Les variances inter-sujets étaient un peu plus élevées avec le modèle conjoint qu'avec le modèle univarié (e.g. respectivement 0,39 *versus* 0,38 à 3 mois et 4,59 *versus* 4,53 à 5 ans).

Pour la taille, la moyenne des variances intra-sujets était, pour chaque âge de prédiction, plus élevée lorsque la taille était modélisée avec le poids sur l'échantillon complet comprenant les 25 enfants sans mesures de taille (cf. « Modélisation Jointe », Tableau 28) comparé à lorsqu'elle était modélisée seule (cf. « Modélisation de la Taille », Tableau 28). Les variances étaient respectivement de 0,29 et 0,28 à 3 mois ; de 2,41 et 2,17 à 5 ans). Ces résultats, surprenant d'un premier abord, étaient expliqués par la présence des 25 individus n'ayant aucune donnée de taille dans le modèle conjoint et dont les prédictions ont été faites en utilisant les mesures de poids uniquement. L'imprécision des prédictions obtenues pour ces individus augmentait la moyenne des variances intra-sujet dans le modèle conjoint. Après suppression de ces 25 individus dans les estimations (cf. « Modélisation Jointe Excluant les sujets sans valeurs de Taille », Tableau 28), des résultats similaires à ceux observés pour le poids étaient obtenus. La moyenne des variances intra-sujets était plus faible dans le modèle conjoint que dans le modèle statural univarié (e.g. respectivement 0,23 *versus* 0,28 à 3 mois et 1,98 *versus* 2,17 à 5 ans). Les variances inter-sujets étaient également plus élevées ou comparables

avec le modèle conjoint par rapport au modèle univarié (e.g. respectivement 4,44 *versus* 4,39 à 3 mois et 17,3 pour les deux à 5 ans).

On peut noter que pour l'IMC, comme attendu, la moyenne des variances intra-sujets étaient réduites après exclusion des individus sans valeurs de taille du calcul des estimations. Les résultats observés sur les prédictions de poids, taille et IMC sont directement liés à la précision sur les paramètres des modèles de croissance. Les prédictions étaient par conséquent également modifiées dans le même sens pour les vitesses de croissance du poids, de la taille et de l'IMC. Les variances inter-sujets étaient plus élevées lorsque le poids et la taille étaient modélisés conjointement que lorsqu'ils étaient modélisés séparément et la moyenne des variances intra-sujets plus faible (après exclusion des individus sans valeurs de taille) (Tableau 28).

Tableau 28 : Variances inter-sujets et moyennes des variances intra-sujet pour les prédictions des caractéristiques de croissance issues de la modélisation conjointe et séparée.

Variable	Modélisation Jointe			Modélisation Jointe Exclusion Sujet			Modélisation du Poids			Modélisation de la Taille		
	Sans Valeurs de Taille			Sans Valeurs de Taille								
	Variance Inter-Sujet	Variance Intra-Sujet Moyenne	Variance Inter/Intra	Variance Inter-Sujet Moyenne	Variance Intra-Sujet Moyenne	Variance Inter/Intra	Variance Inter-Sujet Moyenne	Variance Intra-Sujet Moyenne	Variance Inter/Intra	Variance Inter-Sujet Moyenne	Variance Intra-Sujet Moyenne	Variance Inter/Intra
Poids, kg												
3 mois	0,3861	0,0365	10,6				0,3806	0,0405	9,4			
1 an	0,8591	0,0557	15,4				0,8580	0,0563	15,2			
2 ans	1,4238	0,1089	13,1				1,4171	0,1127	12,6			
3 ans	2,1468	0,1891	11,4				2,1318	0,1943	11,0			
4 ans	3,1873	0,3389	9,4				3,1571	0,3460	9,1			
5 ans	4,5883	0,5736	8,0				4,5347	0,5848	7,8			
Taille, cm												
3 mois	4,3516	0,2913	14,9	4,4386	0,2297	19,3				4,3916	0,2819	15,6
1 an	5,4268	0,4065	13,4	5,5614	0,2795	19,9				5,5288	0,3068	18,0
2 ans	7,5840	0,7696	9,9	7,7803	0,5767	13,5				7,7077	0,6449	12,0
3 ans	10,0971	1,0786	9,4	10,3617	0,8184	12,7				10,2836	0,8953	11,5
4 ans	13,0063	1,5293	8,5	13,3494	1,1940	11,2				13,3245	1,2951	10,3
5 ans	16,8152	2,4113	7,0	17,2615	1,9782	8,7				17,2696	2,1741	7,9
IMC, kg/m2												
3 mois	1,1274	0,2530	4,5	1,1574	0,2226	5,2						
1 an	1,2815	0,1611	8,0	1,3160	0,1259	10,5						
2 ans	1,1254	0,1572	7,2	1,1557	0,1271	9,1						
3 ans	1,0733	0,1589	6,8	1,1022	0,1310	8,4						
4 ans	1,0889	0,1844	5,9	1,1183	0,1568	7,1						
5 ans	1,1318	0,2333	4,9	1,1623	0,2050	5,7						
Vitesse de croissance pondérale, kg/mois												
3 mois	0,0109	0,0016	6,8									
1 an	0,0025	0,0005	4,7				0,0100	0,0021	4,7			
2 ans	0,0012	0,0002	5,3				0,0025	0,0006	4,3			
3 ans	0,0013	0,0003	4,5				0,0012	0,0002	5,1			
4 ans	0,0013	0,0003	4,3				0,0012	0,0003	4,4			
5 ans	0,0013	0,0003	4,2				0,0013	0,0003	4,1			
Vitesse de croissance staturale, cm/mois												
3 mois	0,0751	0,0215	3,5	0,0769	0,0199	3,9				0,0715	0,0243	2,9
1 an	0,0154	0,0050	3,1	0,0158	0,0045	3,5				0,0156	0,0053	2,9
2 ans	0,0044	0,0013	3,4	0,0045	0,0012	3,8				0,0043	0,0013	3,3
3 ans	0,0031	0,0013	2,4	0,0032	0,0012	2,7				0,0031	0,0013	2,3
4 ans	0,0036	0,0017	2,1	0,0037	0,0016	2,3				0,0036	0,0019	1,9
5 ans	0,0039	0,0019	2,0	0,0040	0,0018	2,2				0,0039	0,0022	1,8
Vitesse de croissance de l'IMC, kg/m2/mois												
3 mois	0,0330	0,0098	3,4	0,0339	0,0090	3,8						
1 an	0,0024	0,0015	1,6	0,0024	0,0014	1,7						
2 ans	0,0008	0,0003	2,7	0,0008	0,0003	3,0						
3 ans	0,0004	0,0002	2,1	0,0004	0,0002	2,3						
4 ans	0,0002	0,0001	1,7	0,0003	0,0001	1,8						
5 ans	0,0002	0,0001	1,5	0,0002	0,0001	1,6						

Apport de l'information des mesures de poids sur les prédictions de croissance staturale

La Figure 35 illustre une trajectoire de croissance staturale ajustée par le modèle conjoint et le modèle univarié. Les barres horizontales représentent les prédictions et leurs intervalles de crédibilité à 95%. Les barres noires et rouges correspondent à des prédictions faites aux mêmes âges, le léger décalage de temps permet de mieux distinguer les estimations correspondant au modèle conjoint (représenté en rouge) et au modèle univarié (en noir). Pour ce sujet, on peut voir que les variances des estimations sont meilleures lorsque la taille est modélisée avec le poids (respectivement 0,56 ; 0,89 ; 2,53 ; 3,92 ; 5,39 et 7,74 pour les prédictions de taille à 3 mois, 1, 2, 3, 4 et 5 an(s)) que lorsqu'elle est modélisée seule (respectivement 0,76 ; 1,02 ; 3,57 ; 5,80 ; 7,63 et 10,1 pour les prédictions de taille à 3 mois, 1, 2, 3, 4 et 5 an(s)).

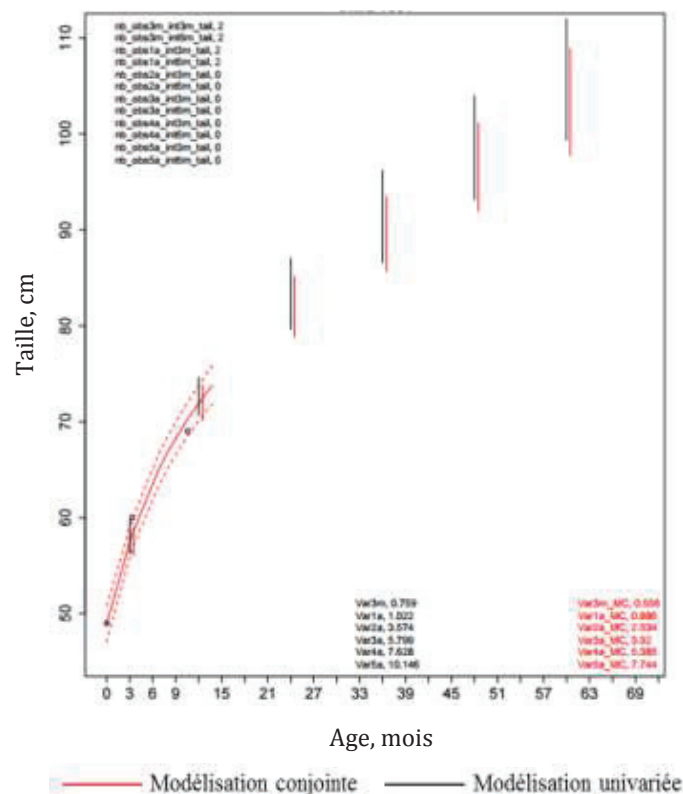


Figure 35 : Trajectoire de croissance staturale, modélisation conjointe et modélisation univariée

La Figure 36 représente les résidus des mesures de taille obtenus avec la modélisation univariée en fonction de ceux obtenus avec la modélisation conjointe pour les 20% de mesures de taille sélectionnées pour être considérées manquantes. Les

résidus se répartissaient le long de la diagonale de pente 1 ce qui indique l'absence de différence systématique entre la modélisation conjointe et la modélisation univariée .

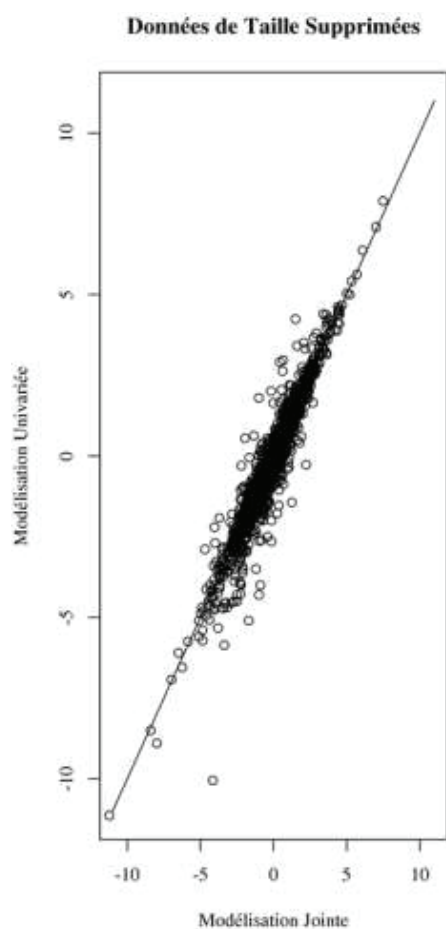


Figure 36 : Comparaison des résidus obtenus avec la modélisation conjointe poids/taille et la modélisation univariée de la taille dans l'échantillon où 20% des tailles ont été supprimés

Le Tableau 29 compare les distributions des résidus obtenues pour les 20% de mesures de taille sélectionnées pour être considérées manquantes ; (1) à partir d'une modélisation conjointe du poids et de la taille et (2) à partir d'une modélisation univariée de la taille. Il montre qu'en absence de données de taille, la variance des résidus est un peu plus faible lorsque de l'information est apportée par la trajectoire de taille et par la trajectoire de poids plutôt que lorsqu'elle est apportée uniquement par la trajectoire de taille (variance de 2,10 et 2,36 respectivement).

Tableau 29 : Distribution des résidus de taille dans le « sous-échantillon de 20% » obtenue avec la modélisation conjointe poids/taille ou avec la modélisation univariée de la taille

Données	Modélisation	Minimum	1 ^{er} quartile	Médiane	Moyenne	3 ^e quartile	Maximum	Variance
« Echantillon de 20 % » (n=3113)	conjointe	-11,23	-0,82	0,04	-0,03484	0,84	7,49	2,10
	Univariée	-12,27	-0,89	0,01	-0,06536	0,85	7,90	2,36

5.4 Discussion

Il a été mis en évidence dans ce chapitre que la variance intra-sujet des prédictions individuelles de poids, taille et IMC aux âges sélectionnés et des vitesses instantanées correspondantes étaient beaucoup plus faibles que les variances inter-sujets ce qui laisse penser que les mesures d'associations entre les déterminants d'intérêt et les caractéristiques de la croissance prédites seraient peu biaisées dans les analyses en 2 étapes. L'analyse des variances intra-sujet a également permis de mettre en évidence que l'incertitude sur les prédictions des caractéristiques de croissance staturale et pondérale était légèrement plus faible lorsque les deux profils de croissance étaient modélisés conjointement, ce qui incite à utiliser cette méthode.

Sur les 20% de mesures de taille sélectionnées pour être considérées manquantes, l'analyse des résidus montrait que l'utilisation conjointe des informations disponibles sur les trajectoires de croissance staturale et pondérale permettait de diminuer légèrement la variance résiduelle obtenue en ne prenant en compte que l'information sur la trajectoire de croissance apportée par les autres mesures de taille. Ce travail n'a pas été réalisé pour évaluer l'impact de la prise en compte de l'information de la croissance staturale pour la prédiction de la croissance pondérale, mais on peut penser que les résultats seraient similaires.

**Travail exploratoire : étude des
associations entre la croissance
pondérale et staturale et modélisation
de la matrice de variances-covariances**

6.1 Contexte

Un des objectifs de recherche de cette thèse était l'étude de la co-évolution de la croissance staturale et de la croissance pondérale. C'est en partie afin d'avoir l'opportunité d'étudier les associations entre la croissance du poids et de la taille qu'une stratégie de modélisation conjointe, modélisant simultanément les données longitudinales du poids et de la taille, a été adoptée. Cette méthode de modélisation permet en effet, notamment, de suivre l'évolution conjointe de deux processus (Henderson et al., 2000). La modélisation conjointe a constitué un champ de recherche particulièrement actif en biostatistiques ces dernières années pour modéliser conjointement des données longitudinales (e.g: un biomarqueur) et un temps d'événement (e.g: une rechute clinique) (Rizopoulos and Ghosh, 2011; tsiatis and Davidian, 2004). Cependant, elle peut être utilisée pour différents types de données et a également été utilisée pour la modélisation conjointe de données longitudinales incluant des données de croissance (Fieuws and Verbeke, 2004b; Willemsen et al., 2015). Plusieurs méthodes ont été proposées dans la littérature pour la construction de modèles conjoints. Parmi ces méthodes, les modèles conjoints à effets aléatoires partagés, permettent d'associer deux réponses d'intérêt, chacune modélisée à l'aide de modèles mixtes, en imposant une distribution conjointe de leurs effets aléatoires (Fieuws and Verbeke, 2004b; Huang et al., 2011; Li et al., 2014; Luo, 2014; Suvak et al., 2009; Thiébaud et al., 2002; Wagner and Tüchler, 2010).

Dans le chapitre 4, nous avons introduit le modèle conjoint de la croissance pondérale et staturale qui utilisait cette approche : les croissances du poids et de la taille étaient chacune modélisée par un modèle non-linéaire mixte, les deux étant reliées par une structure commune des effets aléatoires associés à chaque paramètre de croissance. La matrice de variances-covariances des effets aléatoires des paramètres de croissance du poids et de la taille fournit les éléments permettant d'étudier de quelle manière l'évolution de chaque paramètre de croissance du poids est associée à l'évolution des différents paramètres de croissance de la taille. Ce modèle a été construit selon l'hypothèse de l'existence d'une matrice de variances-covariances commune pour l'ensemble des enfants inclus dans l'étude et quelles que soient leurs caractéristiques (e.g : socio-démographiques) individuelles. Cette matrice des paramètres individuels est considérée comme suivant une loi de Wishart inverse, qui est la loi a priori la plus

commune pour les matrices de covariances, bien qu'elle présente certaines limites (cf discussion chapitre 4, p164).

Outre l'obtention d'informations sur l'évolution conjointe de la croissance pondérale et staturale, un des intérêts de cette approche est d'étudier si certaines caractéristiques (déterminants) modifient la structure des associations entre l'évolution du poids et de la taille. Certains travaux ont, de plus, montré qu'en cas d'hétérogénéité de cette matrice pour les différents sujets ou pour différents sous-groupes caractérisés par des caractéristiques spécifiques à chaque individu, les estimations des écarts-types des effets fixes et aléatoires et, par conséquent, les inférences et les prédictions des trajectoires propres à chaque individu, peuvent être altérées en traitant celle-ci comme homogène (Daniels and Zhao, 2003; Heagerty and Kurland, 2001). Compte tenu du manque de flexibilité de la loi de Wishart inverse, Pourahmadi (Pourahmadi, 1999) a proposé une méthode de modélisation de la matrice de variances-covariances des effets aléatoires et Daniels et coll l'ont illustrée (Daniels and Zhao, 2003). Cette paramétrisation fournit des paramètres directement interprétables et garantit que la matrice soit définie positive.

6.2 Objectifs spécifiques

Dans ce travail préliminaire, l'objectif était d'explorer l'approche proposée par Pourahmadi permettant de modéliser la matrice de variances-covariances des effets aléatoires des paramètres de croissance pondérale et staturale.

Cette méthode a, dans un premier temps, été explorée pour la modélisation de la matrice de variances-covariances globale des enfants, en considérant celle-ci comme homogène pour l'ensemble de la population, puis, dans un second temps, comme différant selon le statut pondéral (normal *versus* surpoids / obèses) des enfants.

6.3 Méthodes

6.3.1 Population d'étude

Le surpoids et l'obésité ont été définis selon les seuils de l'IOTF (Cole et al., 2000). Un enfant est identifié comme étant en surpoids lorsque son IMC est entre les centiles IOTF-25 et IOTF-30 de l'IMC, spécifiques de l'âge et du sexe (centiles de la croissance de l'IMC qui rejoignent à 18 ans ceux qui définissent les seuils chez l'adulte) et il est obèse

lorsque son IMC dépasse le centile IOTF-30. Parmi les garçons de l'échantillon décrit dans le chapitre 5, page 171, à l'âge de 5 ans, 42 étaient en surpoids ou obèse. Ces 42 garçons en surpoids ou obèse, ainsi qu'une sélection aléatoire de 100 garçons ayant un poids normal à 5 ans, ont été inclus dans les analyses.

6.3.2 Méthode de modélisation de la matrice de variances-covariances

Dans le paragraphe suivant, nous reprenons les notations telles qu'elles ont été introduites par Daniels et coll (Daniels and Zhao, 2003). Pour un modèle univarié (i.e. non conjoint) à effets aléatoires, pour chaque sujet i le vecteur de réponses Y_i de dimension n_i suit le modèle suivant :

$$Y_i = X_i\beta + Z_ib_i + \varepsilon_i$$

Avec $b_i \sim N(\mathbf{0}, \Sigma_i)$ et $\varepsilon_i \sim N(0, \tau^2 I)$ et où β est un vecteur de dimension p d'effets fixes X_i , b_i est un vecteur de dimension q d'effets aléatoires avec une matrice Z_i et la matrice de variances-covariances Σ_i est, ici, indexée par i .

Dans l'approche de Pourahmadi, les composantes du vecteur b_i des effets aléatoires doivent être ordonnées en fonction de leur signification. En effet, chaque composante est exprimée en fonction d'une ou plusieurs autres composantes de manière séquentielle et, de ce fait, le choix d'un ordre logique doit être fait. Daniels et coll. montrent par exemple que, pour un polynôme (orthogonal), un ordre naturel est la moyenne globale, puis la tendance linéaire, la tendance quadratique, etc. La matrice de variances-covariances ainsi définie est décomposée à l'aide de la méthode décrite par Pourahmadi et il en résulte un jeu de paramètres de dépendance et un jeu de paramètres de variances.

Ainsi, on définit \widetilde{b}_{ik} comme le prédicteur linéaire des moindres carrés de b_{ik} basé sur ses prédécesseurs b_{ik-1}, \dots, b_{i1} et $e_{ik} = b_{ik} - \widetilde{b}_{ik}$ est son erreur de prédiction de variance σ_{ik}^2 . On a donc :

$$b_{ik} = \sum_{j=1}^{k-1} \phi_{i,kj} b_{ij} + e_{ik}, k = 1, \dots, q$$

Les paramètres ϕ font référence aux paramètres de dépendance et les σ^2 aux paramètres de variance. La seule contrainte pour que la matrice de variances-covariances soit définie positive est que les paramètres de variance soient positifs.

Daniels et coll. montrent que pour un polynôme quadratique incluant 3 effets aléatoires, l'un représentant la déviation individuelle de la moyenne globale, un autre la déviation de la tendance linéaire moyenne, et le 3^e la déviation de la tendance quadratique moyenne, les égalités suivantes peuvent être écrites :

$$\begin{aligned} b_{i1} &= e_{i1} \\ b_{i2} &= \phi_{i,(2,1)}b_{i1} + e_{i2} \\ b_{i3} &= \phi_{i,(3,1)}b_{i1} + \phi_{i,(3,2)}b_{i2} + e_{i3} \end{aligned}$$

Et $\sigma_{ik}^2 = \text{var}(e_{ik})$.

La première équation correspond à la distribution marginale de l'effet aléatoire correspondant à la déviation du sujet i par rapport à la moyenne globale. La deuxième équation correspond à la distribution conditionnelle de l'effet aléatoire de la tendance linéaire sachant la moyenne globale. Enfin, la troisième équation décrit la distribution conditionnelle de l'effet aléatoire associé à la tendance quadratique sachant les effets aléatoires de la moyenne globale et de la tendance linéaire. Les paramètres $\phi_{i,kj}$ représentent donc ici la dépendance entre les différentes composantes du polynôme. Les paramètres de variance quantifient, quant à eux, la précision de l'estimation des déviations (i.e. σ_{i3}^2 décrira la précision avec laquelle la déviation de la tendance quadratique moyenne est prédite sachant les déviations de la tendance linéaire et de la moyenne globale).

Dans cette approche, les paramètres de dépendance et de variance peuvent être modélisés respectivement avec des fonctions linéaires et log.

$$\begin{aligned} \phi_{i,kj} &= W_{i,kj}\gamma \\ \log(\sigma_{ik}^2) &= H_{i,k}\lambda \end{aligned}$$

Avec $W_{i,kj}$ et $H_{i,k}$ des vecteurs de dimensions q_1 et q_2 respectivement, utilisés pour modéliser respectivement les paramètres de dépendance et de variance en fonction de covariables sujet-spécifiques et/ou pour modéliser la structure des $\phi_{i,kj}$ et σ_{ik}^2 d'un sujet.

6.3.3 Application de cette méthode à l'étude de la croissance

Cette méthode a été appliquée pour analyser les liens entre croissance staturale et pondérale et étudier l'effet modificateur du surpoids de l'enfant. Pour ces deux questions de recherche, les trajectoires de croissance staturo-pondérale ont été modélisées en incluant comme covariable le statut pondéral à 5 ans.

Dans le cadre de cette étude, la matrice de variances-covariances est de dimension 8×8, chaque croissance (celle du poids et celle de la taille) étant modélisée par un modèle non-linéaire impliquant 4 paramètres et, de ce fait, 4 effets aléatoires associés à chacun d'entre eux. Nous résumons brièvement la signification de chacun des paramètres : les paramètres A du poids et de la taille, représentent le poids et la taille dans les 4 premiers jours de vie et à la naissance respectivement ; les paramètres B, leur vitesse de croissance après 2 ans ; les paramètres C, leur poussée de croissance de la naissance à 2 ans (« gain pondéral de la naissance à 2 ans » et « gain statural de la naissance à 2 ans ») et les paramètres D leur courbure. Compte tenu de leur signification, il a été choisi de les ordonner de la manière suivante, en tenant compte de leur chronologie dans le temps : paramètres A, paramètres C, paramètres D et paramètres B. Il a ensuite été considéré que la croissance du poids dépend directement de celle de la taille (plutôt que l'inverse), de telle manière que l'ordre final étudié est paramètre A de la taille (A_T), paramètre A du poids (A_P), paramètre C de la taille (C_T), paramètre C du poids (C_P), paramètre D de la taille (D_T), paramètre D du poids (D_P), paramètre B de la taille (B_T), paramètre B du poids (B_P).

Ainsi, les équations suivantes sont obtenues

$$\begin{aligned}
 b_{iA_T} &= e_{iA_T} \\
 b_{iA_P} &= \phi_{i,(A_P,A_T)} b_{iA_T} + e_{iA_P} \\
 b_{iC_T} &= \phi_{i,(C_T,A_T)} b_{iA_T} + \phi_{i,(C_T,A_P)} b_{iA_P} + e_{iC_T} \\
 &\dots\dots\dots \\
 b_{iB_P} &= \phi_{i,(B_P,A_T)} b_{iA_T} \\
 &\quad + \phi_{i,(B_P,A_P)} b_{iA_P} + \phi_{i,(B_P,C_T)} b_{iC_T} \\
 &\quad + \phi_{i,(B_P,C_P)} b_{iC_P} + \phi_{i,(B_P,D_T)} b_{iD_T} + \phi_{i,(B_P,D_P)} b_{iD_P} + \phi_{i,(B_P,B_T)} b_{iB_T} + e_{iB_P} \\
 \text{Et } \sigma_{ik}^2 &= \text{var}(e_{ik}).
 \end{aligned}$$

Où la première équation représente la distribution marginale de l'effet aléatoire (b_{i1} qui a été pour faciliter la compréhension nommé b_{iA_T}) associé à A_T ; la deuxième la distribution de l'effet aléatoire de A_P ($b_{i2} = b_{iA_P}$) sachant A_T ; la troisième, la déviation du paramètre C_T ($b_{i3} = b_{iC_T}$) sachant A_T et A_P ; la quatrième, la distribution de C_P ($b_{i4} = b_{iC_P}$) sachant C_T , A_T et A_P ; la cinquième, la distribution de D_T ($b_{i5} = b_{iD_T}$) sachant C_P , C_T , A_T et A_P ; la sixième, la distribution de D_P ($b_{i6} = b_{iD_P}$) sachant D_T , C_P , C_T , A_T et A_P ; la septième, la distribution de B_T ($b_{i7} = b_{iB_T}$) sachant D_P , D_T , C_P , C_T , A_T et A_P et la dernière, la distribution de B_P ($b_{i8} = b_{iB_P}$) sachant B_T , D_P , D_T , C_P , C_T , A_T et A_P . Les paramètres de variance correspondent à la précision des estimations des effets aléatoires selon le même principe.

Comme énoncé précédemment, les paramètres de dépendance et de variance ont été modélisés respectivement avec des fonctions linéaires et log.

$$\begin{aligned}\phi_{i,kj} &= W_{i,kj}\gamma \\ \log(\sigma_{ik}^2) &= H_{i,k}\lambda\end{aligned}$$

Les structures de $W_{i,kj}$ et $H_{i,k}$ ont été choisies en fonction des deux questions énoncées en début de paragraphe.

1. Pour l'analyse globale simple de la structure d'association des paramètres de croissance pondérale et staturale, afin de mieux comprendre les relations entre paramètres et vérifier la cohérence de ces résultats avec la matrice de variances-covariances précédemment estimée sur la population. Pour cela, la matrice de variances-covariances est considérée homogène et ainsi :

$$W_{i,kj} = 1 \text{ et par conséquent } \gamma = \gamma_{kj}$$

Nous avons également considéré les paramètres de variance, communs à l'ensemble de la population étudiée :

$$H_{i,k} = 1 \text{ et de ce fait, } \lambda = \lambda_{kj}$$

2. Pour étudier les associations avec le statut pondéral de l'enfant à 5 ans, il a été considéré que les paramètres de dépendance pouvaient varier en fonction du statut pondéral.

$$W_{i,kj} = (1, \text{surpoids}_i) \text{ et de ce fait } \gamma = \frac{\gamma_{kjint}}{\gamma_{surpoids}}$$

Où surpoids_i prend la valeur 0 si l'enfant i a un poids normal et 1 sinon.

Le paramètre de dépendance reliant l'effet aléatoire k à l'effet aléatoire j sera donc :

$\phi_{i,kj} = \gamma_{kj_{int}} + \text{surpoids}_i * \gamma_{kj_{surpoids}}$ où $\gamma_{kj_{int}}$ représente la variation de l'effet aléatoire k avec la variation de l'effet aléatoire j pour les enfants de poids normal à 5 ans; $\gamma_{kj_{int}} + \gamma_{kj_{surpoids}}$ pour les enfants en surpoids ou obèse à 5 ans.

Les paramètres de variance ont été considérés comme ne dépendant pas du statut pondéral de l'enfant à 5 ans, on a donc :

$$H_{i,k} = 1 \text{ et de ce fait } \lambda = \lambda_{kj_{int}}$$

6.3.3.1 Ajustement des modèles

Les modèles sont des modèles bayésiens ajustés à l'aide du logiciel rjags sous R. Une phase de chauffe de 300000 itérations a été réalisée et les résultats ont été récupérés à partir de 3 chaînes lancées pour 10000 itérations et échantillonnées toutes les 20 itérations afin de réduire les autocorrélations.

Daniels et coll. recommandaient les distributions a priori suivantes respectivement pour les paramètres $\lambda, \beta, \gamma, \tau$:

$$\lambda \sim N(\lambda^*, \Omega_\lambda)$$

$$\beta \sim N(\beta^*, \Omega_\beta)$$

$$\gamma \sim N(\gamma^*, \Omega_\gamma)$$

$$\tau^2 \sim IG(a, b)$$

Ils recommandaient que ces lois soient non-informatives, par exemple en spécifiant λ^* et γ^* comme des vecteurs de zéros avec comme matrices respectives Ω_s une matrice identité multipliée par un scalaire de 10000 (correspondant à une variance de 10000 soit une précision de 0,0001).

En choisissant ces lois a priori, des problèmes d'estimation des paramètres ont été rencontrés. Il a donc été choisi de prendre les matrices Ω_s comme des matrices identité multipliées par 10 (précision de 0,1) qui reste assez peu informative. Pour les effets fixes β , les lois a priori décrites dans le chapitre 4 page 140 ont été considérées à la fois pour les intercepts des paramètres de croissance et les coefficients de régression associés au surpoids. En ce qui concerne les lois a priori sur les composantes de la variance, il était suggéré par Daniels et coll. de spécifier des lois en accord avec les

résultats attendus sur les données modélisées. Une loi gamma $\gamma(360, 60)$ et une loi gamma $\gamma(5, 7)$ ont été spécifiées sur la précision résiduelle ($1/\tau^2$) du poids et de la taille respectivement.

6.3.3.2 Résultats

Les résultats du Tableau 30 présentent, dans le modèle considérant une structure de variances-covariances des effets aléatoires commune à l'ensemble de la population étudiée (sans considérer le statut pondéral de l'enfant à 5 ans), les associations entre le statut pondéral de l'enfant à 5 ans et les paramètres de croissance (effets fixes). Ce tableau illustre que la principale différence entre les enfants en surpoids ou obèses à l'âge de 5 ans et les enfants de poids normal au même âge est qu'ils ont un paramètre B du poids plus élevé ($\beta=0,365$, IC95% [0,264; 0,451]), ce qui caractérise une vitesse de croissance plus rapide à partir de 2 ans.

Tableau 30 : Différences moyennes des paramètres de croissances staturale et pondérale de la naissance à 5 ans entre les enfants en surpoids ou obèses à 5 ans et les enfants de poids normal, étude EDEN, France, 2003-2012

Variable	A		B		C		D	
	Taille à la naissance		Vitesse de croissance		Poussée de croissance		Degré de courbure	
	Approximation du Poids minimal dans les 5 premiers jours		-Enfance-		-0-2 ans-		-Premiers mois-	
	β_A	IC 95%	β_B	IC 95%	β_C	IC 95%	β_D	IC 95%
Taille	0,006	-0,016; 0,030	0,049	-0,028; 0,111	-0,001	-0,075; 0,069	-0,040	-0,179; 0,070
Poids	0,010	-0,087; 0,107	0,365*	0,264; 0,451	-0,030	-0,144; 0,067	0,137	-0,022; 0,322

Abréviations: IC, Intervalle de Crédibilité

*Intervalles de crédibilité à 95% n'incluant pas la valeur 0

Tableau 31 : Moyenne a posteriori et intervalle de crédibilité des paramètres de dépendance entre les effets aléatoires des paramètres de croissances pondérale et staturale, matrice de variances-covariances homogène^a.

Paramètres de Dépendance	Moyenne a posteriori	IC 95% (Borne Inf)	IC 95% (Borne Sup)
γ			
$\gamma_{A_P A_T}^*$	3,50	3,01	4,05
$\gamma_{C_T A_T}$	1,20	-2,74	5,26
$\gamma_{C_T A_P}$	-0,45	-1,58	0,66
$\gamma_{C_P A_T}$	-1,44	-6,26	3,65
$\gamma_{C_P A_P}$	0,41	-1,04	1,82
$\gamma_{C_P C_T}^*$	1,19	0,60	1,81
$\gamma_{D_T A_T}$	0,30	-4,78	5,41
$\gamma_{D_T A_P}$	-0,83	-2,32	0,60
$\gamma_{D_T C_T}^*$	-1,67	-2,39	-1,06
$\gamma_{D_T C_P}$	0,19	-0,05	0,43
$\gamma_{D_P A_T}$	-1,09	-5,54	3,64
$\gamma_{D_P A_P}$	0,42	-0,99	1,71
$\gamma_{D_P C_T}$	-0,17	-1,36	0,92
$\gamma_{D_P C_P}^*$	-0,82	-1,13	-0,48
$\gamma_{D_P D_T}$	0,79	0,42	1,20
$\gamma_{B_T A_T}$	1,91	-2,56	5,89
$\gamma_{B_T A_P}$	-0,51	-1,91	0,88
$\gamma_{B_T C_T}$	0,41	-0,50	1,83
$\gamma_{B_T C_P}$	0,29	-0,22	1,09
$\gamma_{B_T D_T}$	-0,10	-1,01	0,45
$\gamma_{B_T D_P}$	0,49	-0,10	1,47
$\gamma_{B_P A_T}$	-1,65	-6,02	4,03
$\gamma_{B_P A_P}$	0,64	-1,17	2,03
$\gamma_{B_P C_T}^*$	1,14	0,08	2,32
$\gamma_{B_P C_P}$	-0,44	-1,16	0,51
$\gamma_{B_P D_T}$	-0,05	-1,08	0,70
$\gamma_{B_P D_P}$	0,15	-0,91	1,52
$\gamma_{B_P B_T}^*$	1,04	0,03	1,94

*Intervalles de crédibilité à 95% n'incluant pas la valeur 0

^aModèle considérant une structure de variances-covariances des effets aléatoires commune à l'ensemble de la population étudiée (sans considérer le statut pondéral de l'enfant à 5 ans)

Dans ce même modèle, nous interprétons les relations de dépendance les plus fortes (Intervalles de crédibilité à 95% n'incluant pas la valeur zéro) identifiées entre les effets aléatoires des paramètres. Les relations de dépendance entre les effets aléatoires (Tableau 31) montrent, par exemple, que le paramètre de régression de l'effet aléatoire du paramètre A du poids sur l'effet aléatoire du paramètre A de la taille ($\gamma_{A_P A_T}$) est

positif et la valeur 0 est nettement en dehors son intervalle de crédibilité à 95% ($\beta=3,5$, *Intervalle de Crédibilité à 95%* [3,0; 4,1]), ce qui est lié à la relation très positive entre le poids et la taille du nouveau-né.

La Figure 37 illustre pour 5 sujets fictifs, de quelle manière la déviation du paramètre A_T individuel (effet aléatoire b_{iA_T}) par rapport au paramètre A_T moyen dans la population (effet fixe associé à un effet aléatoire nul) se traduit en terme de déviation du paramètre A_P individuel par rapport au paramètre A_P moyen populationnel (fixé ici à une valeur de 1,10 en accord avec nos précédents résultats). Elle permet d'illustrer plus clairement les répercussions de ces co-variations sur les tailles à la naissance et poids à t_0 . Pour cela, la valeur de A_T moyen a été fixé à 3,9 sur la base de nos précédents résultats. Les valeurs des paramètres individuels ont été fixées à 3,7 ; 3,8 ; 3,9 ; 4,0 et 4,1 pour les sujets 1 à 5 et sont de ce fait associés à des effets aléatoires de -0,2 ; -0,1 ; 0,0 ; 0,1 et 0,2 respectivement. Le paramètre de dépendance entre le paramètre A_P et A_T ($\gamma_{A_P A_T}$) étant de 3,5, si l'on considère le sujet 1, l'effet aléatoire moyen associé à A_P pour ce sujet peut être calculé à partir de l'égalité suivante :

$$b_{1A_P} = \phi_{1,(A_P,A_T)} b_{1A_T}$$

$$\text{avec } \phi_{i,(A_P,A_T)} = \gamma_{A_P A_T}$$

d'où $b_{1A_P} = 3,5 * (-0,2) = -0,7004$

La valeur du paramètre A_P pour cet individu peut donc être calculée en sommant l'effet fixe de A_P et l'effet aléatoire associé à A_P de cet individu (soit $1,10-0,7004=0,40$). La même démarche a été appliquée à chaque individu. Afin de représenter les trajectoires, les valeurs des autres paramètres de croissance sont identiques pour l'ensemble des sujets. Elles ont été choisies de manière à produire des trajectoires de croissance plausibles.

Cette figure illustre le fait que lorsque l'effet aléatoire associé au paramètre A de la taille augmente (i.e. la taille à la naissance augmente), cette augmentation est associée à une augmentation de l'effet aléatoire associé au paramètre A du poids (i.e. du poids dans les premiers jours).

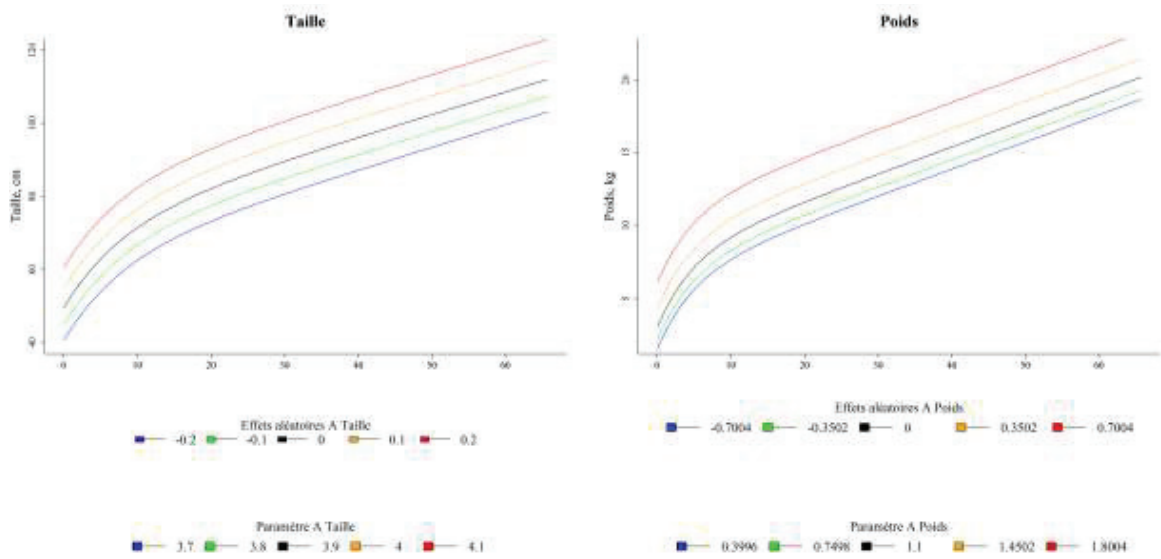


Figure 37 : Association entre le paramètre A de la taille et du poids (association marginale)

Un résultat similaire est observé pour le paramètre de régression de l'effet aléatoire du paramètre C du poids sur le paramètre C de la taille ($\gamma_{C_P C_T}$; $\beta=1,19$, *Intervalle de Crédibilité à 95%* [0,60; 1,81]). Ceci est illustré dans la Figure 38, construite de manière similaire à la Figure 37, par le fait que, lorsque la différence entre la taille à la naissance (commune pour l'ensemble des courbes) et l'ordonnée à l'origine de la l'asymptote à la courbe augmente (« gain statural de la naissance à 2 ans », maximum avec la courbe rouge), la différence entre le poids minimal dans les 1^{ers} jours et l'ordonnée à l'origine de l'asymptote de la courbe de poids augmente également (gain pondéral de la naissance à 2 ans).

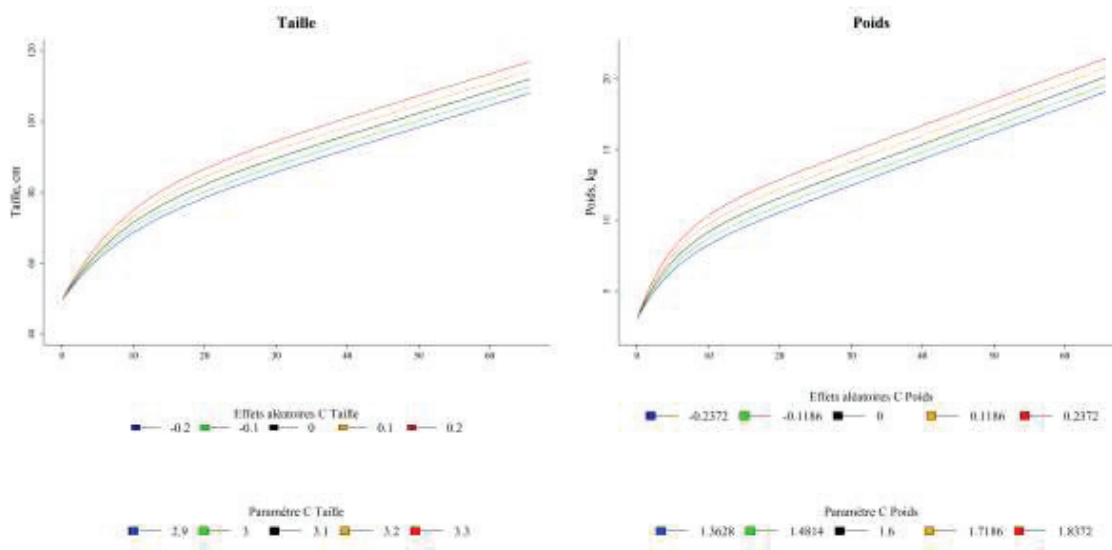


Figure 38 : Association entre le paramètre C de la taille et C du poids (ajustée sur les paramètres de naissance)

Ajustée sur les déviations des paramètres A du poids et de la taille et du paramètre C du poids, l'augmentation de la déviation du paramètre C de la taille était associée à une diminution de la déviation du paramètre D de la taille ($\gamma_{D_T C_T}$; $\beta = -1,67$, *Intervalle de Crédibilité à 95%* [-2,39 ; -1,06]). Ceci signifie que plus le gain statural de la naissance à 2 ans augmente, moins la courbure de la courbe de croissance staturale est élevée dans les premiers mois (indépendamment des paramètres de naissance et du gain pondéral de la naissance à 2 ans). On voit par exemple, dans la Figure 39, que la courbe rouge qui représente la courbe ayant le plus grand gain statural de la naissance à 2 ans, est également la courbe dont la courbure est la plus faible (la trajectoire semble presque linéaire de 0 à 8 mois environ). A l'inverse, la courbe bleue a la courbure la plus forte et le gain statural de la naissance à 2 ans le plus faible.

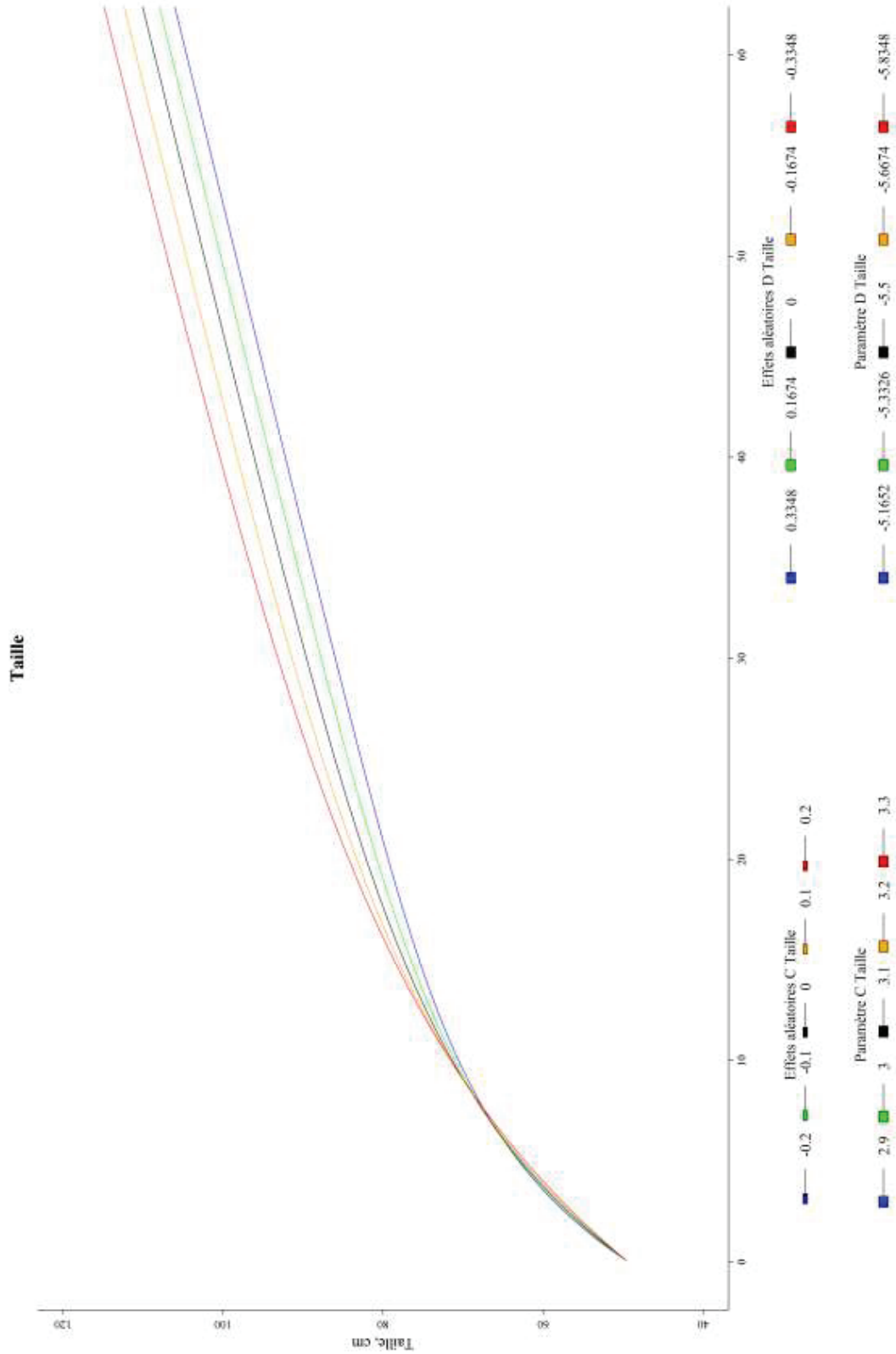


Figure 39 : Association entre le paramètre C de la taille et D de la taille (ajustée sur les paramètres de naissance, et le paramètre C du poids)

La même observation était faite pour l'association entre les paramètres C du poids et D du poids ($\gamma_{D_{PCP}}$; $\beta=-0,82$, *Intervalle de Crédibilité à 95%* [-1,13 ; -0,48]), ce qui est en faveur d'une diminution de la courbure de la croissance pondérale lorsque le gain pondéral de la naissance à 2 ans augmente à paramètres de naissance, gain statural de la naissance à 2 ans et courbure de la courbe de croissance de la taille égaux (Figure 40).

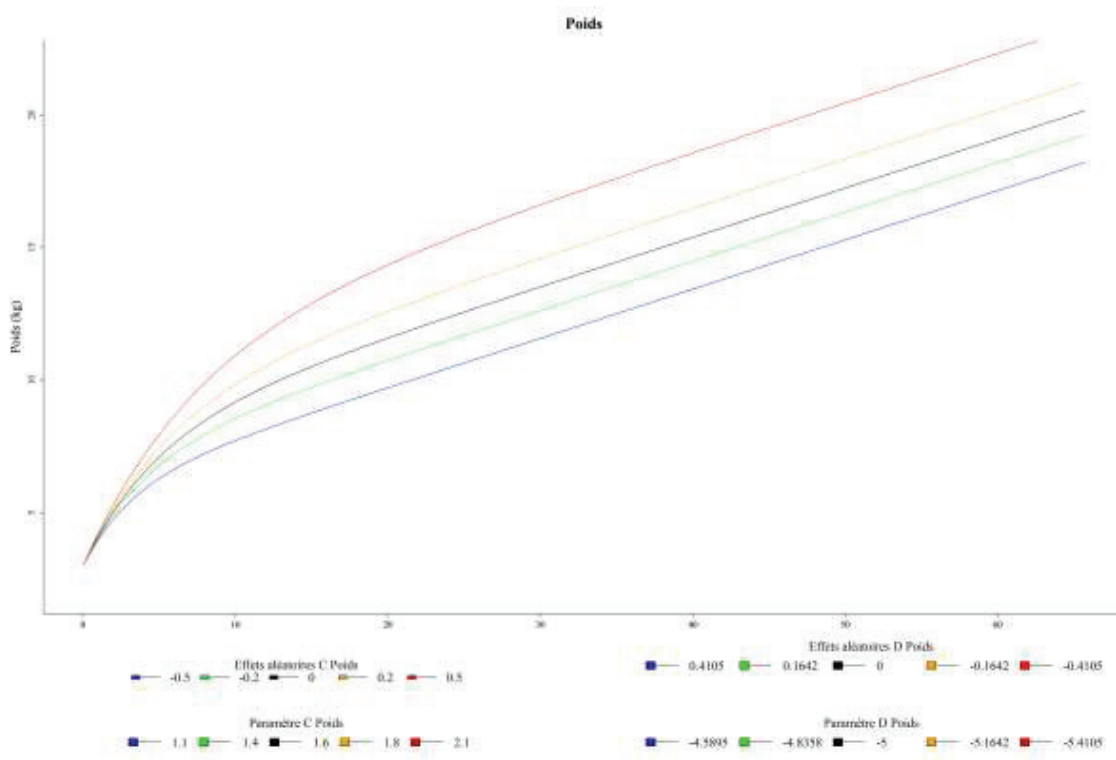


Figure 40 : Association entre le paramètre C du poids et D du poids (ajustée sur les paramètres de naissance, et les paramètres C et D de la taille)

Ajustée sur les déviations des paramètres A du poids et de la taille, C du poids et D de la taille et du poids et B de la taille, une association positive était observée entre le paramètre C de la taille et B du poids ($\gamma_{B_{TC_T}}$; $\beta=1,14$, *Intervalle de Crédibilité à 95%* [0,08; 2,32]). Ceci illustre que la vitesse de croissance pondérale après deux ans est associée positivement au gain statural de la naissance à 2 ans, indépendamment des paramètres de naissance et de la vitesse de croissance pondérale dans les premiers mois de vie et la vitesse de croissance staturale dans les premiers mois. Ceci est illustré par la Figure 41, construite de manière similaire à la Figure 37, qui montre que lorsque le gain

statural de la naissance à 2 ans augmente, la vitesse de croissance du poids après deux ans augmente.

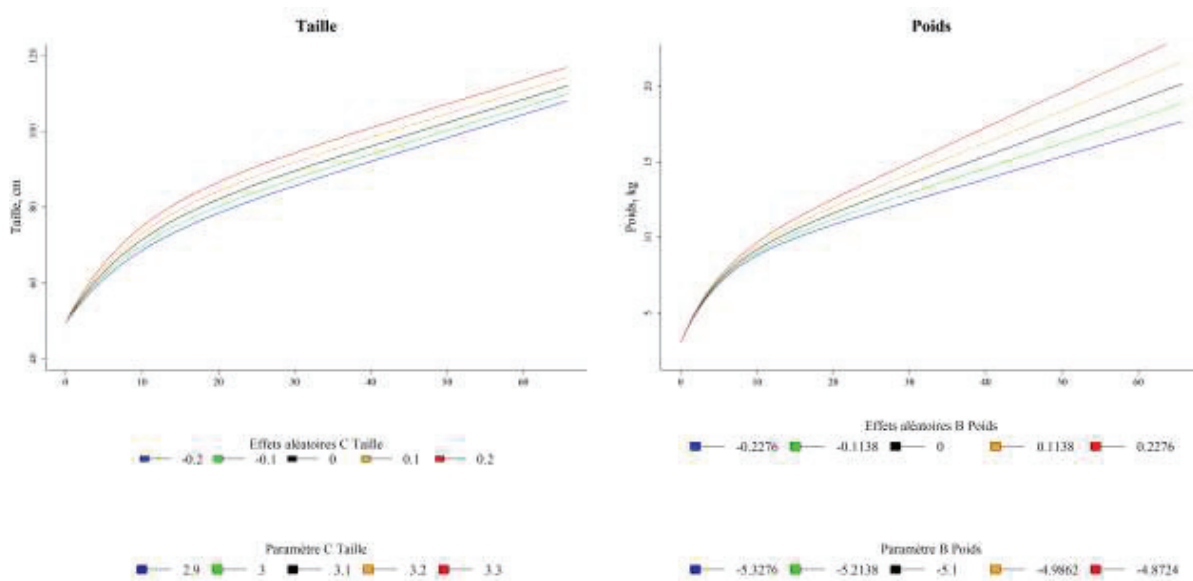


Figure 41 : Association entre le paramètre C de la taille et B du poids (ajustée sur l'ensemble des autres paramètres de croissance)

Enfin, une association positive était mise en évidence entre les paramètres B de la taille et B du poids. La vitesse de croissance pondérale après deux ans augmenterait avec la vitesse de croissance staturale après deux ans ($\gamma_{B_p B_T}$; $\beta=1,04$, *Intervalle de Crédibilité à 95%* [0,03; 1,94]) après prise en compte des autres paramètres caractérisant la croissance pondérale et staturale (paramètres de naissance, gain de poids pondéral et statural de la naissance et 2 ans et courbures des courbes de croissance). La Figure 42, construite de manière similaire à la Figure 37, illustre que la courbe de croissance staturale ayant la vitesse de croissance la plus forte après deux ans correspond à une courbe de croissance pondérale dont la vitesse de croissance après deux ans est également la plus forte (courbes rouges).

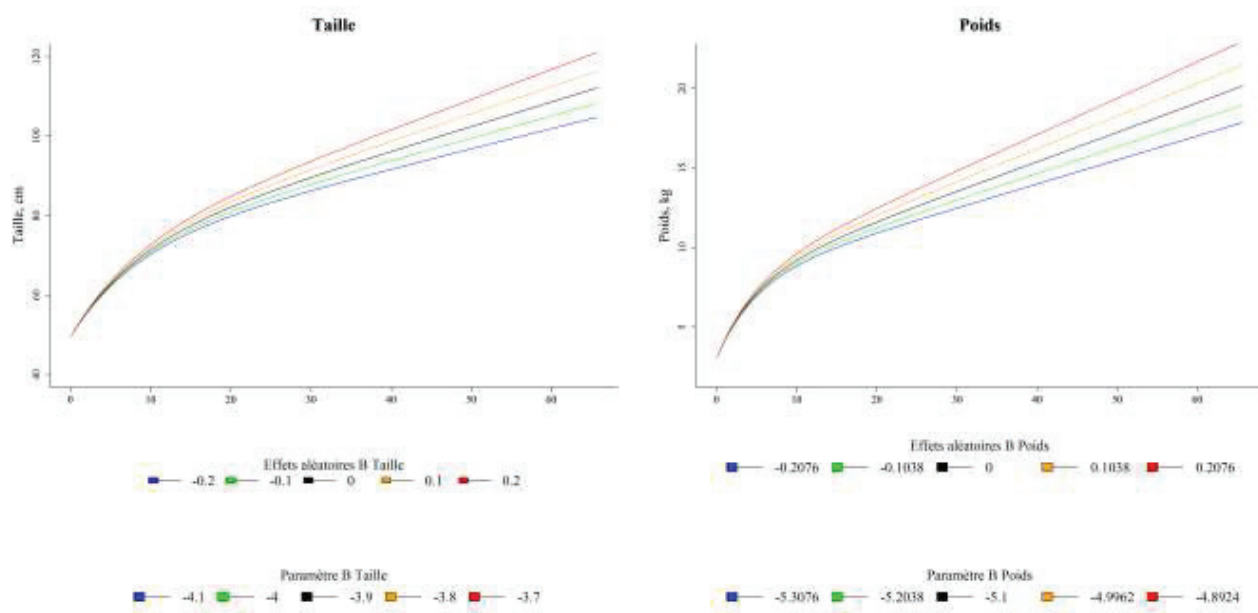


Figure 42 : Association entre le paramètre B de la taille et B du poids (ajusté sur l'ensemble des autres paramètres de croissance)

Le Tableau 32 décrit les paramètres de dépendance entre les effets aléatoires des paramètres de croissance pondérale et staturale obtenus en considérant une structure de variances-covariances des effets aléatoires différente selon le statut pondéral de l'enfant à 5 ans.

Tableau 32 : Moyenne a posteriori et intervalle de crédibilité des paramètres de dépendance entre les effets aléatoires des paramètres de croissances pondérale et staturale, matrice de variances-covariances dépendant du statut pondéral de l'enfant^a.

Paramètres de Dépendance	Moyenne a Posteriori	IC 95% (Borne Inf)	IC 95% (Borne Sup)
$\gamma_{APATint}$	3,10	2,55	3,67
$\gamma_{APATsurpoids}$	0,82	-0,16	1,86
$\gamma_{CTATint}$	1,30	-2,15	4,67
$\gamma_{CTATsurpoids}$	-0,12	-4,31	4,77
$\gamma_{CTAPint}$	-0,71	-1,82	0,37
$\gamma_{CTAPsurpoids}$	0,55	-0,79	1,77
$\gamma_{CPATint}$	-1,04	-5,47	3,51
$\gamma_{CPATsurpoids}$	-2,20	-8,28	4,62
$\gamma_{CPAPint}$	0,50	-0,93	2,00
$\gamma_{CPAPsurpoids}$	-0,02	-1,80	1,75
γ_{CPCint}	1,28	0,62	1,97
$\gamma_{CPCsurpoids}$ *	1,36	0,19	2,61
$\gamma_{DTATint}$	1,54	-3,05	5,90
$\gamma_{DTATsurpoids}$	-0,61	-6,14	4,56
$\gamma_{DTAPint}$	-1,56	-3,02	0,12
$\gamma_{DTAPsurpoids}$	0,74	-0,76	2,22
γ_{DTCint}	-2,19	-3,20	-1,29
$\gamma_{DTCsurpoids}$	1,03	-0,78	2,85
γ_{DTCint}	0,39	0,01	0,80
$\gamma_{DTCsurpoids}$	-0,41	-0,95	0,13
γ_{DPAint}	-0,84	-4,56	2,88
γ_{DPAint}	-0,75	-6,02	4,54
γ_{DPAint}	0,24	-1,24	1,87
γ_{DPAint}	0,81	-0,83	2,48
γ_{DPCint}	-0,62	-1,86	0,68
$\gamma_{DPCsurpoids}$	0,34	-2,19	2,70
γ_{DPCint}	-0,56	-1,05	-0,11
$\gamma_{DPCsurpoids}$	-0,16	-0,88	0,53
γ_{DPDint}	0,70	0,30	1,23
$\gamma_{DPDsurpoids}$	0,69	-0,17	1,63
$\gamma_{BTATint}$	2,04	-2,03	5,93
$\gamma_{BTATsurpoids}$	-0,93	-5,62	3,66
$\gamma_{BTAPint}$	-0,47	-2,02	1,22
$\gamma_{BTAPsurpoids}$	-0,22	-2,07	1,47
γ_{BTCint}	0,92	-0,58	2,53
$\gamma_{BTCsurpoids}$	0,19	-1,73	2,42
γ_{BTCint}	0,50	-0,22	1,62
$\gamma_{BTCsurpoids}$	-0,57	-1,64	0,28
γ_{BTDint}	-0,46	-1,75	0,27
$\gamma_{BTDsurpoids}$	0,09	-1,58	1,56
γ_{BDPint}	1,11	0,35	2,31
$\gamma_{BDPsurpoids}$	-0,77	-2,02	0,14
$\gamma_{BPATint}$	-1,78	-6,54	3,49
$\gamma_{BPATsurpoids}$	-0,21	-5,24	5,21
$\gamma_{BPAPint}$	1,22	-0,86	2,82
$\gamma_{BPAPsurpoids}$	-0,67	-2,69	1,29
γ_{BPCint}	2,48	-0,14	4,92
$\gamma_{BPCsurpoids}$	-0,95	-4,13	2,37
γ_{BPCint}	-0,29	-1,58	1,19
$\gamma_{BPCsurpoids}$	-0,17	-1,83	1,29
γ_{BPDint}	-0,22	-1,64	1,09
$\gamma_{BPDsurpoids}$	0,12	-1,77	2,07
γ_{BDPint}	0,95	-1,73	3,16
$\gamma_{BDPsurpoids}$	-0,79	-2,78	1,67
γ_{BBTint}	0,39	-1,32	2,20
$\gamma_{BBTsurpoids}$	0,68	-1,18	2,62

*Intervalles de crédibilité à 95% n'incluant pas la valeur 0

^aModèle considérant une structure de variances-covariances des effets aléatoires différente selon le statut pondéral de l'enfant à 5 ans.

Le paramètre de dépendance $\gamma_{C_P C_T_{surpoids}}$ s'élevait à 1,36 ($\beta=1,36$, *Intervalle de Crédibilité à 95%* [0,19; 2,61]). Ceci indique que le fait d'être en surpoids ou obèse à 5 ans était associé à une association significativement plus forte entre les déviations du paramètre C de la taille (C_T , gain statural de la naissance à 2 ans) et du paramètre C du poids (C_P , gain pondéral de la naissance à 2 ans) après ajustement sur les déviations des paramètres de naissance (paramètres A du poids et de la taille). A variation égale du gain statural de la naissance à 2 ans, le gain pondéral augmentait donc plus rapidement dans cette même période chez les enfants en surpoids ou obèse à 5 ans que chez les enfants de poids normal.

En ce qui concerne les autres associations, les intervalles de crédibilité à 95% contenaient tous la valeur zéro.

6.4 Discussion

Le gain pondéral dans les 2 premières années de vie augmentait plus fortement avec le gain statural sur cette même période après considération des paramètres à la naissance chez les enfants en surpoids ou obèse comparé aux enfants de poids normal à 5 ans. Dans, les 2 premières années de vie, à variation égale de la croissance staturale, les enfants en surpoids ou obèses à 5 ans avaient donc tendance à grossir plus rapidement que les enfants de poids normal. Ajusté sur les paramètres (poids, taille) de naissance, ainsi que sur les gains de poids et de taille de 0 à 2 ans, le fait d'être en surpoids ou obèse n'était pas associé à une modification notable des associations entre les déviations des vitesses de croissance staturale et pondérale après 2 ans (paramètres B).

Cette méthode de décomposition de la matrice de variances-covariances des effets aléatoires des paramètres de croissance pondérale et staturale offre un moyen intéressant d'étudier de quelle manière leur co-variation dépend de facteurs d'intérêt, en garantissant l'obtention de matrices définies positives. Lorsque les paramètres de variances et de covariances diffèrent selon des variables catégorielles, des matrices séparées pourraient aussi être spécifiées pour les sous-groupes d'intérêt sans avoir recours à ce type de modélisation. Dans ce cas, la modélisation peut permettre de bénéficier de la part commune entre les sous-groupes (s'il existe une tendance commune) et diminuer ainsi le nombre de paramètres à estimer. L'approche que nous avons adoptée reste particulièrement intéressante dans le cas où le facteur d'intérêt est

une variable continue pour lequel l'ajustement de matrices spécifiques par groupe ne sera pas possible.

Ce travail exploratoire nécessitera cependant d'être approfondi. En effet, les qualités d'ajustement des modèles n'ont pas été comparées. Une prochaine étape devrait donc consister à comparer les DIC et/ou WAIC du modèle simple (considérant une matrice de variances-covariances commune entre les sujets) et du modèle modélisant la matrice de variances-covariances en fonction du statut pondéral de l'enfant à 5 ans, afin de sélectionner la structure la plus adéquate. Il a été mis en évidence dans la littérature que la prise en compte des facteurs susceptibles de modifier la structure de variances-covariances pouvait améliorer à la fois la précision des estimateurs des effets fixes et la prédiction des trajectoires sujet-spécifiques (liée également à l'estimation des effets aléatoires) (Daniels and Zhao, 2003; Heagerty and Kurland, 2001). Les intervalles de crédibilité des différents modèles pour les effets aléatoires et effets fixes pourront également être comparés entre les deux modèles.

Les paramètres de variance (précisions des estimations des effets aléatoires) ont, dans les deux modèles ajustés, été considérés homogènes selon le statut pondéral. Il pourrait être intéressant d'étudier si la structure des paramètres de variance est, elle aussi, associée au statut pondéral de l'enfant à 5 ans.

D'une manière générale, cette méthode de décomposition offre une reparamétrisation de la matrice de covariances ayant une interprétation directe, mais elle impose un ordre entre les différents paramètres impliqués dans la modélisation. Cependant, compte tenu de la signification des paramètres de l'équation de Jentsch-Bayley, la séquence temporelle choisie entre les différents coefficients est intéressante. D'autres combinaisons pourraient également être considérées en fonction de la question d'intérêt.

Enfin, une des difficultés de cette décomposition réside dans le choix de la fonction de lien c'est-à-dire la fonction qui transforme les paramètres de la matrice de variances-covariances initiale (les variances et covariances) en un jeu de paramètres pour lesquels les paramètres de régression seront linéaires (Daniels and Zhao, 2003). Il a été choisi, dans ce travail, de modéliser les paramètres de dépendance et de variance en utilisant des fonctions linéaires et log respectivement en accord avec les travaux de Daniels et coll. (Daniels and Zhao, 2003). Cependant, d'autres fonctions ont été décrites

par Pourahmadi et coll. (Pourahmadi, 1999) et pourraient être comparées à celles choisies, notamment en comparant les DIC et/ou WAIC des différents modèles.

Discussion générale

L'objectif de cette thèse était d'explorer de nouvelles méthodes permettant de décrire la croissance staturo-pondérale et celle de l'IMC dans la petite enfance et de les appliquer à l'étude de ses déterminants en choisissant comme exemple le tabagisme maternel pendant la grossesse.

Dans la première partie, une méthode de modélisation par régression non-linéaire a été décrite et deux méthodes l'utilisant pour l'étude des associations entre le tabagisme maternel pendant la grossesse et la croissance du poids, de la taille et de l'IMC ont été comparées : une méthode en « 1 étape » et une méthode en « 2 étapes ». La méthode en « 1 étape » a permis de résumer la croissance pondérale et staturale des enfants avec un nombre restreint de paramètres ayant une interprétation concrète et d'en étudier les associations avec des déterminants d'intérêt et de prédire les trajectoires d'IMC pour des sous-groupes définis. La méthode en « 2 étapes » a permis de quantifier les associations avec le poids, la taille et l'IMC en s'affranchissant de problèmes de convergence éventuels lorsque le nombre de covariables est conséquent.

Dans la deuxième partie, l'intérêt à étudier conjointement la croissance du poids et de la taille par une approche bayésienne, et à prédire, dans le même temps, les trajectoires d'IMC avec une information sur l'incertitude associée à ces prédictions, a été exploré.

Dans la troisième partie, il a été montré que les prédictions de la croissance staturale étaient améliorées en prenant en compte l'information disponible sur la croissance pondérale.

Enfin, dans la dernière partie, une méthode permettant d'étudier l'effet de covariables sur la structure de variances-covariances des paramètres de croissance du poids et de la taille a été présentée.

Forces et limites de l'étude EDEN pour l'étude de la croissance et de ses déterminants précoces

L'étude de la dynamique de croissance staturo-pondérale (et d'IMC) des enfants requiert un suivi régulier et la collecte de données anthropométriques régulière au

cours du temps. Le nombre de couples mère-enfant ayant accepté de participer à l'étude était important (2002 couples inclus) et un nombre élevé de mesures de poids et de taille (environ 30000 pour le poids ainsi que pour la taille) a permis de disposer d'une information riche pour l'étude des trajectoires de croissance dans la population. Un des avantages de l'étude EDEN comparée aux autres cohortes mères-enfants était la disponibilité de mesures de poids et de taille dans les premiers jours de vie permettant de mieux modéliser la croissance pondérale. De plus, le suivi longitudinal mis en place pour les enfants de la cohorte et le nombre important de données collectées (e.g. données socio-démographiques, alimentaires, etc.) permet de caractériser finement l'environnement des enfants, que ce soit en période prénatale ou en postnatale.

Une des limites de l'étude EDEN est que les mères incluses ne sont pas représentatives de la population française (elles ont notamment un niveau d'études plus élevé que le niveau d'études moyen dans la population française). De ce fait, les résultats obtenus doivent être généralisés à l'ensemble de la population avec prudence, même si dans le cadre de cette thèse d'ordre méthodologique, cet aspect n'est pas vraiment limitant.

L'attrition de la population, inhérente aux études de cohorte, peut également être source de biais supplémentaires. Cependant, dans notre cas, le nombre d'enfants pour lesquels nous disposons de données de croissance après 4 ans restait important (72 % de l'échantillon total).

Importance de caractériser les trajectoires de croissance

La croissance de l'enfant est un processus dynamique indicateur de la santé à court terme et déterminant pour le développement et la santé à plus long terme. Plusieurs études ont montré que des variations de cinétique de croissance pourraient être associées à des désordres métaboliques à l'âge adulte. Une croissance pondérale rapide dans la petite enfance a notamment été associée à un risque accru d'obésité, de diabète de type 2 et de maladies cardiovasculaires (Eriksson, 2011). Lorsque les études analysent les associations entre un déterminant d'intérêt et les caractéristiques anthropométriques à un ou plusieurs âges donnés cela ne permet pas de caractériser finement la cinétique de croissance et les éventuelles phases de changements des profils de croissance.

Ces caractéristiques sont pourtant importantes pour apporter des éléments de compréhension et éventuellement suggérer des hypothèses sur les mécanismes pouvant expliquer des différences de trajectoires de croissance. Il a par exemple été mis en évidence dans la littérature que, *in utero*, à la fois une restriction alimentaire et une surnutrition, chacune induite par des déterminants différents, tels que, respectivement, le tabagisme maternel pendant la grossesse (Brisbois et al., 2012b; Ino, 2010; Oken et al., 2008; Weng et al., 2012) et un gain de poids gestationnel excessif (Mamun et al., 2014), pouvaient être associés à un risque de surpoids accru dans l'enfance et à l'âge adulte. Cependant, les cinétiques de croissances staturale et pondérale pouvant expliquer les divergences d'IMC semblaient différentes. Cela laisse supposer que les mécanismes biologiques sous-jacents amenant au surpoids seraient distincts.

Les études conduites sur le tabagisme maternel pendant la grossesse suggèrent, pour expliquer le risque de surpoids ultérieur, une part déterminante du phénomène de rattrapage pondéral lié à un faible poids à la naissance associé à un déficit de taille persistant (Chen et al., 2006; Durmus et al., 2011; Howe et al., 2012; Timmermans et al., 2013). Parmi les études conduites sur le gain de poids gestationnel, la plupart, suggèrent que les enfants de mères ayant un gain de poids gestationnel excessif ont un poids de naissance accru comparés aux enfants dont le gain de poids gestationnel est normal et conservent ce différentiel de poids au cours de la vie (phénomène de « tracking ») expliquant les différences d'IMC (Mamun et al., 2014). L'excès de poids serait alors, selon cette hypothèse, une conséquence de l'environnement intra-utérin, sans que ne soit observée d'augmentation de la vitesse de croissance pondérale. Ces exemples illustrent, en outre, l'importance de caractériser les trajectoires et cinétiques de croissance du poids et de la taille, en plus de celles de l'IMC. Dans l'exemple spécifique de l'IMC, l'effet différentiel de la croissance de la taille et du poids dans la croissance de l'IMC permet de mieux comprendre la « part imputable » aux caractéristiques de croissance du poids et de la taille dans l'évolution de l'IMC et, ainsi, de formuler des hypothèses sur les mécanismes impliqués.

Modèle de Jenss-Bayley : justification et intérêt

Les travaux réalisés dans le cadre de cette thèse se sont principalement basés sur la modélisation des croissances staturale et pondérale à l'aide de l'équation de Jenss-

Bayley. Ce choix a été guidé par la comparaison des performances des différents modèles paramétriques, adaptés à la modélisation de la croissance dans la petite enfance, à s'ajuster aux données de croissance des enfants de l'étude EDEN. En effet, comparé aux modèles de Reed et Count, le modèle de Jentsch-Bayley était plus adapté aux données des enfants inclus dans cette étude.

D'autres types de modèles, tels que les polynômes et les splines, auraient pu être considérés dans ces travaux, mais le modèle de Jentsch-Bayley présente des avantages importants pour l'étude des caractéristiques de la croissance et de ses déterminants. En effet, ce modèle (et plus généralement les modèles structuraux) suppose une forme de trajectoire de croissance commune à l'ensemble des enfants, guidée par des considérations physiologiques. De ce fait, il est plus robuste que les modèles par splines ou polynômes fractionnaires, en particulier aux âges extrêmes. Les paramètres ont une interprétation et fournissent une information directe sur les cinétiques de croissance. Le pendant de cette robustesse est que la forme générale de la trajectoire de croissance imposée par le modèle peut ne pas correspondre à l'ensemble des enfants et contraindre les enfants ayant des trajectoires très atypiques à avoir des trajectoires de croissance proches de celles des enfants ayant une croissance « normale ». Des méthodes comme celle de Nagin ou Muthén et coll. (Muthén and Shedden, 1999; Nagin et al., 1999) pourraient être, dans le cas où l'on soupçonne l'existence de sous-groupes ayant des trajectoires de croissance très différentes, plus appropriées. Nous avons vérifié que les qualités d'ajustement des modèles ne montraient pas d'écart trop importants entre les trajectoires de croissance observées et les trajectoires de croissance prédites par le modèle.

L'ensemble des travaux réalisés dans cette thèse a exploité plusieurs aspects du modèle de Jentsch-Bayley pour décrire la croissance pondérale et staturale des enfants. En particulier, en autorisant les paramètres du modèle à varier en fonction des valeurs des caractéristiques des individus, le modèle permet de fournir une information directement interprétable sur la modification des profils de croissance potentiellement induite par ces caractéristiques. Ainsi, nous avons mis en évidence que l'exposition au tabagisme tout au long de la grossesse était associée à un poids et une taille plus faible à la naissance, une croissance plus rapide du poids et de la taille dans la petite enfance et une croissance pondérale du poids plus rapide après 2 ans.

Avantages de la modélisation séparée du poids et de la taille pour en déduire l'IMC

La croissance de l'IMC suit une trajectoire non monotone. Sa modélisation est plus complexe que celle du poids ou de la taille. Dans la littérature, les associations entre des déterminants et l'IMC sont le plus souvent étudiées à des âges fixes sans que les trajectoires individuelles ne soient caractérisées (Wen et al., 2012). Les études s'intéressant aux profils de croissance longitudinaux d'IMC utilisent le plus souvent des polynômes fractionnaires (Chivers et al., 2010; Guo et al., 2000; Tilling et al., 2014; Whitaker et al., 1998; Williams et al., 1999) ou des splines (linéaires ou non)(Anderson et al., 2013; Ben-Shlomo et al., 2008; Howe et al., 2013, 2010; Tilling et al., 2014). Ces modèles, qui n'ont pas d'a priori sur la forme des trajectoires de croissance théorique, peuvent être délicats à ajuster : ils nécessitent notamment de trouver respectivement le degré de polynôme approprié ou le nombre et l'emplacement des nœuds. De plus, l'interprétation de leurs paramètres ou des associations avec des caractéristiques individuelles dans une modélisation en « 1 étape » n'est souvent pas possible. Enfin, la modélisation directe de l'IMC ne permet pas de conserver l'information sur les croissances séparées du poids et de la taille, même s'il est vrai que cette information pourrait être apportée par trois modélisations séparées (poids, taille et IMC). Notre méthode permet de n'en faire que deux et d'en déduire la troisième. Les profils de croissance du poids et de la taille évoluent de manière monotone au cours de l'enfance ; leur modélisation est probablement plus adaptée (surtout pour les individus disposant de peu de mesures) qu'une modélisation directe de l'IMC qui évolue de manière non monotone et dont les mesures sont plus imprécises.

De plus, comme les trajectoires d'IMC sont obtenues à partir de profils de croissance théoriques de la croissance du poids et de la croissance de la taille, le profil de croissance d'IMC est assez fidèle à la trajectoire de croissance attendue dans une population dont la croissance est « normale ». La « modélisation indirecte » permet également de conserver l'ensemble des mesures de poids et de taille dans les analyses sans limiter le jeu de données aux mesures de poids et de taille réalisées aux mêmes âges, tandis que pour modéliser l'IMC directement, les sujets doivent avoir été pesés et mesurés au même moment.

Deux méthodes utilisant cette méthode ont été adoptées dans cette thèse pour l'étude de l'association entre le tabagisme de la mère pendant la grossesse et la croissance du poids, de la taille et de l'IMC.

La première méthode « en 1 étape », a permis, dans les chapitres 3 et 4, de tester les associations entre ce déterminant et les paramètres des modèles de croissance du poids et de la taille donnant une interprétation directe des modifications des profils de croissance associées à ce déterminant et permettant d'étudier les associations entre ce déterminant et le profil de croissance global du poids ou de la taille. La méthode présentée dans le chapitre 3 permettait de dériver les trajectoires de croissance de l'IMC pour des sous-groupes d'intérêt, mais avait pour limite de ne pas permettre de tester les différences entre ces sous-groupes de manière directe. Dans le chapitre 4, la modélisation bayésienne conjointe du poids et de la taille a permis de pallier cette limite en permettant de donner un intervalle de crédibilité sur les prédictions d'IMC. Elle fournit ainsi un reflet de la confiance que l'on peut avoir sur les valeurs des différences (de poids, de taille) et d'IMC existant entre les sous-groupes étudiés.

Cette méthode a ici été utilisée pour caractériser les différences des mesures « cumulées » de poids, taille ou d'IMC. Les différences de poids, taille ou IMC à un temps données sont le reflet d'ensembles de phénomènes ayant marqué la croissance de l'enfant jusqu'à cet âge. Ainsi, une croissance très rapide du poids dans la petite enfance dans un groupe comparé à un autre avec un maintien du différentiel de poids plus tard dans l'enfance peut expliquer des différences ultérieures. Il serait donc intéressant de prédire également les vitesses de croissance instantanées dans les différents sous-groupes. Les vitesses de croissance fourniraient une mesure directe des différences de cinétiques de croissance dans les groupes étudiées et, ainsi, d'anticiper leurs répercussions sur les différences de poids, de taille et d'IMC.

Nous pourrions également envisager de complexifier ce modèle. Il a été choisi d'illustrer cette méthode pour l'étude des associations avec l'exposition au tabagisme maternel pendant la grossesse en ajustant sur un nombre de covariables restreint et d'autres covariables pourraient être incluses. En pratique, le temps de calcul est toutefois assez conséquent dans les modèles décrits précédemment malgré un jeu de covariables relativement petit ; la complexification des modèles pourrait être freinée par ces aspects pratiques. Enfin, d'autres expositions principales pourraient être également étudiées.

La deuxième méthode, « en 2 étapes », a permis de dériver les trajectoires de croissance individuelles du poids, de la taille et de l'IMC et, dans un second temps, de tester les différences de poids, taille, IMC prédits à des âges donnés associées aux covariables d'intérêt (chapitre 3). Cette méthode permet également de quantifier les différences entre caractéristiques anthropométriques d'intérêt à des âges donnés. Nous avons ici mis en évidence que les qualités d'ajustement aux données de croissance d'IMC obtenues à partir d'une modélisation séparée du poids et de la taille ne montrent pas de défaut majeur. L'étude des associations entre le tabagisme maternel et les caractéristiques de croissance prédites à partir de cette modélisation indirecte de l'IMC (cf. méthode en « 2 étapes », chapitre 3) montre également que cette méthode permet d'obtenir des estimations très proches de celles obtenues avec les méthodes de modélisation directe de l'IMC les plus utilisées dans la littérature, telles que les polynômes fractionnaires et les splines. Cependant, cette méthode consistant à prédire les caractéristiques de croissance dans un premier temps et à en étudier ensuite les associations avec des facteurs d'intérêt ignore l'incertitude sur les prédictions individuelles. Dans le chapitre 5, il a été mis en évidence que les moyennes des variances intra-sujets des prédictions étaient faibles par rapport aux variances inter-sujets. Les inférences ne devraient donc pas être biaisées par cette méthodologie, et nous retrouvons d'ailleurs des résultats cohérents entre la méthode en « 1 étape » et celle en « 2 étapes ».

Comparée à la méthode en 1 étape, cette méthode en 2 étapes offre l'avantage d'être moins limitée dans le nombre de covariables pouvant être inclus dans les modèles. Elle offre en outre la possibilité d'étudier les associations entre la croissance et d'autres déterminants ou événements de santé ultérieurs par des modèles de régression standards. Dans le cadre de ce travail de thèse, la modélisation des trajectoires individuelles de croissance du poids, de la taille a été réalisée et les trajectoires d'IMC prédites pour l'ensemble des enfants inclus dans l'étude EDEN lorsque le nombre de données disponibles le permettait. Les poids, taille, IMC et les vitesses instantanées correspondantes pour des âges définis ainsi que l'âge et l'IMC au pic d'IMC ont été calculés. Ces prédictions ont ainsi pu être utilisées (suivant la méthode en « 2 étapes »), au sein de l'équipe, pour répondre à d'autres questions de recherche sur l'étude des associations avec d'autres déterminants d'intérêt. Cela a été le cas pour l'étude des

associations entre l'exposition aux phtalates durant la grossesse et la croissance ; l'étude du rôle des vitesses de croissance précoces dans les associations entre le poids de naissance et la pression artérielle, l'étude des associations entre les comportements impliqués dans la balance énergétique et l'adiposité ultérieure.

Avantages de la modélisation conjointe du poids et de la taille

La modélisation bayésienne de la croissance pondérale et staturale mise en œuvre dans ce travail offre la possibilité de décrire simultanément l'évolution du poids et de la taille au cours du temps. Un des avantages de cette méthode est qu'elle ne nécessite pas que les variables modélisées soient mesurées aux mêmes âges, ni de disposer du même nombre de mesures pour chacune d'elle (Fieuws and Verbeke, 2004a). Ainsi, elle permet de conserver l'ensemble des mesures de poids et de taille dans les analyses (comme dans le cas de la « modélisation indirecte »). De plus, comme cela a pu être illustré dans le chapitre 5, elle permet en cas de données manquantes pour un des événements d'intérêt de bénéficier de l'information apportée par l'événement modélisé simultanément pour prédire cette donnée. Il a été illustré que les précisions des estimations des mesures de taille supprimées étaient, dans la majorité des cas, meilleures avec la modélisation conjointe plutôt qu'avec des modélisations univariées du poids et de la taille. Elle permet également, lorsqu'aucune donnée n'a pu être collectée pour un individu, de se servir de la trajectoire de croissance de l'événement disponible pour prédire sa trajectoire de croissance manquante. Ainsi, la trajectoire de croissance staturale de 25 individus sans mesure de taille a été modélisée à partir des données de poids de ces individus. Ceci est un atout majeur de cette méthode dans le cadre d'études de cohorte où le nombre de perdus de vue peut être conséquent et ce type de modèle a été décrit pour corriger sur le biais lié à l'exclusion de données informatives (liée à la présence de données manquantes) dans les études longitudinales (Follmann and Wu, 1995). Cependant, avec peu de mesures de taille pour un nombre de mesures de poids importants, la trajectoire de croissance de la taille sera guidée par les relations existant entre croissance pondérale et staturale dans la population étudiée. On peut ainsi penser qu'elle sera susceptible de contraindre l'évolution de la taille à avoir un profil « moyen », fonction du poids, et ainsi ne pas refléter l'existence d'une évolution conjointe atypique réelle.

Enfin, nous avons exploré, dans un modèle conjoint à effets aléatoires partagés, de quelle manière les associations entre les paramètres de croissance du poids et de la taille et leurs associations avec des caractéristiques sujet-spécifiques pouvaient être étudiées en utilisant une modification de la décomposition de Cholesky pour la matrice de variances-covariances des paramètres de croissance (Daniels and Zhao, 2003; Pourahmadi, 1999). Il serait intéressant de poursuivre ce travail : d'autres méthodes ont été décrites dans la littérature à la fois pour la modélisation conjointe et pour l'étude des déterminants associés à des différences d'associations entre les paramètres étudiés et pourraient également être explorées (Dantan et al., 2008; Proust-Lima et al., 2014). Les modèles conjoints à classes latentes font partie des modèles utilisés pour modéliser plusieurs événements simultanément ; ils reposent sur des hypothèses différentes de celles des modèles à effets aléatoires partagés que nous avons utilisés: ils supposent qu'une population hétérogène d'individus est composée de sous-groupes latents homogènes d'individus partageant les mêmes profils pour les événements modélisés (Lin et al., 2002; Proust-Lima et al., 2009). Dans les modèles à effets aléatoires partagés, la structure de corrélation des effets aléatoires considère une population homogène dans les profils des événements étudiés et dans leur co-évolution. Dans les modèles conjoints à classes latentes, les deux structures de dépendance sont séparées : chaque événement est modélisé grâce à un modèle à effets aléatoires permettant de prendre en compte la corrélation entre les différentes mesures d'un même individu pour chaque événement modélisé ; les classes latentes quant à elles, permettent de prendre en compte la dépendance entre les événements modélisés. Cette méthode a été décrite comme plus flexible comparée à une modélisation basée sur la structure de corrélation des effets aléatoires et dans sa possibilité à considérer des covariables expliquant les structures de corrélation entre les événements modélisés (Proust-Lima et al., 2014). Elle pourrait donc être appliquée à la modélisation conjointe de la croissance staturale et pondérale afin d'identifier, s'ils existent, des sous-groupes dont les structures de dépendance entre la croissance pondérale et staturale diffèrent et afin d'identifier les covariables expliquant l'appartenance à ces groupes. En considérant les modèles conjoint à effets aléatoires partagés plus généralement, il est également nécessaire d'avoir des hypothèses a priori sur les facteurs associés à des modifications des structures d'évolution, les variables devant être a priori incluses dans la modélisation.

Dans les modèles conjoints à classes latentes, la caractérisation (en terme de déterminants) des différents sous-groupes peut être faite a posteriori.

La modélisation conjointe étudiée ici impliquait la croissance pondérale et staturale mais pourrait être appliquée à d'autres événements de santé. Il pourrait également être envisagé de modéliser conjointement la croissance de la taille, du poids avec un troisième événement de santé collecté au cours du temps. En particulier, ce modèle pourrait être utilisé dans le but de prédire de façon dynamique la survenue d'un événement d'intérêt en fonction de plusieurs profils de croissance. Les modèles conjoints ont été majoritairement utilisés dans le but de prédire la survenue d'un événement de santé en fonction de l'évolution d'un biomarqueur associé à ce risque de survenue (Rizopoulos and Ghosh, 2011).

Autres perspectives de recherche

Outre les perspectives de recherche décrites au sein des différents paragraphes discutés ci-dessus, d'autres perspectives peuvent s'appliquer à la fois aux méthodes de modélisation et au travail réalisé sur les associations entre le tabagisme maternel pendant la grossesse et la croissance.

Dans l'ensemble de ce travail de thèse, la modélisation de la croissance pondérale et staturale s'est basée sur l'équation de Jenss-Bayley pour les raisons qui ont été évoquées précédemment. D'autres modèles, dont les avantages en terme d'interprétabilité des paramètres sont comparables, pourraient être utilisés, comme le modèle SITAR (Cole et al., 2010), utilisé pour la modélisation bayésienne conjointe de la croissance prénatale caractérisée par le volume embryonnaire, la longueur vertex-coccyx et la longueur estimée entre la tête et la longueur totale du fœtus (Willemsen et al., 2015).

Concernant le tabagisme maternel pendant la grossesse, la précocité des associations mises en évidence avec les caractéristiques de la croissance soutient un effet direct via l'environnement intra-utérin. Cependant, les mécanismes expliquant l'altération de la croissance fœtale et postnatale ne sont pas encore bien caractérisés et d'autres facteurs pourraient en partie expliquer ces associations. Il serait intéressant de compléter les analyses réalisées dans le cadre de cette thèse par des analyses utilisant des modèles à équations structurelles qui permettraient d'évaluer les associations

directes du tabagisme maternel sur la croissance des enfants et les associations indirectes (effet de ce facteur via un ou plusieurs facteurs intermédiaires). Une étude récente a utilisé cette méthodologie pour étudier les associations entre la croissance dans la petite enfance et des caractéristiques maternelles (dont le tabagisme maternel), de naissance et de la petite enfance. Les résultats de cette étude suggéraient la coexistence d'un effet direct et d'un effet indirect du tabagisme maternel pendant la grossesse via le développement fœtal et les pratiques d'allaitement (Zheng et al., 2015). Cette approche est modélisable par une approche bayésienne.

Perspectives en terme de Santé Publique

Les travaux méthodologiques de cette thèse ont été illustrés avec le tabagisme maternel pendant la grossesse qui, bien qu'il soit un facteur de risque établi d'issues de grossesse indésirables, reste un comportement fréquent dans les pays industrialisés. Nous avons confirmé l'association de ce facteur avec le risque de surpoids de l'enfant en décrivant finement la cinétique de croissance des enfants exposés par rapport aux non exposés. Si les mécanismes ne sont pas encore connus, les méfaits du tabagisme pendant la grossesse sur plusieurs aspects du développement du fœtus sont bien établis (Iñiguez et al., 2013; Jaddoe et al., 2007b).

Alors qu'il est pourtant un des principaux facteurs modifiables permettant d'améliorer la santé de l'enfant à court et à long terme, la prévalence du tabagisme pendant la grossesse dépassait 10% en 2001 dans les pays européens (europeristat). L'information des mères sur les dangers liés au tabagisme et l'accompagnement de celles-ci pour trouver des mesures adéquates, devraient donc être renforcés pour aider les femmes à arrêter de fumer et devrait principalement cibler les femmes issues des niveau-socioéconomiques les plus faibles, qui semblent les plus susceptibles de fumer pendant leur grossesse. La mise en place de campagnes de prévention pour aider et inciter les femmes à arrêter de fumer apparaît d'autant plus important que celles-ci semblent moins allaiter leurs enfants ou pour une durée plus courte (Horta et al., 2001). Cette fréquence d'allaitement plus faible a notamment été attribuée à une production plus faible de lait (Bahadori et al., 2013). Or, les bénéfices de l'allaitement maternel sur la santé de l'enfant sont nombreux. Si le choix de ne pas allaiter est lié à ces facteurs biologiques (ou à une volonté de ne pas faire perdurer l'exposition au tabac via le lait

maternel]), plutôt que lié à un réel choix de leur part, le sevrage tabagique chez ces femmes pourrait également aider à promouvoir l'allaitement maternel.

D'une manière générale, les méthodes de modélisation de la croissance utilisées dans cette thèse permettent de caractériser finement les trajectoires de croissance et d'identifier les facteurs pré- et postnataux associés à des modifications de ces trajectoires très tôt dans l'enfance. Elles pourront être utilisées pour l'étude d'autres déterminants susceptibles d'affecter la croissance postnatale. Elles constituent également un outil permettant d'anticiper les répercussions des modifications des cinétiques de croissance sur des événements de santé délétères. En cela, elles constituent un outil utile pour identifier des facteurs susceptibles d'avoir des conséquences en terme de Santé Publique. En effet, les périodes prénatale et postnatale précoce constituent des fenêtres de susceptibilité, car l'enfant est en plein développement, et ce sont donc des périodes à considérer pour la prévention de maladies chroniques.

REFERENCES

- Anderson, E.L., Fraser, A., Martin, R.M., Kramer, M.S., Oken, E., Patel, R., Tilling, K., PROBIT Study, 2013. Associations of postnatal growth with asthma and atopy: the PROBIT Study. *Pediatr Allergy Immunol* 24, 122–130. doi:10.1111/pai.12049
- Anderson, T., 1984. *An Introduction to Multivariate Statistical Analysis*. second edition. [WWW Document]. URL <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0471360910.html> (accessed 1.14.16).
- Andres, R.L., Day, M.C., 2000. Perinatal complications associated with maternal tobacco use. *Semin Neonatol* 5, 231–241. doi:10.1053/siny.2000.0025
- Ayatollahi, S.M.T., 2005. Growth velocity standards from longitudinally measured infants of age 0-2 years born in Shiraz, southern Iran. *Am. J. Hum. Biol.* 17, 302–309. doi:10.1002/ajhb.20126
- Bahadori, B., Riediger, N.D., Farrell, S.M., Uitz, E., Moghadasian, M.F., 2013. Hypothesis: smoking decreases breast feeding duration by suppressing prolactin secretion. *Med. Hypotheses* 81, 582–586. doi:10.1016/j.mehy.2013.07.007
- Baptiste-Roberts, K., Nicholson, W.K., Wang, N.-Y., Brancati, F.L., 2012. Gestational Diabetes and Subsequent Growth Patterns of Offspring: The National Collaborative Perinatal Project. *Matern Child Health J* 16, 125–132. doi:10.1007/s10995-011-0756-2
- Barker, D.J., 2007. The origins of the developmental origins theory. *Journal of internal medicine* 261, 412–7. doi:10.1111/j.1365-2796.2007.01809.x
- Beath, K.J., 2007. Infant growth modelling using a shape invariant model with random effects. *Stat Med* 26, 2547–2564. doi:10.1002/sim.2718
- Ben-Shlomo, Y., McCarthy, A., Hughes, R., Tilling, K., Davies, D., Smith, G.D., 2008. Immediate postnatal growth is associated with blood pressure in young adulthood: the Barry Caerphilly Growth Study. *Hypertension* 52, 638–644. doi:10.1161/HYPERTENSIONAHA.108.114256
- Berkey, C.S., 1982. Comparison of two longitudinal growth models for preschool children. *Biometrics* 38, 221–234.
- Berk, M., 2013. *sme: Smoothing-splines Mixed-effects Models*.
- Besse, P., Thomas-Agnan, C., 1989. Le lissage par fonctions splines en statistique, revue bibliographique. *Statistique et analyse des données* 14, 55–84.
- Betoko, A., Charles, M.-A., Hankard, R., Forhan, A., Bonet, M., Saurel-Cubizolles, M.-J., Heude, B., de Lauzon-Guillain, B., EDEN mother-child cohort study group, 2013. Infant feeding patterns over the first year of life: influence of family characteristics. *Eur J Clin Nutr* 67, 631–637. doi:10.1038/ejcn.2012.200
- Binder, H., Sauerbrei, W., Royston, P., 2013. Comparison between splines and fractional polynomials for multivariable model building with continuous covariates: a simulation study with continuous response. *Statist. Med.* 32, 2262–2277. doi:10.1002/sim.5639
- Blondel, B., Kermarrec, M., 2011. *Enquête Nationale Périnatale 2010*.
- Botton, J., Heude, B., Maccario, J., Borys, J.-M., Lommez, A., Ducimetière, P., Charles, M.A., 2010. Parental body size and early weight and height growth velocities in their offspring. *Early Human Development* 86, 445–450. doi:10.1016/j.earlhumdev.2010.06.001
- Botton, J., Heude, B., Maccario, J., Ducimetière, P., Charles, M.-A., FLVS Study Group, 2008. Postnatal weight and height growth velocities at different ages between birth and

- 5 y and body composition in adolescent boys and girls. *Am. J. Clin. Nutr.* 87, 1760–1768.
- Botton, J., Scherdel, P., Regnault, N., Heude, B., Charles, M.-A., 2014a. Postnatal weight and height growth modeling and prediction of body mass index as a function of time for the study of growth determinants. *Ann. Nutr. Metab.* 65, 156–166. doi:10.1159/000362203
- Botton, J., Scherdel, P., Regnault, N., Heude, B., Charles, M.-A., 2014b. Postnatal weight and height growth modeling and prediction of body mass index as a function of time for the study of growth determinants. *Ann. Nutr. Metab.* 65, 156–166. doi:10.1159/000362203
- Brisbois, T.D., Farmer, A.P., McCargar, L.J., 2012a. Early markers of adult obesity: a review. *Obesity Reviews* 13, 347–367. doi:10.1111/j.1467-789X.2011.00965.x
- Brisbois, T.D., Farmer, A.P., McCargar, L.J., 2012b. Early markers of adult obesity: a review. *Obesity Reviews* 13, 347–367. doi:10.1111/j.1467-789X.2011.00965.x
- Castles, A., Adams, E.K., Melvin, C.L., Kelsch, C., Boulton, M.L., 1999. Effects of smoking during pregnancy. Five meta-analyses. *Am J Prev Med* 16, 208–215.
- Cavelaars, A.E., Kunst, A.E., Geurts, J.J., Crialesi, R., Grötvedt, L., Helmer, U., Lahelma, E., Lundberg, O., Mielck, A., Rasmussen, N.K., Regidor, E., Spuhler, T., Mackenbach, J.P., 2000. Persistent variations in average height between countries and between socio-economic groups: an overview of 10 European countries. *Ann. Hum. Biol.* 27, 407–421.
- Chen, A., Pennell, M.L., Klebanoff, M.A., Rogan, W.J., Longnecker, M.P., 2006. Maternal smoking during pregnancy in relation to child overweight: follow-up to age 8 years. *Int J Epidemiol* 35, 121–130. doi:10.1093/ije/dyi218
- Chen, H., Morris, M.J., 2007. Maternal smoking—A contributor to the obesity epidemic? *Obesity Research & Clinical Practice* 1, 155–163. doi:10.1016/j.orcp.2007.07.004
- Chirwa, E.D., Griffiths, P.L., Maleta, K., Norris, S.A., Cameron, N., 2014. Multi-level modelling of longitudinal child growth data from the Birth-to-Twenty Cohort: a comparison of growth models. *Ann Hum Biol* 41, 168–179. doi:10.3109/03014460.2013.839742
- Chivers, P., Hands, B., Parker, H., Beilin, L., Kendall, G., Bulsara, M., 2009. Longitudinal modelling of body mass index from birth to 14 years. *Obes Facts* 2, 302–310. doi:10.1159/000235561
- Chivers, P., Hands, B., Parker, H., Bulsara, M., Beilin, L.J., Kendall, G.E., Oddy, W.H., 2010. Body mass index, adiposity rebound and early feeding in a longitudinal cohort (Raine Study). *Int J Obes (Lond)* 34, 1169–1176. doi:10.1038/ijo.2010.61
- Chrestani, M.A., Santos, I.S., Horta, B.L., Dumith, S.C., de Oliveira Dode, M.A.S., 2013. Associated factors for accelerated growth in childhood: a systematic review. *Matern Child Health J* 17, 512–519. doi:10.1007/s10995-012-1025-8
- Cnattingius, S., 2004. The epidemiology of smoking during pregnancy: smoking prevalence, maternal characteristics, and pregnancy outcomes. *Nicotine Tob. Res.* 6 Suppl 2, S125–140. doi:10.1080/14622200410001669187
- Cole, T.J., Bellizzi, M.C., Flegal, K.M., Dietz, W.H., 2000. Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ* 320, 1240–1243.
- Cole, T.J., Donaldson, M.D.C., Ben-Shlomo, Y., 2010. SITAR—a useful instrument for growth curve analysis. *Int J Epidemiol* 39, 1558–1566. doi:10.1093/ije/dyq115
- Comets, E., Lavenu, A., Lavielle, M., 2012. SAEMIX, an R version of the SAEM algorithm for parameter estimation in nonlinear mixed effect models.

- Costet, N., Pelé, F., Comets, E., Rouget, F., Monfort, C., Bodeau-Livinec, F., Linganiza, E.M., Bataille, H., Kadhel, P., Multigner, L., Cordier, S., 2015. Perinatal exposure to chlordecone and infant growth. *Environ. Res.* 142, 123–134. doi:10.1016/j.envres.2015.06.023
- Cowles, M.K., Carlin, B.P., 1996. Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *Journal of the American Statistical Association* 91, 883. doi:10.2307/2291683
- Daniels, M.J., Zhao, Y.D., 2003. Modelling the random effects covariance matrix in longitudinal data. *Stat Med* 22, 1631–1647. doi:10.1002/sim.1470
- Dantan, E., Proust-Lima, C., Letenneur, L., Jacqmin-Gadda, H., 2008. Pattern mixture models and latent class models for the analysis of multivariate longitudinal data with informative dropouts. *Int J Biostat* 4, Article 14.
- Davidian, M., Giltinan, D.M., 1995. *Nonlinear Models for Repeated Measurement Data*. CRC Press.
- De la Cruz-Mesía, R., Marshall, G., 2006. Non-linear random effects models with continuous time autoregressive errors: a Bayesian approach. *Stat Med* 25, 1471–1484. doi:10.1002/sim.2290
- Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 39, 1–38.
- Dennis, C.-L., 2002. Breastfeeding Initiation and Duration: A 1990-2000 Literature Review. *Journal of Obstetric, Gynecologic, & Neonatal Nursing* 31, 12–32. doi:10.1111/j.1552-6909.2002.tb00019.x
- Dennison, B.A., Edmunds, L.S., Stratton, H.H., Pruzek, R.M., 2006. Rapid infant weight gain predicts childhood overweight. *Obesity (Silver Spring)* 14, 491–499. doi:10.1038/oby.2006.64
- Dewey, K.G., 1998. Growth characteristics of breast-fed compared to formula-fed infants. *Biol. Neonate* 74, 94–105.
- Dewey, K.G., Peerson, J.M., Brown, K.H., Krebs, N.F., Michaelsen, K.F., Persson, L.A., Salmenpera, L., Whitehead, R.G., Yeung, D.L., 1995. Growth of breast-fed infants deviates from current reference data: a pooled analysis of US, Canadian, and European data sets. *World Health Organization Working Group on Infant Growth. Pediatrics* 96, 495–503.
- DiFranza, J.R., Aligne, C.A., Weitzman, M., 2004. Prenatal and postnatal environmental tobacco smoke exposure and children's health. *Pediatrics* 113, 1007–1015.
- Donnet, S., Foulley, J.-L., Samson, A., 2010. Bayesian analysis of growth curves using mixed models defined by stochastic differential equations. *Biometrics* 66, 733–741. doi:10.1111/j.1541-0420.2009.01342.x
- Durmus, B., Kruithof, C.J., Gillman, M.H., Willemssen, S.P., Hofman, A., Raat, H., Eilers, P.H.C., Steegers, E.A.P., Jaddoe, V.W.V., 2011. Parental smoking during pregnancy, early growth, and risk of obesity in preschool children: the Generation R Study. *Am. J. Clin. Nutr.* 94, 164–171. doi:10.3945/ajcn.110.009225
- Duval, M., Robert-Granié, C., Foulley, J.-L., 2009. Estimation of heterogeneous variances in nonlinear mixed models via the SAEM-MCMC algorithm with applications to growth curves in poultry. *Journal de la Société Française de Statistique* 150, 65–83.
- Eriksson, J.G., 2011. Early growth and coronary heart disease and type 2 diabetes: findings from the Helsinki Birth Cohort Study (HBCS). *Am. J. Clin. Nutr.* 94, 1799S–1802S. doi:10.3945/ajcn.110.000638

- Fieuws, S., Verbeke, G., 2004a. Joint modelling of multivariate longitudinal profiles: pitfalls of the random-effects approach. *Statist. Med.* 23, 3093–3104. doi:10.1002/sim.1885
- Fieuws, S., Verbeke, G., 2004b. Joint modelling of multivariate longitudinal profiles: pitfalls of the random-effects approach. *Statist. Med.* 23, 3093–3104. doi:10.1002/sim.1885
- Florath, I., Kohler, M., Weck, M.N., Brandt, S., Rothenbacher, D., Schöttker, B., Moß, A., Gottmann, P., Wabitsch, M., Brenner, H., 2014. Association of pre- and post-natal parental smoking with offspring body mass index: an 8-year follow-up of a birth cohort. *Pediatr Obes* 9, 121–134. doi:10.1111/j.2047-6310.2012.00146.x
- Follmann, D., Wu, M., 1995. An Approximate Generalized Linear Model with Random Effects for Informative Missing Data. *Biometrics* 51, 151–168. doi:10.2307/2533322
- Foulley, J.L., San Cristobal, M., Gianola, D., Im, S., 1992. Marginal likelihood and Bayesian approaches to the analysis of heterogeneous residual variances in mixed linear Gaussian models. *Computational Statistics & Data Analysis* 13, 291–305. doi:10.1016/0167-9473(92)90137-5
- Garner, P., Panpanich, R., Logan, S., 2000. Is routine growth monitoring effective? A systematic review of trials. *Arch. Dis. Child.* 82, 197–201.
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. *Bayesian Data Analysis, Third Edition*. CRC Press.
- Gelman, A., Hill, J., 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Gelman, A., Rubin, D.B., 1992. Inference from Iterative Simulation Using Multiple Sequences. *Statist.Sci.* 7, 457–472. doi:10.1214/ss/1177011136
- Gilles Celeux, F.F., 2006. Deviance Information Criteria for Missing Data Models. *Bayesian Analysis* 1. doi:10.1214/06-BA122
- Goldstein, H., 1986. Efficient statistical modelling of longitudinal data. *Ann. Hum. Biol.* 13, 129–141.
- Guo, S.S., Huang, C., Maynard, L.M., Demerath, E., Towne, B., Chumlea, W.C., Siervogel, R.M., 2000. Body mass index during childhood, adolescence and young adulthood in relation to adult overweight and adiposity: the Fels Longitudinal Study. *Int. J. Obes. Relat. Metab. Disord.* 24, 1628–1635.
- Harrell, F.E., 2001. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer Science & Business Media.
- Hauspie, R.C., Cameron, N., Molinari, L., 2004. *Methods in Human Growth Research*. Cambridge University Press.
- Heagerty, P.J., Kurland, B.F., 2001. Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika* 88, 973–985. doi:10.1093/biomet/88.4.973
- Hediger, M.L., Overpeck, M.D., Ruan, W.J., Troendle, J.F., 2000. Early infant feeding and growth status of US-born infants and children aged 4-71 mo: analyses from the third National Health and Nutrition Examination Survey, 1988-1994. *Am. J. Clin. Nutr.* 72, 159–167.
- Henderson, R., Diggle, P., Dobson, A., 2000. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 1, 465–480. doi:10.1093/biostatistics/1.4.465
- Heude, B., Forhan, A., Slama, R., Douhaud, L., Bedel, S., Saurel-Cubizolles, M.-J., Hankard, R., Thiebaugeorges, O., Agostini, M.D., Annesi-Maesano, I., Kaminski, M., Charles, M.-A., Group, on behalf of the E. mother-child cohort study, Annesi-Maesano, I.,

- Bernard, J.Y., Botton, J., Charles, M.-A., Dargent-Molina, P., Lauzon-Guillain, B. de, Ducimetière, P., Agostini, M. de, Foliguet, B., Forhan, A., Fritel, X., Germa, A., Goua, V., Hankard, R., Heude, B., Kaminski, M., Larroque, B., Lelong, N., Lepeule, J., Magnin, G., Marchand, L., Nabet, C., Pierre, F., Slama, R., Saurel-Cubizolles, M.J., Schweitzer, M., Thiebaugeorges, O., 2015. Cohort Profile: The EDEN mother-child cohort on the prenatal and early postnatal determinants of child health and development. *Int. J. Epidemiol.* dyv151. doi:10.1093/ije/dyv151
- Heude, B., Thiébaugeorges, O., Goua, V., Forhan, A., Kaminski, M., Foliguet, B., Schweitzer, M., Magnin, G., Charles, M.-A., 2012. Pre-pregnancy body mass index and weight gain during pregnancy: relations with gestational diabetes and hypertension, and birth outcomes. *Matern Child Health J* 16, 355–363. doi:10.1007/s10995-011-0741-9
- Holzhauser, S., Hokken Koelega, A.C.S., Ridder, M. de, Hofman, A., Moll, H.A., Steegers, E.A.P., Witteman, J.C.M., Jaddoe, V.W.V., 2009. Effect of birth weight and postnatal weight gain on body composition in early infancy: The Generation R Study. *Early Hum. Dev.* 85, 285–290. doi:10.1016/j.earlhumdev.2008.11.002
- Horta, B.L., Kramer, M.S., Platt, R.W., 2001. Maternal smoking and the risk of early weaning: a meta-analysis. *Am J Public Health* 91, 304–307.
- Hosseini, S.-M., Maracy, M.-R., Sarrafzade, S., Kelishadi, R., 2014. Child Weight Growth Trajectory and its Determinants in a Sample of Iranian Children from Birth until 2 Years of Age. *Int J Prev Med* 5, 348–355.
- Howe, L.D., Matijasevich, A., Tilling, K., Brion, M.-J., Leary, S.D., Smith, G.D., Lawlor, D.A., 2012. Maternal smoking during pregnancy and offspring trajectories of height and adiposity: comparing maternal and paternal associations. *Int J Epidemiol* 41, 722–732. doi:10.1093/ije/dys025
- Howe, L.D., Tilling, K., Benfield, L., Logue, J., Sattar, N., Ness, A.R., Smith, G.D., Lawlor, D.A., 2010. Changes in ponderal index and body mass index across childhood and their associations with fat mass and cardiovascular risk factors at age 15. *PLoS ONE* 5, e15186. doi:10.1371/journal.pone.0015186
- Howe, L.D., Tilling, K., Galobardes, B., Smith, G.D., Ness, A.R., Lawlor, D.A., 2011. Socioeconomic disparities in trajectories of adiposity across childhood. *Int J Pediatr Obes* 6, e144–153. doi:10.3109/17477166.2010.500387
- Howe, L.D., Tilling, K., Matijasevich, A., Petherick, E.S., Santos, A.C., Fairley, L., Wright, J., Santos, I.S., Barros, A.J.D., Martin, R.M., Kramer, M.S., Bogdanovich, N., Matush, L., Barros, H., Lawlor, D.A., 2013. Linear spline multilevel models for summarising childhood growth trajectories: A guide to their application using examples from five birth cohorts. *Stat Methods Med Res.* doi:10.1177/0962280213503925
- Huang, Y., Dagne, G., Wu, L., 2011. Bayesian inference on joint models of HIV dynamics for time-to-event and longitudinal data with skewness and covariate measurement errors. *Stat Med* 30, 2930–2946. doi:10.1002/sim.4321
- Iliadou, A.N., Koupil, I., Villamor, E., Altman, D., Hultman, C., Långström, N., Cnattingius, S., 2010. Familial factors confound the association between maternal smoking during pregnancy and young adult offspring overweight. *Int J Epidemiol* 39, 1193–1202. doi:10.1093/ije/dyq064
- Iñiguez, C., Ballester, F., Costa, O., Murcia, M., Souto, A., Santa-Marina, L., Aurekoetxea, J.J., Espada, M., Vrijheid, M., Alvarez-Avellón, S.M., Alvarez-Pedrerol, M., Rebagliato, M., INMA Study Investigators, 2013. Maternal smoking during pregnancy and fetal biometry: the INMA Mother and Child Cohort Study. *Am. J. Epidemiol.* 178, 1067–1075. doi:10.1093/aje/kwt085

- Ino, T., 2010. Maternal smoking during pregnancy and offspring obesity: meta-analysis. *Pediatr Int* 52, 94–99. doi:10.1111/j.1442-200X.2009.02883.x
- Jaddoe, V.W.V., Verburg, B.O., de Ridder, M. a. J., Hofman, A., Mackenbach, J.P., Moll, H.A., Steegers, E.A.P., Witteman, J.C.M., 2007a. Maternal smoking and fetal growth characteristics in different periods of pregnancy: the generation R study. *Am. J. Epidemiol.* 165, 1207–1215. doi:10.1093/aje/kwm014
- Jaddoe, V.W.V., Verburg, B.O., de Ridder, M. a. J., Hofman, A., Mackenbach, J.P., Moll, H.A., Steegers, E.A.P., Witteman, J.C.M., 2007b. Maternal smoking and fetal growth characteristics in different periods of pregnancy: the generation R study. *Am. J. Epidemiol.* 165, 1207–1215. doi:10.1093/aje/kwm014
- Jenss, R., Bayley, N., 1937. A mathematical method for studying the growth of a child. *Hum Biol* 9, 556–63.
- John Barnard, McCulloch, R., Meng, X.-L., 2000. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. Vol.10 No.4 [WWW Document]. URL <http://www3.stat.sinica.edu.tw/statistica/j10n4/j10n416/j10n416.htm> (accessed 9.24.15).
- Johnson, L., Llewellyn, C.H., van Jaarsveld, C.H.M., Cole, T.J., Wardle, J., 2011. Genetic and environmental influences on infant growth: prospective analysis of the Gemini twin birth cohort. *PLoS ONE* 6, e19918. doi:10.1371/journal.pone.0019918
- Jones-Smith, J.C., Neufeld, L.M., Laraia, B., Ramakrishnan, U., Garcia-Guerra, A., Fernald, L.C.H., 2013. Early life growth trajectories and future risk for overweight. *Nutr Diabetes* 3, e60. doi:10.1038/nutd.2012.32
- Karaolis-Danckert, N., Buyken, A.E., Bolzenius, K., Perim de Faria, C., Lentze, M.J., Kroke, A., 2006. Rapid growth among term children whose birth weight was appropriate for gestational age has a longer lasting effect on body fat percentage than on body mass index. *Am. J. Clin. Nutr.* 84, 1449–1455.
- Karlberg, J., 1989. A biologically-oriented mathematical model (ICP) for human growth. *Acta Paediatr Scand Suppl* 350, 70–94.
- Koletzko, B., Chourdakis, M., Grote, V., Hellmuth, C., Prell, C., Rzehak, P., Uhl, O., Weber, M., 2014. Regulation of early human growth: impact on long-term health. *Ann. Nutr. Metab.* 65, 101–109. doi:10.1159/000365873
- Kramer, M.S., Guo, T., Platt, R.W., Vanilovich, I., Sevkovskaya, Z., Dzikovich, I., Michaelsen, K.F., Dewey, K., Promotion of Breastfeeding Intervention Trials Study Group, 2004. Feeding effects on growth during infancy. *J. Pediatr.* 145, 600–605. doi:10.1016/j.jpeds.2004.06.069
- Kwok, M.K., Schooling, C.M., Lam, T.H., Leung, G.M., 2010. Paternal smoking and childhood overweight: evidence from the Hong Kong “Children of 1997.” *Pediatrics* 126, e46–56. doi:10.1542/peds.2009-2642
- Leary, S.D., Smith, G.D., Rogers, I.S., Reilly, J.J., Wells, J.C.K., Ness, A.R., 2006. Smoking during pregnancy and offspring fat and lean mass in childhood. *Obesity (Silver Spring)* 14, 2284–2293. doi:10.1038/oby.2006.268
- Levine, R.A., Casella, G., 2001. Implementations of the Monte Carlo EM Algorithm. *Journal of Computational and Graphical Statistics* 10, 422–439. doi:10.1198/106186001317115045
- Lin, H., Turnbull, B.W., McCulloch, C.E., Slate, E.H., 2002. Latent Class Models for Joint Analysis of Longitudinal Biomarker and Event Process Data. *Journal of the American Statistical Association* 97, 53–65. doi:10.1198/016214502753479220

- Li, Q., Pan, J., Belcher, J., 2014. Bayesian inference for joint modelling of longitudinal continuous, binary and ordinal events. *Stat Methods Med Res.* doi:10.1177/0962280214526199
- Lucas, A., Fewtrell, M.S., Cole, T.J., 1999. Fetal origins of adult disease-the hypothesis revisited. *BMJ* 319, 245–249.
- Lundgren, E.M., Cnattingius, S., Jonsson, B., Tuvemo, T., 2001. Intellectual and psychological performance in males born small for gestational age with and without catch-up growth. *Pediatr. Res.* 50, 91–96. doi:10.1203/00006450-200107000-00017
- Lunn, D., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D., 2012. *The BUGS Book: A Practical Introduction to Bayesian Analysis.* CRC Press.
- Luo, S., 2014. A Bayesian approach to joint analysis of multivariate longitudinal data and parametric accelerated failure time. *Stat Med* 33, 580–594. doi:10.1002/sim.5956
- MacCallum, R.C., Kim, C., Malarkey, W.B., Kiecolt-Glaser, J.K., 1997. Studying Multivariate Change Using Multilevel Models and Latent Curve Models. *Multivariate Behavioral Research* 32, 215–253. doi:10.1207/s15327906mbr3203_1
- Macdonald-Wallis, C., Lawlor, D.A., Palmer, T., Tilling, K., 2012. Multivariate multilevel spline models for parallel growth processes: application to weight and mean arterial pressure in pregnancy. *Stat Med* 31, 3147–3164. doi:10.1002/sim.5385
- Makowski, D., 2012. *Méthodes statistiques bayésiennes. Bases théoriques et applications en alimentation, environnement et génétique.* INRA/FormaSciences.
- Mamun, A.A., Mannan, M., Doi, S. a. R., 2014. Gestational weight gain in relation to offspring obesity over the life course: a systematic review and bias-adjusted meta-analysis. *Obes Rev* 15, 338–347. doi:10.1111/obr.12132
- Mei, Z., Grummer-Strawn, L.M., Pietrobelli, A., Goulding, A., Goran, M.I., Dietz, W.H., 2002. Validity of body mass index compared with other body-composition screening indexes for the assessment of body fatness in children and adolescents. *Am J Clin Nutr* 75, 978–985.
- Mendez, M.A., Torrent, M., Ferrer, C., Ribas-Fitó, N., Sunyer, J., 2008. Maternal smoking very early in pregnancy is related to child overweight at age 5-7 y. *Am. J. Clin. Nutr.* 87, 1906–1913.
- Monteiro, P.O.A., Victora, C.G., 2005. Rapid growth in infancy and childhood and obesity in later life – a systematic review. *Obesity Reviews* 6, 143–154. doi:10.1111/j.1467-789X.2005.00183.x
- Muthén, B., Shedden, K., 1999. Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics* 55, 463–469.
- Nagin, D.S., Lynam, D., Raudenbush, S., Roeder, K., 1999. Analyzing Developmental Trajectories: A Semiparametric, Group-Based Approach. *Psychological Methods* 4, 139–157.
- Nielsen, G.A., Thomsen, B.L., Michaelsen, K.F., 1998. Influence of breastfeeding and complementary food on growth between 5 and 10 months. *Acta Paediatr.* 87, 911–917.
- Oken, E., Levitan, E.B., Gillman, M.W., 2008. Maternal smoking during pregnancy and child overweight: systematic review and meta-analysis. *Int J Obes (Lond)* 32, 201–210. doi:10.1038/sj.ijo.0803760
- Ong, K.K., Langkamp, M., Ranke, M.B., Whitehead, K., Hughes, I.A., Acerini, C.L., Dunger, D.B., 2009. Insulin-like growth factor I concentrations in infancy predict differential gains in body length and adiposity: the Cambridge Baby Growth Study. *Am. J. Clin. Nutr.* 90, 156–161. doi:10.3945/ajcn.2008.27408

- Ong, K.K.L., Preece, M.A., Emmett, P.M., Ahmed, M.L., Dunger, D.B., ALSPAC Study Team, 2002. Size at birth and early childhood growth in relation to maternal smoking, parity and infant breast-feeding: longitudinal birth cohort study and analysis. *Pediatr. Res.* 52, 863–867. doi:10.1203/00006450-200212000-00009
- Pan, H., Goldstein, H., 1998. Multi-level repeated measures growth modelling using extended spline functions. *Stat Med* 17, 2755–2770.
- Panpanich, R., Garner, P., 2000. Growth monitoring in children. *Cochrane Database Syst Rev* CD001443. doi:10.1002/14651858.CD001443
- Pickles, A., Croudace, T., 2010. Latent mixture models for multivariate and longitudinal outcomes. *Stat Methods Med Res* 19, 271–289. doi:10.1177/0962280209105016
- Pinheiro, J., Bates, D.B., Saikat, D., Deepayan, S., R-core, EISPACK authors, 2015. nlme: Linear and Nonlinear Mixed Effects Models.
- Pinheiro, J.C., Bates, D.M., 2000. *Mixed-Effects Models in S and S-PLUS*. Springer Science & Business Media.
- Pizzi, C., Cole, T.J., Richiardi, L., dos-Santos-Silva, I., Corvalan, C., De Stavola, B., 2014. Prenatal Influences on Size, Velocity and Tempo of Infant Growth: Findings from Three Contemporary Cohorts. *PLoS ONE* 9, e90291. doi:10.1371/journal.pone.0090291
- Plummer, M., Stukalov, A., 2014. rjags: Bayesian graphical models using MCMC.
- Plurphanswat, N., Rodu, B., 2014. The association of smoking and demographic characteristics on body mass index and obesity among adults in the U.S., 1999-2012. *BMC Obes* 1, 18. doi:10.1186/s40608-014-0018-0
- Pourahmadi, M., 1999. Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika* 86, 677–690. doi:10.1093/biomet/86.3.677
- Prader, A., Tanner, J.M., von HARNACK, G., 1963. Catch-up growth following illness or starvation. An example of developmental canalization in man. *J. Pediatr.* 62, 646–659.
- Proust-Lima, C., Joly, P., Dartigues, J.-F., Jacqmin-Gadda, H., 2009. Joint modelling of multivariate longitudinal outcomes and a time-to-event: A nonlinear latent class approach. *Computational Statistics & Data Analysis* 53, 1142–1154. doi:10.1016/j.csda.2008.10.017
- Proust-Lima, C., Séne, M., Taylor, J.M., Jacqmin-Gadda, H., 2014. Joint latent class models for longitudinal and time-to-event data: A review. *Stat Methods Med Res* 23, 74–90. doi:10.1177/0962280212445839
- Regnault, N., Botton, J., Blanc, L., Hankard, R., Forhan, A., Goua, V., Thiebaugeorges, O., Kaminski, M., Heude, B., Charles, M.-A., EDEN mother-child cohort study group, 2011. Determinants of neonatal weight loss in term-infants: specific association with pre-pregnancy maternal body mass index and infant feeding mode. *Arch. Dis. Child. Fetal Neonatal Ed.* 96, F217–222. doi:10.1136/adc.2010.185546
- Regnault, N., Botton, J., Forhan, A., Hankard, R., Thiebaugeorges, O., Hillier, T.A., Kaminski, M., Heude, B., Charles, M.-A., 2010. Determinants of early ponderal and statural growth in full-term infants in the EDEN mother-child cohort study. *Am. J. Clin. Nutr.* 92, 594–602. doi:10.3945/ajcn.2010.29292
- Regnault, N., Gillman, M.W., Kleinman, K., Rifas-Shiman, S., Botton, J., 2014. Comparative study of four growth models applied to weight and height growth data in a cohort of US children from birth to 9 years. *Ann. Nutr. Metab.* 65, 167–174. doi:10.1159/000365894

- Reilly, J.J., Armstrong, J., Dorosty, A.R., Emmett, P.M., Ness, A., Rogers, I., Steer, C., Sherriff, A., Avon Longitudinal Study of Parents and Children Study Team, 2005. Early life risk factors for obesity in childhood: cohort study. *BMJ* 330, 1357. doi:10.1136/bmj.38470.670903.E0
- Riedel, C., Schönberger, K., Yang, S., Koshy, G., Chen, Y.-C., Gopinath, B., Ziebarth, S., von Kries, R., 2014. Parental smoking and childhood obesity: higher effect estimates for maternal smoking in pregnancy compared with paternal smoking--a meta-analysis. *Int J Epidemiol* 43, 1593–1606. doi:10.1093/ije/dyu150
- Rizopoulos, D., Ghosh, P., 2011. A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Stat Med* 30, 1366–1380. doi:10.1002/sim.4205
- Rolland-Cachera, M.F., Deheeger, M., Bellisle, F., Sempé, M., Guilloud-Bataille, M., Patois, E., 1984. Adiposity rebound in children: a simple indicator for predicting obesity. *Am. J. Clin. Nutr.* 39, 129–135.
- Rolland-Cachera, M.F., Deheeger, M., Guilloud-Bataille, M., Avons, P., Patois, E., Sempé, M., 1987. Tracking the development of adiposity from one month of age to adulthood. *Ann. Hum. Biol.* 14, 219–229.
- Royston, P., Altman, D.G., 1994. Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling. *Applied Statistics* 43, 429. doi:10.2307/2986270
- Samson, A., Lavielle, M., Mentré, F., 2007. The SAEM algorithm for group comparison tests in longitudinal data analysis based on non-linear mixed-effects model. *Stat Med* 26, 4860–4875. doi:10.1002/sim.2950
- Silverwood, R.J., De Stavola, B.L., Cole, T.J., Leon, D.A., 2009. BMI peak in infancy as a predictor for later BMI in the Uppsala Family Study. *Int J Obes (Lond)* 33, 929–937. doi:10.1038/ijo.2009.108
- Simondon, K., 1992. Modélisation de la courbe de croissance staturo-pondérale de la première année de la vie, in: *Séminfor 5 : Cinquième Séminaire Informatique de l'ORSTOM, Montpellier (FRA), 1991/09/2-4 (Ed.), Statistique impliquée, Colloques et Séminaires. ORSTOM, Paris, pp. 355–363.*
- Simondon, K.B., Simondon, F., Delpuech, F., Cornu, A., 1992. Comparative study of five growth models applied to weight data from congolese infants between birth and 13 months of age. *American journal of human biology* 4, 327–335.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64, 583–639. doi:10.1111/1467-9868.00353
- Suvak, M.K., Walling, S.M., Iverson, K.M., Taft, C.T., Resick, P.A., 2009. Multilevel regression analyses to investigate the relationship between two variables over time: examining the longitudinal association between intrusion and avoidance. *J Trauma Stress* 22, 622–631. doi:10.1002/jts.20476
- Thiébaud, R., Jacqmin-Gadda, H., Chêne, G., Leport, C., Commenges, D., 2002. Bivariate linear mixed models using SAS proc MIXED. *Computer Methods and Programs in Biomedicine* 69, 249–256. doi:10.1016/S0169-2607(02)00017-2
- Tilling, K., Macdonald-Wallis, C., Lawlor, D.A., Hughes, R.A., Howe, L.D., 2014. Modelling childhood growth using fractional polynomials and linear splines. *Ann. Nutr. Metab.* 65, 129–138. doi:10.1159/000362695
- Timmermans, S.H., Mommers, M., Gubbels, J.S., Kremers, S.P.J., Stafleu, A., Stehouwer, C.D.A., Prins, M.H., Penders, J., Thijs, C., 2013. Maternal smoking during pregnancy

- and childhood overweight and fat distribution: the KOALA Birth Cohort Study. *Pediatr Obes.* doi:10.1111/j.2047-6310.2012.00141.x
- Toschke, A.M., Montgomery, S.M., Pfeiffer, U., von Kries, R., 2003. Early intrauterine exposure to tobacco-inhaled products and obesity. *Am. J. Epidemiol.* 158, 1068–1074.
- Touger, L., Looker, H.C., Krakoff, J., Lindsay, R.S., Cook, V., Knowler, W.C., 2005. Early growth in offspring of diabetic mothers. *Diabetes Care* 28, 585–589.
- tsiatis, A., Davidian, M., 2004. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* 14, 809–834.
- Turcksin, R., Bel, S., Galjaard, S., Devlieger, R., 2014. Maternal obesity and breastfeeding intention, initiation, intensity and duration: a systematic review. *Matern Child Nutr* 10, 166–183. doi:10.1111/j.1740-8709.2012.00439.x
- Veldhuis, J.D., Roemmich, J.N., Richmond, E.J., Rogol, A.D., Lovejoy, J.C., Sheffield-Moore, M., Mauras, N., Bowers, C.Y., 2005. Endocrine control of body composition in infancy, childhood, and puberty. *Endocr. Rev.* 26, 114–146. doi:10.1210/er.2003-0038
- Victora, C.G., Barros, F.C., Horta, B.L., Martorell, R., 2001. Short-term benefits of catch-up growth for small-for-gestational-age infants. *Int J Epidemiol* 30, 1325–1330.
- von Kries, R., Bolte, G., Baghi, L., Toschke, A.M., GME Study Group, 2008. Parental smoking and childhood obesity--is maternal smoking in pregnancy the critical exposure? *Int J Epidemiol* 37, 210–216. doi:10.1093/ije/dym239
- Wagner, H., Tüchler, R., 2010. Bayesian estimation of random effects models for multivariate responses of mixed data. *Computational Statistics & Data Analysis* 54, 1206–1218. doi:10.1016/j.csda.2009.12.007
- Watanabe, S., 2010. Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. arXiv:1004.2316 [cs].
- Wei, G.C.G., Tanner, M.A., 1990. A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms. *Journal of the American Statistical Association* 85, 699–704. doi:10.1080/01621459.1990.10474930
- Weng, S.F., Redsell, S.A., Swift, J.A., Yang, M., Glazebrook, C.P., 2012. Systematic review and meta-analyses of risk factors for childhood overweight identifiable during infancy. *Arch. Dis. Child.* 97, 1019–1026. doi:10.1136/archdischild-2012-302263
- Wen, X., Kleinman, K., Gillman, M.W., Rifas-Shiman, S.L., Taveras, E.M., 2012. Childhood body mass index trajectories: modeling, characterizing, pairwise correlations and socio-demographic predictors of trajectory characteristics. *BMC Med Res Methodol* 12, 38. doi:10.1186/1471-2288-12-38
- Whitaker, R.C., Pepe, M.S., Wright, J.A., Seidel, K.D., Dietz, W.H., 1998. Early adiposity rebound and the risk of adult obesity. *Pediatrics* 101, E5.
- Whittaker, E.T., 1922. On a New Method of Graduation. *Proceedings of the Edinburgh Mathematical Society* 41, 63–75. doi:10.1017/S001309150000359X
- Willemsen, S.P., Eilers, P.H.C., Steegers-Theunissen, R.P.M., Lesaffre, E., 2015. A multivariate Bayesian model for embryonic growth. *Statist. Med.* 34, 1351–1365. doi:10.1002/sim.6411
- Williams, S., Davie, G., Lam, F., 1999. Predicting BMI in young adults from childhood data using two approaches to modelling adiposity rebound. *Int. J. Obes. Relat. Metab. Disord.* 23, 348–354.

- World Health Organization Working Group on Infant Growth, 1995. An evaluation of infant growth: the use and interpretation of anthropometry in infants. WHO Working Group on Infant Growth. Bull. World Health Organ. 73, 165-174.
- Zhang, Z., 2013. Bayesian growth curve models with the generalized error distribution. Journal of Applied Statistics 40, 1779-1795. doi:10.1080/02664763.2013.796348
- Zheng, W., Suzuki, K., Shinohara, R., Sato, M., Yokomichi, H., Yamagata, Z., 2015. Maternal smoking during pregnancy and growth in infancy: a covariance structure analysis. J Epidemiol 25, 44-49. doi:10.2188/jea.JE20140040

ANNEXES

Annexe 1 : Comparaison des modèles décrivant la croissance du poids et de la taille selon les modèles d'erreur utilisés

Variable	Modele Erreur	Garçons		Filles	
		AIC	BIC	AIC	BIC
Poids	Constant	24242	24314	21466	21536
	Proportionnel	20644	20715	17472	17542
	Combiné	20507	20584	17391	17466
Taille	Constant	56086	56158	50540	50610
	Proportionnel	56284	56355	50766	50836
	Combiné	55983	56059	50441	50516

Annexe 2 : Code rjags, modèle bayésien décrivant conjointement la croissance du poids et de la taille

```

model
{
  for (j in 1:N)
  {

#Sous-modèle de croissance pondérale
  y_pds[j]~dnorm(mu_pds[j],tau.e_pds[j])
  mu_pds[j]<-exp(beta[id[j],1])+exp(beta[id[j],2])*t[j]+exp(beta[id[j],3])*(1-exp(-exp(beta[id[j],4])*t[j]))

#Sous-modèle de croissance staturale
  y_tail[j]~dnorm(mu_tail[j],tau.e_tail[j])
  mu_tail[j]<-exp(beta[id[j],5])+exp(beta[id[j],6])*t[j]+exp(beta[id[j],7])*(1-exp(-exp(beta[id[j],8])*t[j]))

#Modèle d'erreur poids
  tau.e_pds[j]<-1/exp(a_pds+b_pds*t[j]+c_pds*t[j]*t[j])
  var.e_pds[j]<-1/tau.e_pds[j]
  sigma.e_pds[j]<-pow(tau.e_pds[j],-0.5)

#Modèle d'erreur taille
  tau.e_tail[j]<-1/exp(a_tail+b_tail*t[j])
  var.e_tail[j]<-1/tau.e_tail[j]
  sigma.e_tail[j]<-pow(tau.e_tail[j],-0.5)

  }

#Paramètres de croissance individuels

  for (f in 1:F)
  {
    theta[f,1]<-
    gamma[1]+betaSG_feminin[1]*feminin[f]+betaSG_tabgro1[1]*tabgrosse_1[f]+betaSG_tabgro2[1]*tabgrosse_2[f]
    +
    betaSG_niveau1[1]*c24_Niveau_tab3cat_1[f]+betaSG_niveau2[1]*c24_Niveau_tab3cat_2[f]+betaSG_obes[1]*obes3[f]
    +betaSG_centre[1]*centre01[f]+betaSG_allait[1]*allait[f]+betaSG_age[1]*(age_acc[f]-30)

    theta[f,2]<-
    gamma[2]+betaSG_feminin[2]*feminin[f]+betaSG_tabgro1[2]*tabgrosse_1[f]+betaSG_tabgro2[2]*tabgrosse_2[f]
    +
    betaSG_niveau1[2]*c24_Niveau_tab3cat_1[f]+betaSG_niveau2[2]*c24_Niveau_tab3cat_2[f]+betaSG_obes[2]*obes3[f]
    +betaSG_centre[2]*centre01[f]+betaSG_allait[2]*allait[f]+betaSG_age[2]*(age_acc[f]-30)

    theta[f,3]<-
    gamma[3]+betaSG_feminin[3]*feminin[f]+betaSG_tabgro1[3]*tabgrosse_1[f]+betaSG_tabgro2[3]*tabgrosse_2[f]
    +
    betaSG_niveau1[3]*c24_Niveau_tab3cat_1[f]+betaSG_niveau2[3]*c24_Niveau_tab3cat_2[f]+betaSG_obes[3]*obes3[f]
    +betaSG_centre[3]*centre01[f]+betaSG_allait[3]*allait[f]+betaSG_age[3]*(age_acc[f]-30)
  }
}

```

```

    theta[f,4]<-
gamma[4]+betaSG_feminin[4]*feminin[f]+betaSG_tabgro1[4]*tabgrosse_1[f]+betaSG_tabgro2[4]*tabgrosse_2[f
]+
betaSG_niveau1[4]*c24_Niveau_tab3cat_1[f]+betaSG_niveau2[4]*c24_Niveau_tab3cat_2[f]+betaSG_obes[4]*obes3[f
]+betaSG_centre[4]*centre01[f]+betaSG_allait[4]*allait[f]+betaSG_age[4]*(age_acc[f]-30)

    theta[f,5]<-
gamma[5]+betaSG_feminin[5]*feminin[f]+betaSG_tabgro1[5]*tabgrosse_1[f]+betaSG_tabgro2[5]*tabgrosse_2[f
]+
betaSG_niveau1[5]*c24_Niveau_tab3cat_1[f]+betaSG_niveau2[5]*c24_Niveau_tab3cat_2[f]+betaSG_obes[5]*obes3[f
]+betaSG_centre[5]*centre01[f]+betaSG_allait[5]*allait[f]+betaSG_age[5]*(age_acc[f]-30)

    theta[f,6]<-
gamma[6]+betaSG_feminin[6]*feminin[f]+betaSG_tabgro1[6]*tabgrosse_1[f]+betaSG_tabgro2[6]*tabgrosse_2[f
]+
betaSG_niveau1[6]*c24_Niveau_tab3cat_1[f]+betaSG_niveau2[6]*c24_Niveau_tab3cat_2[f]+betaSG_obes[6]*obes3[f
]+betaSG_centre[6]*centre01[f]+betaSG_allait[6]*allait[f]+betaSG_age[6]*(age_acc[f]-30)

    theta[f,7]<-
gamma[7]+betaSG_feminin[7]*feminin[f]+betaSG_tabgro1[7]*tabgrosse_1[f]+betaSG_tabgro2[7]*tabgrosse_2[f
]+
betaSG_niveau1[7]*c24_Niveau_tab3cat_1[f]+betaSG_niveau2[7]*c24_Niveau_tab3cat_2[f]+betaSG_obes[7]*obes3[f
]+betaSG_centre[7]*centre01[f]+betaSG_allait[7]*allait[f]+betaSG_age[7]*(age_acc[f]-30)

    theta[f,8]<-
gamma[8]+betaSG_feminin[8]*feminin[f]+betaSG_tabgro1[8]*tabgrosse_1[f]+betaSG_tabgro2[8]*tabgrosse_2[f
]+
betaSG_niveau1[8]*c24_Niveau_tab3cat_1[f]+betaSG_niveau2[8]*c24_Niveau_tab3cat_2[f]+betaSG_obes[8]*obes3[f
]+betaSG_centre[8]*centre01[f]+betaSG_allait[8]*allait[f]+betaSG_age[8]*(age_acc[f]-30)

    beta[f,1:8]~dmnorm(theta[f,1:8],T[1:8,1:8])
}

#Lois a priori sur les paramètres des Modeles d'erreur
#Poids
a_pds~dnorm(0,16)
b_pds~dnorm(0,16)
c_pds~dnorm(0,16)

#Taille
a_tail~dnorm(0,16)
b_tail~dnorm(0,16)

```

```

#A priori sur les intercepts des paramètres de croissance

#Ap
  gamma[1]~dnorm(1,16)
#Bp
  gamma[2]~dnorm(5,16)
#Cp
  gamma[3]~dnorm(1.5,16)
#Dp
  gamma[4]~dnorm(5,16)
#At
  gamma[5]~dnorm(4,16)
#Bt
  gamma[6]~dnorm(4,16)
#Ct
  gamma[7]~dnorm(3,16)
#Dt
  gamma[8]~dnorm(5,16)

#Lois a priori sur coefficients de régression

for (z in 1:8){
  betaSG_feminin[z]~dnorm(0,16)
  betaSG_tabgro1[z]~dnorm(0,16)
  betaSG_tabgro2[z]~dnorm(0,16)
  betaSG_niveau1[z]~dnorm(0,16)
  betaSG_niveau2[z]~dnorm(0,16)
  betaSG_obes[z]~dnorm(0,16)
  betaSG_centre[z]~dnorm(0,16)
  betaSG_allait[z]~dnorm(0,16)
  betaSG_age[z]~dnorm(0,16)
}

#Loi A priori Paramètres individuels
sigma.u[1:8,1:8]<-inverse(T[1:8,1:8])
T[1:8,1:8]~dwish(S[1:8,1:8],9)

for (z in 1:30000){
  for (k in 1:nbmesure){

#Poids Prédit
pred_pds[z,k]<-exp(beta[idfictif_num[z],1])+exp(beta[idfictif_num[z],2])*jpred[k]+exp(beta[idfictif_num[z],3])*(1-
exp(-exp(beta[idfictif_num[z],4])*jpred[k]))

#Taille Prédite
pred_tail[z,k]<-exp(beta[idfictif_num[z],5])+exp(beta[idfictif_num[z],6])*jpred[k]+exp(beta[idfictif_num[z],7])*(1-
exp(-exp(beta[idfictif_num[z],8])*jpred[k]))

#IMC Prédit
pred_bmi[z,k]<-pred_pds[z,k]/(pred_tail[z,k]/100)^2

```

```

}}

for (k in 1:nbmesure){

#Poids moyen par sous-groupe de tabagisme maternel
mean_pds_NF[k]<-mean(pred_pds[first_NF:last_NF,k])
mean_pds_F1T[k]<-mean(pred_pds[first_F1T:last_F1T,k])
mean_pds_F23T[k]<-mean(pred_pds[first_F23T:last_F23T,k])
mean_diff_pds_F1T_NF[k]<-mean_pds_F1T[k]-mean_pds_NF[k]
mean_diff_pds_F23T_NF[k]<-mean_pds_F23T[k]-mean_pds_NF[k]

#Taille moyenne par sous-groupe de tabagisme maternel
mean_tail_NF[k]<-mean(pred_tail[first_NF:last_NF,k])
mean_tail_F1T[k]<-mean(pred_tail[first_F1T:last_F1T,k])
mean_tail_F23T[k]<-mean(pred_tail[first_F23T:last_F23T,k])
mean_diff_tail_F1T_NF[k]<-mean_tail_F1T[k]-mean_tail_NF[k]
mean_diff_tail_F23T_NF[k]<-mean_tail_F23T[k]-mean_tail_NF[k]

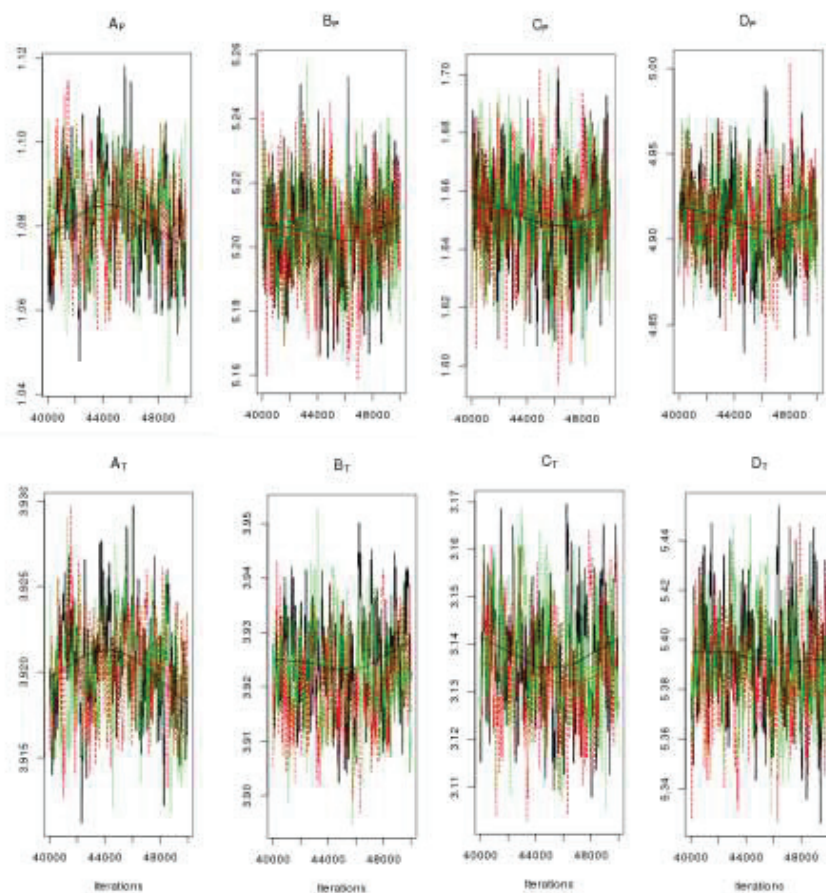
## moyen par sous-groupe de tabagisme maternel

mean_bmi_NF[k]<-mean(pred_bmi[first_NF:last_NF,k])
mean_bmi_F1T[k]<-mean(pred_bmi[first_F1T:last_F1T,k])
mean_bmi_F23T[k]<-mean(pred_bmi[first_F23T:last_F23T,k])
mean_diff_bmi_F1T_NF[k]<-mean_bmi_F1T[k]-mean_bmi_NF[k]
mean_diff_bmi_F23T_NF[k]<-mean_bmi_F23T[k]-mean_bmi_NF[k]
}

}

```


Annexe 3 : Mélange des trois chaînes pour les intercepts des paramètres de croissance



Annexe 4 : Différences moyennes d'IMC entre les enfants exposés au tabagisme maternel durant le 1^{er} trimestre exclusivement ou tout au long de la grossesse et les enfants non exposés de la naissance à 5 ans

Variable	Tabagisme Maternel			
	1er Trimestre Exclusivement		Tout au long de la grossesse	
	Coefficient a posteriori	Intervalle de Crédibilité à 95%	Coefficient a posteriori	Intervalle de Crédibilité à 95%
Différence d'IMC moyen, kg/m²				
5 jours	0,14	[-0,09 ; 0,38]	-0,05	[-0,23 ; 0,12]
15 jours	0,12	[-0,10 ; 0,34]	0,01	[-0,15 ; 0,17]
25 jours	0,09	[-0,12 ; 0,31]	0,06	[-0,10 ; 0,22]
30 jours	0,08	[-0,13 ; 0,29]	0,07	[-0,09 ; 0,24]
60 jours	0,02	[-0,21 ; 0,26]	0,14	[-0,05 ; 0,32]
90 jours	-0,01	[-0,25 ; 0,24]	0,16	[-0,04 ; 0,36]
120 jours	-0,02	[-0,27 ; 0,23]	0,16	[-0,04 ; 0,36]
150 jours	-0,03	[-0,28 ; 0,23]	0,16	[-0,04 ; 0,36]
180 jours	-0,02	[-0,27 ; 0,24]	0,15	[-0,04 ; 0,36]
210 jours	-0,01	[-0,27 ; 0,25]	0,15	[-0,05 ; 0,36]
240 jours	0,00	[-0,26 ; 0,26]	0,15	[-0,05 ; 0,36]
270 jours	0,01	[-0,24 ; 0,27]	0,15	[-0,05 ; 0,35]
300 jours	0,03	[-0,23 ; 0,28]	0,15	[-0,05 ; 0,35]
330 jours	0,04	[-0,21 ; 0,30]	0,15	[-0,05 ; 0,35]
360 jours	0,05	[-0,20 ; 0,31]	0,15	[-0,04 ; 0,35]
365 jours	0,06	[-0,20 ; 0,31]	0,15	[-0,04 ; 0,35]
390 jours	0,07	[-0,19 ; 0,32]	0,15	[-0,04 ; 0,35]
420 jours	0,08	[-0,18 ; 0,33]	0,16	[-0,04 ; 0,35]
450 jours	0,08	[-0,17 ; 0,33]	0,16	[-0,03 ; 0,36]
480 jours	0,09	[-0,15 ; 0,34]	0,17	[-0,02 ; 0,36]
510 jours	0,10	[-0,15 ; 0,35]	0,17	[-0,02 ; 0,36]
540 jours	0,11	[-0,14 ; 0,35]	0,17	[-0,01 ; 0,36]
570 jours	0,11	[-0,13 ; 0,35]	0,18	[-0,01 ; 0,36]
600 jours	0,11	[-0,13 ; 0,36]	0,18	[0,00 ; 0,36]
630 jours	0,12	[-0,12 ; 0,36]	0,18	[0,00 ; 0,36]
660 jours	0,12	[-0,12 ; 0,36]	0,19	[0,01 ; 0,37]
690 jours	0,12	[-0,12 ; 0,36]	0,19	[0,01 ; 0,37]
720 jours	0,12	[-0,12 ; 0,36]	0,19	[0,02 ; 0,37]
730 jours	0,12	[-0,12 ; 0,36]	0,19	[0,02 ; 0,37]
750 jours	0,12	[-0,11 ; 0,36]	0,20	[0,02 ; 0,37]
780 jours	0,12	[-0,11 ; 0,35]	0,20	[0,02 ; 0,37]
810 jours	0,12	[-0,11 ; 0,35]	0,20	[0,03 ; 0,37]
840 jours	0,12	[-0,11 ; 0,35]	0,20	[0,03 ; 0,37]
870 jours	0,11	[-0,11 ; 0,34]	0,20	[0,03 ; 0,38]

Variable	Tabagisme Maternel			
	1er Trimestre Exclusivement		Tout au long de la grossesse	
	Coefficient a posteriori	Intervalle de Crédibilité à 95%	Coefficient a posteriori	Intervalle de Crédibilité à 95%
Différence d'IMC moyen, kg/m²				
900 jours	0,11	[-0,11 ; 0,34]	0,21	[0,04 ; 0,38]
930 jours	0,11	[-0,11 ; 0,33]	0,21	[0,04 ; 0,38]
960 jours	0,11	[-0,12 ; 0,33]	0,21	[0,04 ; 0,38]
990 jours	0,10	[-0,12 ; 0,32]	0,21	[0,04 ; 0,38]
1020 jours	0,10	[-0,12 ; 0,32]	0,21	[0,05 ; 0,38]
1050 jours	0,09	[-0,13 ; 0,31]	0,21	[0,05 ; 0,38]
1080 jours	0,09	[-0,13 ; 0,31]	0,21	[0,05 ; 0,38]
1095 jours	0,09	[-0,13 ; 0,31]	0,21	[0,05 ; 0,38]
1110 jours	0,09	[-0,13 ; 0,30]	0,21	[0,05 ; 0,38]
1140 jours	0,08	[-0,14 ; 0,30]	0,22	[0,05 ; 0,38]
1170 jours	0,08	[-0,14 ; 0,29]	0,22	[0,05 ; 0,38]
1200 jours	0,07	[-0,14 ; 0,29]	0,22	[0,05 ; 0,38]
1230 jours	0,07	[-0,15 ; 0,29]	0,22	[0,05 ; 0,38]
1260 jours	0,06	[-0,15 ; 0,28]	0,22	[0,05 ; 0,38]
1290 jours	0,06	[-0,16 ; 0,27]	0,22	[0,05 ; 0,38]
1320 jours	0,05	[-0,16 ; 0,27]	0,22	[0,05 ; 0,39]
1350 jours	0,05	[-0,17 ; 0,27]	0,22	[0,05 ; 0,39]
1380 jours	0,05	[-0,17 ; 0,26]	0,22	[0,05 ; 0,39]
1410 jours	0,04	[-0,17 ; 0,26]	0,22	[0,05 ; 0,39]
1440 jours	0,04	[-0,18 ; 0,25]	0,22	[0,05 ; 0,39]
1470 jours	0,03	[-0,19 ; 0,25]	0,22	[0,05 ; 0,39]
1500 jours	0,03	[-0,19 ; 0,24]	0,22	[0,04 ; 0,39]
1530 jours	0,02	[-0,20 ; 0,24]	0,21	[0,04 ; 0,39]
1560 jours	0,02	[-0,20 ; 0,24]	0,21	[0,04 ; 0,39]
1590 jours	0,01	[-0,21 ; 0,23]	0,21	[0,04 ; 0,39]
1620 jours	0,01	[-0,21 ; 0,23]	0,21	[0,04 ; 0,39]
1650 jours	0,00	[-0,22 ; 0,23]	0,21	[0,04 ; 0,39]
1680 jours	0,00	[-0,22 ; 0,22]	0,21	[0,04 ; 0,39]
1710 jours	0,00	[-0,23 ; 0,22]	0,21	[0,03 ; 0,40]
1740 jours	-0,01	[-0,23 ; 0,22]	0,21	[0,03 ; 0,40]
1770 jours	-0,01	[-0,24 ; 0,21]	0,21	[0,03 ; 0,40]
1800 jours	-0,02	[-0,24 ; 0,21]	0,21	[0,03 ; 0,40]
1825 jours	-0,02	[-0,25 ; 0,21]	0,21	[0,03 ; 0,40]

Annexe 5 : Articles

1 **A novel method to describe early offspring body mass index (BMI) trajectories:**
2 **application to maternal smoking in pregnancy.**

3 Sophie Carles^{a,b,c*}, Marie-Aline Charles^{a,b}, Anne Forhan^{a,b}, Rémy Slama^{d,e}, Barbara Heude^{a,b},
4 Jérémie Botton^{a,c,f} and the EDEN mother child study group

5 Author affiliations:

6 ^aORCHAD Team, INSERM, UMR1153 Epidemiology and Biostatistics Sorbonne Paris Cité
7 Center (CRESS), Paris, F-75014 France

8 ^bParis Descartes University, Paris, France

9 ^cUniv Paris-Sud, Villejuif, France

10 ^dTeam of Environmental Epidemiology applied to Reproduction and Respiratory Health,
11 Institut Albert Bonniot, U 823, INSERM, Grenoble, France

12 ^eGrenoble University, Institut Albert Bonniot, Grenoble, France

13 ^fLaboratoire de biomathématique, Faculté de Pharmacie, Univ Paris-Sud, Châtenay-Malabry,
14 F-92290, France

15 Corresponding Authorship :

16 *Corresponding author

17 E-mail address: sophie.carles@inserm.fr (SC)

18

19

20 **A novel method to describe early offspring body mass index (BMI) trajectories:**
21 **application to maternal smoking in pregnancy.**
22

23 **Abstract**

24 **Background:** Accurately characterizing children's body mass index (BMI) trajectories and
25 studying their determinants is a statistical challenge. There is a need to identify early public
26 health measures for obesity prevention. We describe a method that allows studies of the
27 determinants of height, weight and BMI growth up to five years of age. We illustrated this
28 method using maternal smoking during pregnancy as one of the early-life factors that is
29 potentially involved in prenatal programming of obesity.

30 **Methods:** Individual height and weight trajectories were fitted using the Jenss-Bayley model
31 on 28,381 and 30,515 measurements, respectively, from 1,666 children to deduce BMI
32 trajectories. We assessed global associations between smoking and growth trajectories and
33 cross-sectional associations at specific ages.

34 **Results:** Children exposed in late pregnancy had a 0.24 kg/m² (95% confidence interval: 0.07,
35 0.41) higher BMI at 5 years of age compared with non-exposed children. Although the BMIs
36 of children exposed during late pregnancy became significantly higher compared with those
37 of non-exposed children from 2 years onwards, the trajectories began to diverge during the
38 first weeks of life.

39 **Conclusion:** Our method is relevant for studies on the relationships between individual-level
40 exposures and the dynamics and shapes of BMI growth during childhood, including key
41 features such as instantaneous growth velocities and the age or BMI value at the BMI infancy
42 peak that benefit from the monotonic pattern of height and weight growth.

43 **Abbreviations:** CI, confidence interval; BMI, body mass index; EDEN, Etude des
44 Déterminants pré et postnatals précoces du développement et de la santé de l'Enfant

45 **Keywords:** Postnatal growth; Maternal smoking; Prenatal exposure; Body Mass Index;
46 Developmental origins of health and disease; critical windows

47 **Introduction**

48 Maternal smoking during pregnancy has been consistently associated with an
49 increased risk of obesity in adulthood [1]. There is growing evidence that this association
50 appears in childhood, but uncertainties remain regarding the timing and the existence of
51 critical time-windows of exposure during pregnancy. The paucity of longitudinal studies and
52 the lack of relevant methodologies do not allow the precise characterization of the pattern of
53 growth of children from smoking mothers and the identification of when the difference in the
54 body mass index (BMI) between exposed and non-exposed children emerges.

55 In a broad context, group- and individual-based approaches are the main methods used
56 to model BMI in childhood. Group-based approaches consider population heterogeneity in
57 trajectories to identify subgroups sharing similar growth patterns [2,3], but the relevance of
58 the number of groups chosen by the analyst can be criticized [4]. Individual-based approaches
59 provide individual BMI trajectories, but the non-monotonic shape of the BMI may require
60 complex modeling methods. Models using (fractional) polynomials [5,6] require the
61 determination of the relevant polynomial order and combination of power(s). High order
62 polynomials better fit the data but often exhibit a poor fit at the extremes [6]. Spline models
63 [7,8] are more flexible, but the number and localization of the knots need to be chosen with
64 some degree of arbitrariness. Moreover, these methods do not provide biologically
65 interpretable parameters for growth.

66 We describe a method that allows studies of the determinants of an offspring's height,
67 weight and BMI growth up to five years of age based on non-linear modeling of the
68 individual height and weight growth curves [9]. We illustrated this method with maternal
69 smoking during pregnancy as one of the early-life factors that is potentially involved in the

70 prenatal programming of obesity. We show how our method (later called “indirect modeling,
71 IM”) provides insights into the timing of the emergence of potential BMI differences in early
72 life.

73 **Materials and Methods**

74 **Population**

75 EDEN is an ongoing French mother-child cohort established to assess pre- and early
76 postnatal determinants of child health and development [10]. Between 2003 and 2006, 2,002
77 pregnant women were recruited prior to 24 weeks of gestation defined by the date of the last
78 menstrual period in the Nancy and Poitiers hospitals. Exclusion criteria were multiple
79 pregnancies, personal history of diabetes, illiteracy, and intention to deliver outside the
80 university hospital or to move outside the region within 3 years. Spouses willing to take part
81 were invited to participate; the enrolled pregnant women had to complete a questionnaire and
82 attend a clinical examination where their height was recorded by research midwives. Written
83 informed consent was obtained from the parents for themselves and from both parents for the
84 child after delivery. The study was approved by the ethics committee of Kremlin-Bicêtre and
85 declared to the National Committee for Processed Data and Freedom.

86 Ninety-five enrolled women left the study due to miscarriages, *in utero* deaths or
87 discontinuation of pregnancy for medical reasons. Among the 1,907 enrolled children, 1,666
88 had more than one postnatal length/height and weight measurement and complete data on
89 potential confounders; these children were included in the analyses.

90 **Maternal smoking**

91 During the first visit with the research midwives (between 24 and 28 weeks of
92 gestation), the mothers reported their current smoking status (used to characterize the 2nd
93 trimester status), daily cigarette consumption and smoking habits at the beginning of

94 pregnancy (used for the 1st trimester status). After delivery, similar information for smoking at
95 the end of the 3rd trimester of pregnancy was collected by the research midwives (used for the
96 3rd trimester status). We investigated differences between “Non-smoking mothers”, “Smokers
97 exclusively in the 1st trimester” and “Smokers in late pregnancy” (i.e., smoking at least during
98 the second and/or third trimesters). Additionally, we calculated the average daily number of
99 cigarettes in the second and third trimesters.

100 **Anthropometric measurements**

101 The child’s birthweight, daily weights on days 1 to 5 and birth length were extracted
102 from the obstetric records. Length/height and weight postnatal measurements were collected
103 from routine checks via mailed questionnaires and during the study’s clinical visits at 1, 3 and
104 5 years. Mothers reported lengths/heights and weights recorded by their child’s health
105 practitioners on personal child health records in the 4 preceding months (4-month, 8-month,
106 1-year questionnaires) or trimesters (2-, 3-, 4- and 5-year questionnaires). At the 1-year visit,
107 mothers were weighed (to the nearest 0.1 kg) alone and holding their child and the child’s
108 weight was determined by subtracting the two measurements. The child’s weight was directly
109 measured at the 3- and 5-year visits. The height was measured to the nearest 0.5 cm.

110 The median [min-max] numbers of measurements were 18 [2-31] and 19 [2-32] for height and
111 weight, respectively (Supporting Information S1). Among the enrolled children, 1,203 (72%)
112 and 1,208 (73%) had at least one height and weight measurement after four years,
113 respectively.

114 **Potential confounders**

115 During the first visit, the maternal height was measured to the nearest 0.2 cm by a
116 midwife and the mothers reported their pre-pregnancy weight, age and education level. Their
117 pre-pregnancy BMI was calculated as weight (in kilograms) for measured height (in meters)
118 squared. We determined the overweight (including obesity) status as a BMI ≥ 25 kg/m².

119 Breast-feeding data were extracted from the medical records during the hospital maternity
120 stay, at discharge and through the mothers' self-administered postnatal mailed questionnaires
121 at 4 and 8 months and at 1 and 2 years. The duration of any (exclusive or partial) breast-
122 feeding practices was calculated [11] to identify children breast-fed less than 3 months or
123 longer.

124 **Growth trajectory modeling**

125 Using a recently proposed methodology [9], the individual weight and height
126 trajectories were estimated using the non-linear Jenss-Bayley equation through a mixed
127 effects modeling approach including variance modeling to take heteroscedasticity into
128 account (Supporting Information S2). By adding individual random effects in each equation
129 parameter, a set of parameters was used to characterize each child's growth trajectory. To
130 bypass neonatal weight loss not correctly fitted by the model, birthweight measurements were
131 excluded and the minimal weight in the first four days of life was selected for each child. The
132 models were fitted using the 'SAEMIX' package in the R software [12].

133 Given their non-monotonic shape in early childhood, the BMI trajectories were not
134 directly modeled but were calculated from the modeled height and weight [9]. The mean
135 population and individual BMI trajectories were determined using the height and weight
136 equations (Supporting Information S2).

137 **Statistical analysis**

138 Whether the child's global weight or length/height growth shape differed according to
139 the maternal smoking status was assessed using the "one-step" method. Associations between
140 maternal smoking and anthropometric parameters at birth, the predicted height, weight and
141 BMI, their corresponding velocities at specific ages, and age/BMI at the BMI infancy peak
142 were assessed using the "two-step" method (Fig 1). The analyses were performed on R v3.0
143 and P values <0.05 were considered significant.

144 **Fig 1. Levels of the “One-step” and “Two-step” Modeling Methods of Growth for**
145 **Children from the EDEN study aged from birth to 5 years.**

146 **One-step approach**

147 Gender and potential confounders (maternal education level, age, BMI, breast-feeding
148 duration and recruitment center) were introduced as fixed effects in the length/height and
149 weight models. The general form of the length/height model is given (Equation 1A) and the
150 form decomposing the individual parameters is also detailed (Equation 1B).

151

152 **Equation 1A.**

$$Height_{i,j} = \exp(A_{H_i}) + \exp(B_{H_i}) * t_{i,j} + \exp(C_{H_i}) * (1 - \exp(-\exp(D_{H_i}) * t_{i,j})) + e_{i,j}$$

153

154 **Equation 1B.**

$$Height_{i,j} = \exp(\mu_{A_H} + \beta_{A_H} * Z + u_{A_{H_i}}) + \exp(\mu_{B_H} + \beta_{B_H} * Z + u_{B_{H_i}}) * t_{i,j} \\ + \exp(\mu_{C_H} + \beta_{C_H} * Z + u_{C_{H_i}}) \\ * (1 - \exp(-\exp(\mu_{D_H} + \beta_{D_H} * Z + u_{D_{H_i}}) * t_{i,j})) + e_{i,j}$$

155

156 *Height*_{*i,j*} describes the height (cm) of child *i* at time *t*_{*i,j*} (days), exp(*A_H*) describes the birth
157 length, exp(*B_H*) describes the growth velocity beyond two years, *C_H* is the spurt of growth in
158 the first months of life and *D_H* is the curvature of the trajectories in the first age period. The
159 parameter definitions are illustrated in figure 2 and in Supporting Information S2 (Fig S1).

160 **Fig 2. Weight Growth Trajectories from birth to 5 years by deciles of weight growth**
161 **parameters (Ai, Bi, Ci and Di) in Children of the EDEN Study, France, 2003-2012.**

162 Figure 2 describes the influence of each of the parameters (and more specifically their
163 increase) (Supporting Information S2, Table S2 provides decile values) on the growth of the
164 global shape. β_{A_H} , β_{B_H} , β_{C_H} , and β_{D_H} represent covariate effects from vector *Z* (maternal
165 smoking status during pregnancy; potential confounders) on *A_H*, *B_H*, *C_H*, and *D_H*, respectively.
166 *u_{A_H}*, *u_{B_H}*, *u_{C_H}*, and *u_{D_H}* are individual random effects of *A_H*, *B_H*, *C_H*, and *D_H*, respectively. *e_{i,j}*
167 is the residual error. μ_{A_H} , μ_{B_H} , μ_{C_H} , and μ_{D_H} represent the fixed effects of *A_H*, *B_H*, *C_H*, and *D_H*,
168 respectively, for children with *Z*=0. The weight was modeled similarly. As the first
169 measurement is the minimal weight within the first 4 days, *A_W* corresponds to the
170 extrapolation of the birth weight without considering neonatal weight loss; thus, the values are
171 lower than the observed measurements. Longitudinal associations between maternal smoking
172 and the length/height and weight trajectories were assessed through comparison of the nested
173 models using likelihood ratio tests.

174 The BMI of individual i at time j was predicted as $BMI_{i,j} = \frac{Weight_{i,j}}{(Height_{i,j}/100)^2}$ with the weight
175 expressed in kg.

176 **Two-step approach**

177 Individual height, weight and BMI were predicted at 0.25, 0.5, 1, 2, 4 and 5 years
178 using individual equations fitted separately for boys and girls without covariates. The
179 individual height (Equation 2), weight and BMI (Equation 3) velocities were calculated at
180 these ages by calculating the first-derivative equations.

181 **Equation 2.**

$$HeightVelocity_{i,j} = \exp(B_{H_i}) + \exp(C_{H_i} + D_{H_i} - \exp(D_{H_i}) * t_{i,j})$$

182

183 **Equation 3.**

$$BMIVelocity_{i,j} = \frac{WeightVelocity_{i,j} * \left(\frac{Height_{i,j}}{100}\right)^2 - 2 * \frac{Height_{i,j}}{100} * \frac{HeightVelocity_{i,j}}{100} * Weight_{i,j}}{\left(\frac{Height_{i,j}}{100}\right)^4}$$

184 The individual daily BMI velocities were calculated from birth to five years of age. The age at
185 the BMI infancy peak was the time point when the BMI velocity signs changed
186 ($BMIVelocity_{i,j}/BMIVelocity_{i,j+1} < 0$); the BMI at the BMI infancy peak was calculated for this
187 age. The BMI infancy peaks identified after 2 years of age were excluded from the analyses
188 after the graphical examination of trajectories (n=49).

189 Associations between the smoking status and measured birth length/weight, predicted weight,
190 length/height and BMI at the selected ages as well as the age and BMI at the BMI infancy
191 peak were assessed using multivariable linear regression models adjusted for confounders.

192 **Complementary analysis**

193 To explore the dose-response relationship between maternal smoking in late pregnancy
194 and the child's growth, the analyses were repeated considering the average number of
195 cigarettes smoked daily in the second and third trimesters as the main exposure.

196 To evaluate the association between postnatal smoking and growth, we compared the
197 children's growth from pre- and postnatal non-smoking mothers and non-smoking fathers to
198 the children from non-smoking mothers during pregnancy that (re-)started smoking after birth
199 or whose father smoked postnatally.

200 **Sensitivity analyses**

201 The analyses were repeated after excluding children born before 37 weeks of gestation
202 to assess any confounding by preterm children. To assess the potential selection bias caused
203 by missing length/height or weight data, the analyses were repeated after excluding children
204 lacking measurements after 4 years of age. To discuss potential differences with other BMI
205 modeling approaches, we replicated the 2-step approach using fractional polynomials and
206 splines. Finally, we report analyses on anthropometric measurements from clinical visits at 1,
207 3 and 5 years.

208 **Results**

209 **Population characteristics**

210 Among the enrolled mothers, 24% smoked at least once during pregnancy: 132 (7.9%)
211 exclusively in the first trimester and 275 (16.5%) also during the second and/or third
212 trimester(s) (Table 1). These prevalences were similar for mothers with missing data on
213 confounders (8.5% and 16.5%, respectively).

214

215 **Table 1. Children (n = 1,666) and Parental Characteristics by Maternal Smoking Status**
 216 **during pregnancy from the EDEN study, 2003-2012**

	No smoking (n = 1259, 75.57 %)		Exclusively the 1 st trimester (n = 132, 7.92 %)			Late pregnancy (n = 275, 16.51 %)		
	%	Mean (SD)	%	Mean (SD)	P Value ^a	%	Mean (SD)	P Value ^b
Maternal Characteristics								
Age		30.1 (4.6)		28.5 (5)	<0.001		28.5 (5.4)	<0.001
Height, cm		163.6 (6.1)		163.7 (6)	0.85		163.3 (6.4)	0.54
Weight, kg		62.5 (12.7)		60.5 (12.1)	0.08		61.6 (12.7)	0.33
BMI ^c , kg/m ²		22.2 (14.9 - 60.4)		21.4 (15.6 - 38)	0.02		21.8 (14 - 43)	0.28
Overweight or Obese	25.9		23.5		0.09	26.6		0.08
Center (Poitiers)	45.9		50.8		0.29	53.1		0.03
Education level								
High school or less	38.8		51.5		<0.001	69.8		<0.001
2-year university degree	23.8		20.4			15.6		
3-year university degree	37.4		28			14.6		
Primiparity	43.8		56.8		0.004	43.3		0.86
Number of cigarettes per day								
1 st Trimester				6.0 (1-35)			8.0 (1-40)	
2 nd Trimester							5.0 (1-20)	
3 rd Trimester							5.0 (1-25)	
Paternal Characteristics								
BMI, kg/m ²		24.8 (15.8 - 40.8)		24.5 (18 - 35.2)	0.12		24.3 (17.6 - 35.5)	0.07
Smoking status	29.7		62.1		<0.001	75.1		<0.001
Child Characteristics								
Male Sex	52.3		49.2		0.51	51.6		0.85
Gestational age, weeks ^c		40 (28 - 42)		40 (34 - 42)	0.92		39 (30 - 42)	0.36
Birth length, cm		49.71 (2.31)		49.43 (2.02)	0.19		49.03 (2.39)	<0.001
Birth weight, g		3311 (507)		3316 (455)	0.92		3157 (486)	<0.001
Breast feeding Yes/No	76.9		71.2		0.14	61.8		<0.001
Breast feeding ≥3 months	50.2		40.9		0.04	26.9		<0.001
Breast-feeding duration in breast-fed child, months		4 (0.2 - 13)		3 (0.2 - 13)	<0.001		2 (0.2 - 13)	<0.001
Prematurity	5.8		2.3		0.09	4.4		0.35

Abbreviations: BMI, body mass index; SD, standard deviation

^{a, b}Comparison between characteristics of mothers who smoked exclusively during the 1st trimester, their children and partners and those of non-smokers using the X², t tests or Man-Whitney-Wilcoxon tests as appropriate

^cMedian (min-max) presented due to left or right skew

217

218

219 Maternal smoking was inversely associated with the age at delivery, education level
220 and breast feeding initiation and duration. Smokers in late pregnancy delivered neonates with
221 smaller birth lengths and weights than non-smokers (Table 1).

222 **Growth**

223 The 'one-step' method assesses associations between maternal smoking and
224 length/height or weight trajectories as a whole and the effects on the four-specific Jenss-
225 Bayley model parameters for length/height and weight (Table 2).

226

227 **Table 2. Mean differences in Height and Weight Model Parameters in each Maternal smoking category and the reference category (No**
 228 **smoking) using the (One-step Method)^a in Children aged 0-5 years in the EDEN Study, 2003-2012.**

229

Outcomes	A		B		C		D		Global P-value
	Birth length Extrapolation of Birth weight	β_A CI 95%	Growth Velocity -Childhood-	β_B CI 95%	Spurt of Growth -First Months-	β_C CI 95%	Curvature Degree -First Months-	β_D CI 95%	
Height									<0.001
Exclusively the 1 st trimester	-0.003	[-0.012,0.005]	0.028	[-0.001,0.057]	-0.022	[-0.057,0.013]	0.074*	[0.010,0.137]	
Late pregnancy	-0.015*	[-0.021,-0.008]	0.022	[-0.001,0.046]	0.004	[-0.024,0.031]	0.038	[-0.011,0.087]	
Weight									0.03
Exclusively the 1 st trimester	0.008	[-0.025,0.041]	-0.004	[-0.050,0.042]	0.016	[-0.036,0.068]	-0.023	[-0.099,0.054]	
Late pregnancy	-0.037*	[-0.062,-0.012]	0.035	[-0.001,0.071]	0.008	[-0.032,0.048]	0.022	[-0.037,0.081]	

Abbreviations: BMI, body mass index, CI, confidence interval.

^aAdjusted for maternal education level, BMI and age at delivery; breast-feeding duration; recruitment center; child's gender.

* $P < 0.05$.

230

231

232 The 'two-step' method shows associations between maternal smoking and predicted
233 height, weight, and BMI and their corresponding velocities at specific ages (Table 3).
234

235 **Table 3. Mean Differences in Predicted Anthropometric Outcomes from Birth to 5 Years in each Maternal Smoking Category and the**
 236 **Reference Category (No Smoking)^a (Two-step method).**

237

Outcomes by maternal smoking status	Birth		3 months		6 months		1 year		2 years		4 years		5 years	
	β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI
Length/Height, cm														
Exclusively the 1 st trimester	-0.24	[-0.63,0.15]	0.16	[-0.22,0.53]	0.23	[-0.16,0.62]	0.18	[-0.24,0.60]	0.06	[-0.45,0.57]	0.20	[-0.46,0.87]	0.35	[-0.40,1.11]
Late pregnancy	-0.63*	[-0.92,-0.33]	-0.40*	[-0.68,-0.12]	-0.30*	[-0.59,0.00]	-0.23	[-0.55,0.08]	-0.19	[-0.57,0.19]	-0.06	[-0.56,0.44]	0.02	[-0.55,0.58]
Weight, g														
Exclusively the 1 st trimester	28	[-61,117]	15	[-98,129]	39	[-98,176]	75	[-95,245]	101	[-119,321]	125	[-206,455]	137	[-261,535]
Late pregnancy	-120*	[-187,-54]	-53	[-138,33]	-22	[-125,81]	33	[-95,161]	121	[-44,287]	249*	[1,498]	307*	[8,606]
BMI, kg/m ²														
Exclusively the 1 st trimester	0.25	[-0.01,0.52]	-0.05	[-0.27,0.17]	-0.04	[-0.27,0.19]	0.03	[-0.19,0.26]	0.10	[-0.12,0.31]	0.04	[-0.17,0.25]	0.00	[-0.22,0.22]
Late pregnancy	-0.16	[-0.37,0.04]	0.08	[-0.09,0.24]	0.11	[-0.07,0.28]	0.16	[-0.01,0.33]	0.23*	[0.07,0.39]	0.25*	[0.09,0.41]	0.24*	[0.07,0.41]
Height velocities, cm/month														
Exclusively the 1 st trimester	0.049*	[0.002,0.096]	0.007	[-0.023,0.036]	0.007	[-0.023,0.036]	-0.015	[-0.039,0.009]	-0.003	[-0.015,0.009]	0.011	[0.000,0.023]	0.013*	[0.001,0.025]
Late pregnancy	0.049*	[0.013,0.084]	0.022*	[0.000,0.044]	0.022*	[0.000,0.044]	0.005	[-0.013,0.022]	0.004	[-0.005,0.013]	0.006	[-0.002,0.015]	0.006	[-0.003,0.015]
Weight velocities, kg/month														
Exclusively the 1 st trimester	0.007	[-0.012,0.026]	0.007	[-0.012,0.026]	0.008	[-0.006,0.022]	0.004	[-0.005,0.013]	0.001	[-0.005,0.008]	0.001	[-0.006,0.008]	0.001	[-0.006,0.008]
Late pregnancy	0.011	[-0.003,0.026]	0.010	[-0.001,0.020]	0.010	[-0.001,0.020]	0.009*	[0.002,0.016]	0.006*	[0.001,0.011]	0.005	[0.0,0.01]	0.005	[0.000,0.010]
BMI velocities, kg/m ² /month														
Exclusively the 1 st trimester	-0.068*	[-0.135,-0.001]	0.011	[-0.012,0.035]	0.011	[-0.012,0.035]	0.011	[-0.001,0.023]	0.001	[-0.005,0.006]	-0.003	[-0.007,0.001]	-0.003	[-0.007,0.000]
Late pregnancy	0.023	[-0.028,0.073]	0.009	[-0.009,0.026]	0.009	[-0.009,0.026]	0.009	[0.000,0.017]	0.003	[-0.001,0.008]	0.000	[-0.003,0.003]	-0.001	[-0.003,0.002]
Age at BMI peak ^b , days														
Exclusively the 1 st trimester					7.31	[-11.93,26.55]								
Late pregnancy					4.97	[-9.46,19.39]								
BMI at BMI peak ^b , kg/m ²														
Exclusively the 1 st trimester					-0.04	[-0.27,0.19]								
Late pregnancy					0.12	[-0.05,0.29]								

Abbreviations: BMI, body mass index, CI, confidence interval.

^aAdjusted for maternal education level, BMI and age at delivery, breast-feeding duration, recruitment center, and child's gender.

^bAnalyses restricted to subjects with valid data for age and BMI at the BMI peak (identifiable BMI peak and BMI peak before 732 days). The number of children available for the analysis was 1,617.

* $P < 0.05$.

238

239

240 The overall fits of the models were good (Supporting Information S3, Figs S3 and S4).
241 Although the fit was good for weight because no trend was observed among the residuals
242 across time, the model fit the data somewhat less adequately for height. However, the
243 distribution of the BMI residuals over time and the individual fitted trajectories illustrated the
244 good ability of the indirect modeling to fit the BMI data (Supporting Information S3, Figs S3
245 and S5).

246 **Length/height growth**

247 Figure 3B1 illustrates the height growth curve by smoking categories for children with
248 given socio-demographic characteristics, and figure 3B2 shows the differences in the mean
249 height trajectories between children from non-smoking mothers and children from mothers
250 who smoked exclusively during the 1st trimester or until late pregnancy.

251 **Fig 3. A1) Weight (kg), B1) Height (cm), and C1) BMI (kg/m²): Mean Trajectories in**
252 **each category of Maternal Smoking for children from the EDEN Study aged from birth**
253 **to 5 years. A2) Weight (kg), B2) Height (cm), and C2) BMI (kg/m²): Differences in**
254 **Trajectories between each category of Maternal Smoking and the Reference Category**
255 **(No Smoking) for children from the EDEN Study aged from birth to 5 years. Mean**
256 **Trajectories are illustrated for Boys breastfed less than 3 months with 30 year-old non-**
257 **overweight mothers with a high-school education or less born in Poitiers.**

258 The child length/height trajectories differed among children exposed to maternal
259 smoking during pregnancy and children from non-smokers ($p < 0.001$).

260 Compared with no maternal smoking, smoking during the 1st trimester exclusively was
261 associated with lower C and significantly higher D parameters. This finding illustrates a faster
262 early length growth for a short period of time (up to 3 months). These children also exhibited
263 faster height growth velocities at 5 years compared to the children from non-smokers,
264 although their height was not significantly greater at these ages (Table 3).

265 Compared with no maternal smoking, maternal smoking during late pregnancy was
266 associated with a 0.015 lower A parameter (95% CI: -0.021, -0.008). This difference
267 represented a 0.63 cm shorter length at birth in children whose mothers smoked during late
268 pregnancy compared to children born to non-smoker mothers.

269 Compared to children from non-smokers, children whose mothers smoked during late
270 pregnancy had a lower average length from birth to 6 months. They exhibited a significantly
271 faster length growth velocity immediately after birth (higher D parameter compared to
272 children from non-smokers) up to 6 months (Table 3). These children seemed to catch up to
273 the heights of children from non-smokers only at the end of the study period (Table 3).

274 **Weight growth**

275 Similar to height, Figure 3A1 illustrates the weight growth curve by smoking
276 categories, and figure 3A2 shows the differences in the mean weight trajectories between
277 children from non-smoking mothers and children from mothers who smoked exclusively
278 during the 1st trimester or until late pregnancy.

279 Maternal smoking during pregnancy was significantly associated with a change in the
280 global weight trajectory ($p=0.03$).

281 The average weight trajectory of children whose mothers smoked during the 1st
282 trimester exclusively was similar to that of children from non-smokers (Table 2).

283 However, children born to mothers who smoked during late pregnancy had weight
284 trajectories that differed from the children of non-smokers. These children had birthweights
285 that were 120 grams lower and tended to have a higher D parameter that was in line with
286 faster growth in the beginning of life. Using the two-step approach, we found that their weight
287 growth was significantly higher at one and two years (0.009 and 0.006 kg/month, 95% CI:
288 0.002, 0.016 and 0.001, 0.011, respectively) and higher at 4 and 5 years without reaching
289 statistical significance. The higher B parameter in this subgroup was in line with this higher

290 weight growth during childhood. Thus, they surpassed the weight of unexposed children at
291 approximately 2 years and continued to gain more weight thereafter. At ages 4 and 5 years,
292 they had a 249 gram (95% CI: 1, 498) and 307 gram (95% CI: 8, 606) higher mean weight,
293 respectively, compared to children born to non-smokers.

294 **BMI growth**

295 Figure 3C1 illustrates the resulting BMI growth curve by smoking categories, and
296 figure 3C2 shows the differences in the mean BMI trajectories between children from non-
297 smoking mothers and children from mothers who smoked exclusively during the 1st trimester
298 or until late pregnancy.

299 Children born to non-smokers and to mothers who smoked during the 1st trimester
300 exclusively had quite similar mean BMI trajectories. Smoking during the 1st trimester
301 exclusively was not associated with any age-predicted BMI. The mean BMI trajectory of
302 children from smokers in late pregnancy surpassed the trajectory of the unexposed children in
303 the first couple of months. Although their ages at the BMI infancy peak were comparable,
304 maternal smoking in late pregnancy was associated with a non-significant BMI increase at the
305 BMI infancy peak (0.12, 95% CI: -0.05, 0.29). At 2 years of age, these children had a
306 significantly higher BMI (0.23 kg/m²). This difference reached 0.24 kg/m² at 5 years.

307 **Dose response**

308 A dose-response relationship between maternal smoking in late pregnancy and growth
309 characteristics (Supporting Information S4, Tables S4; S5) was observed. The height and
310 weight A parameters were significantly negatively associated and height D was positively
311 associated with the number of cigarettes smoked per day. A five cigarette per day increase was
312 associated with a 0.39 cm (95% CI: -0.58, -0.20) and 0.21 cm (95% CI: -0.40, -0.02) lower
313 length at birth and at 6 months, respectively, a 82 gram lower measured birthweight, and a
314 155 gram (95% CI: -1, 312) and 188 gram (95% CI: 0, 376) higher predicted weight at 4 and

315 5 years, respectively. These results were in line with a 0.018 higher B parameter (95% CI: -
316 0.005, 0.042).

317 **Sensitivity analysis (Supporting Information S5)**

318 The exclusion of the 88 preterm children (Table S6) and the restriction of the analysis
319 to 1,211 children with data beyond 4 years of age did not change the results. The weights,
320 heights and BMIs of children exposed only postnatally were comparable to these
321 measurements from the non-exposed children (Table S7). Associations between the predicted
322 BMIs at different time points and maternal smoking status gave close results regardless of the
323 modeling method used (Table S8). For the subsample of children attending the clinical visits
324 at 1, 3 and 5 years, the associations with maternal smoking status were comparable for the
325 measurements from clinical visits or the model-predicted data (Table S9).

326 **Discussion**

327 Our approach provides mutually complementary methods to accurately characterize
328 associations between maternal smoking and growth. We showed that although the BMI of
329 children exposed during late pregnancy became significantly higher from 2 years of age
330 onwards compared to unexposed children, their BMI trajectories started to diverge very early
331 in infancy. This finding was explained by a predominant catch-up growth in weight compared
332 to length and reinforced by a faster weight gain from 1 year onwards.

333 The ‘one-step’ method allows the examination of growth trajectories longitudinally
334 and provides a global test for the relationship between the exposure and weight/height
335 trajectories overall and for period-specific characteristics of growth using the Jenss-Bayley
336 model parameters. The ‘two-step’ method provides statistical tests at specific ages (especially
337 for BMI), thereby providing information on the magnitude of the between-subgroups
338 differences.

339 The results regarding the associations between maternal smoking and BMI across
340 childhood were very close using our indirect method or the fractional polynomials and splines
341 However, the monotonic pattern of weight and height (in contrast to BMI growth) can be
342 beneficial using our indirect BMI modeling. Our indirect modeling approach also provides
343 additional assets. First, the height and weight model parameters have biological
344 interpretability. Second, the model is less sensitive to outliers compared to non-structural
345 models [13]. Third, our method does not need to collect weight and height measurements at
346 the same time (in comparison with direct modeling), which increases the number of available
347 measurements for modeling. Compared to the study of associations using observed data, our
348 approach takes into account missing data and limits potential selection bias (Supplementary
349 Information S5, Table S9). Furthermore, in contrast to model-predicted data, analyses using
350 the observed data did not allow the assessment of the association at many time points and
351 consequently the accurate characterization of the evolution of differences between the
352 smoking subgroups over time. Finally, the observed data did not allow the assessment of
353 associations with key growth milestones, such as instantaneous growth velocities and age and
354 BMI at the BMI infancy peak.

355 In terms of weaknesses, the indirect BMI modeling did not allow the assessment of
356 the global association between exposure and BMI trajectories, which can be performed for the
357 height and weight trajectories separately; however, there can be some potential convergence
358 issues when the covariate numbers increase. Additionally, the smoking status at the beginning
359 of pregnancy was taken as a proxy measurement for the 1st trimester status. A high percentage
360 of women quit very early in pregnancy, which could explain the lack of association with birth
361 outcomes in children exposed exclusively during the 1st trimester. Furthermore, assessing the
362 smoking status using a self-reported questionnaire could lead to misclassification bias. We do

363 not expect that the estimated differences in anthropometric measurements between exposed
364 and non-exposed children would be overestimated by underreporting.

365 In recent reviews of early determinants of overweight or obesity, maternal smoking
366 during pregnancy was consistently associated with adult obesity [1] and child overweight and
367 obesity [14,15]. These reviews were in favor of a long-lasting effect of maternal smoking on
368 health. Our study adds to the growing body of evidence suggesting the emergence of these
369 associations in infancy. The few studies investigating early growth explained the increased
370 BMI by a weight deficit at birth followed by a rapid weight catch-up growth prior to 1 year
371 that lead to a similar or higher weight compared to non-exposed children later in childhood
372 [16–18]. Although exposed children tend to catch up in weight, a persistently shorter height in
373 late childhood was observed [17–19]. Our results on the growth velocities and C and D
374 parameters showed a greater impact of maternal smoking on length than weight growth during
375 the first 6 months of life. Because weight depends on length growth, this discrepancy is in
376 favor of a differential effect on the different compartments of body weight. The BMI
377 trajectory in children exposed during late pregnancy suggests a faster fat mass accretion in the
378 first months of life with a higher BMI at the BMI infancy peak that is maintained throughout
379 childhood.

380 In our population, a smaller weight and length at birth were associated with maternal
381 smoking during late pregnancy but not with smoking during the 1st trimester exclusively.
382 These results corroborate the association between maternal smoking and impaired fetal
383 growth reported in previous works [20,21]. Furthermore, maternal smoking exclusively
384 during the 1st trimester was not associated with any age-predicted BMI, supporting a study
385 that found no association between maternal smoking in the 1st trimester and the BMI of
386 children up to 4 years [17] but contrasting with another study where a higher risk of being
387 overweight was found in children aged 5 to 7 years [22].

388 A direct effect of smoking during pregnancy has been postulated to occur through
389 increased maternal and fetal carboxyhemoglobin blood levels, which are responsible for fetal
390 hypoxia and growth retardation [23]. Subsequent associations with postnatal growth might be
391 explained by a deregulation of the hypothalamic centers that regulate food intake, growth and
392 adipocyte numbers [24]. Tobacco might also have a potentiating effect on hypothalamic
393 structures, resulting in impaired insulin signaling and metabolism [24].

394 The persistent positive association between maternal smoking and overweight children
395 in within-sibling analyses [25] and the higher effect estimates for maternal smoking compared
396 with paternal smoking [26] support intrauterine effects, whereas the attenuated associations
397 between maternal smoking and obesity in young adults in within-family analyses [27] backed
398 by studies with similar [19,28–30] or stronger [31] associations between paternal and
399 maternal smoking statuses and the offspring's overweight BMI suggest residual confounding
400 by unmeasured shared familial factors. We did not control for lifestyle characteristics.
401 Although an effect of shared familial factors cannot be excluded, we propose that this effect is
402 less likely because the differences appeared very early in life. The fast increase in the BMI
403 during the first weeks of life could suggest that cessation of exposure to tobacco products
404 (e.g., nicotine) immediately after birth may contribute to the observed association as
405 previously speculated [32].

406 Our findings are consistent with an association between smoking in late pregnancy and
407 upward BMI trajectories as early as the first months of life, resulting in a difference in the
408 BMI that is maintained throughout the infancy and childhood periods. This result supports the
409 concept that infant programming related to fetal exposure to adverse conditions has lasting or
410 lifelong effects. The second and third trimesters of pregnancy could be more critical periods
411 for the induction of lifelong susceptibility to obesity. Our approach based on height and

412 weight modeling offers a convenient method to estimate key trajectory milestones and
413 associations with early-life determinants.
414

415 **Acknowledgments**

416 The EDEN (Etude des Déterminants pré et postnatals précoces du développement et de la
417 santé de l'Enfant) study was provided by the Foundation for Medical Research, the French
418 Ministry of Research (IFR program), the Institut National de la Santé et de la Recherche
419 Médicale (Human Nutrition Research Program), the Diabetes National Research Program (via
420 a collaboration with the French Association for Diabetes Research), the French Ministry of
421 Health Perinatal Program, the French Agency for Environmental Security, the French
422 National Institute for Population Health Surveillance, the Paris-Sud University, the French
423 National Institute for Health Education, Nestlé, the National Education Health Insurance, the
424 French Speaking Association for the Study of Diabetes and Metabolism, the National Agency
425 for Research (nonthematic program), and the National Institute for Research in Public Health
426 (TGIR health cohort 2008 program).

427 We thank Lise Giorgis-Allemand and Anne Forhan for data management, the midwife
428 research assistants (Lorraine Douhaud, Sophie Bedel, Brigitte Lortholary, Sophie Gabriel,
429 Muriel Rogeon, and Monique Malinbaum) for data collection, and the data entry operators
430 (Patricia Lavoine, Josiane Sahuquillo, and Ginette Debotte).

431 The EDEN mother-child cohort study group includes I. Annesi-Maesano, JY. Bernard, J.
432 Botton, M.A. Charles, P. Dargent-Molina, B. de Lauzon-Guillain, P. Ducimetière, M. de
433 Agostini, B. Foliguet, A. Forhan, X. Fritel, A. Germa, V. Goua, R. Hankard, B. Heude, M.
434 Kaminski, B. Larroque†, N. Lelong, J. Lepeule, G. Magnin, L. Marchand, C. Nabet, F. Pierre,
435 R. Slama, M.J. Saurel-Cubizolles, M. Schweitzer, O. Thiebaugeorges.

436 All of the authors have read, critiqued and approved the contents of this article.

437

438 **References**

- 439 1. Brisbois TD, Farmer AP, McCargar LJ. Early markers of adult obesity: a review. *Obesity*
440 *Reviews*. 2012;13:347–67.
- 441 2. Ziyab AH, Karmaus W, Kurukulaaratchy RJ, Zhang H, Arshad SH. Developmental
442 trajectories of Body Mass Index from infancy to 18 years of age: prenatal determinants and
443 health consequences. *Journal of epidemiology and community health*. 2014;68:934–41.
- 444 3. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. *Annual*
445 *Review of Clinical Psychology*. 2010;6:109–38.
- 446 4. Bauer DJ. Observations on the Use of Growth Mixture Models in Psychological Research.
447 *Multivariate Behavioral Research*. 2007;42:757–86.
- 448 5. Wen X, Kleinman K, Gillman MW, Rifas-Shiman SL, Taveras EM. Childhood body mass
449 index trajectories: modeling, characterizing, pairwise correlations and socio-demographic
450 predictors of trajectory characteristics. *BMC Medical Research Methodology*. 2012;12:38.
- 451 6. Royston P, Altman DG. Regression Using Fractional Polynomials of Continuous
452 Covariates: Parsimonious Parametric Modelling. *Journal of the Royal Statistical Society*.
453 *Series C (Applied Statistics)*. 1994;43:429–67.
- 454 7. Silverwood RJ, De Stavola BL, Cole TJ, Leon DA. BMI peak in infancy as a predictor for
455 later BMI in the Uppsala Family Study. *International Journal of Obesity*. 2009;33:929–37.
- 456 8. Howe LD, Tilling K, Benfield L, Logue J, Sattar N, Ness AR, et al. Changes in ponderal
457 index and body mass index across childhood and their associations with fat mass and
458 cardiovascular risk factors at age 15. *PLoS One*. 2010;5:e15186.

- 459 9. Botton J, Scherdel P, Regnault N, Heude B, Charles M-A, EDEN Mother-Child Cohort
460 Study Group. Postnatal weight and height growth modeling and prediction of body mass
461 index as a function of time for the study of growth determinants. *Annals of Nutrition &*
462 *Metabolism*. 2014;65:156–66.
- 463 10. Heude B, Forhan A, Slama R, Douhaud L, Bedel S, Saurel-Cubizolles M-J, et al. Cohort
464 Profile: The EDEN mother-child cohort on the prenatal and early postnatal determinants of
465 child health and development. *International Journal of Epidemiology*. 2015;
- 466 11. Bonet M, Marchand L, Kaminski M, Fohran A, Betoko A, Charles M-A, et al.
467 Breastfeeding duration, social and occupational characteristics of mothers in the French
468 “EDEN mother-child” cohort. *Maternal and Child Health Journal*. 2013;17:714–22.
- 469 12. Comets E, Lavenu A, Lavielle M. SAEMIX, an R version of the SAEM algorithm for
470 parameter estimation in nonlinear mixed effect models [Internet]. [cited 2015 Jan 7].
471 Available from: http://papersjds14.sfds.asso.fr/submission_54.pdf
- 472 13. Chirwa ED, Griffiths PL, Maleta K, Norris SA, Cameron N. Multi-level modelling of
473 longitudinal child growth data from the Birth-to-Twenty Cohort: a comparison of growth
474 models. *Annals of Human Biology*. 2014;41:168–79.
- 475 14. Ino T. Maternal smoking during pregnancy and offspring obesity: meta-analysis.
476 *Pediatrics International: Official Journal of the Japan Pediatric Society*. 2010;52:94–9.
- 477 15. Oken E, Levitan EB, Gillman MW. Maternal smoking during pregnancy and child
478 overweight: systematic review and meta-analysis. *International Journal of Obesity*.
479 2008;32:201–10.
- 480 16. Timmermans SH, Mommers M, Gubbels JS, Kremers SPJ, Stafleu A, Stehouwer CDA, et

- 481 al. Maternal smoking during pregnancy and childhood overweight and fat distribution: the
482 KOALA Birth Cohort Study. *Pediatric Obesity*. 2014;9:e14–25.
- 483 17. Durmus B, Kruithof CJ, Gillman MH, Willemsen SP, Hofman A, Raat H, et al. Parental
484 smoking during pregnancy, early growth, and risk of obesity in preschool children: the
485 Generation R Study. *The American Journal of Clinical Nutrition*. 2011;94:164–71.
- 486 18. Chen A, Pennell ML, Klebanoff MA, Rogan WJ, Longnecker MP. Maternal smoking
487 during pregnancy in relation to child overweight: follow-up to age 8 years. *International*
488 *Journal of Epidemiology*. 2006;35:121–30.
- 489 19. Howe LD, Matijasevich A, Tilling K, Brion M-J, Leary SD, Smith GD, et al. Maternal
490 smoking during pregnancy and offspring trajectories of height and adiposity: comparing
491 maternal and paternal associations. *International Journal of Epidemiology*. 2012;41:722–32.
- 492 20. Iñiguez C, Ballester F, Costa O, Murcia M, Souto A, Santa-Marina L, et al. Maternal
493 smoking during pregnancy and fetal biometry: the INMA Mother and Child Cohort Study.
494 *American Journal of Epidemiology*. 2013;178:1067–75.
- 495 21. Jaddoe VWV, Verburg BO, de Ridder M a. J, Hofman A, Mackenbach JP, Moll HA, et al.
496 Maternal smoking and fetal growth characteristics in different periods of pregnancy: the
497 generation R study. *American Journal of Epidemiology*. 2007;165:1207–15.
- 498 22. Mendez MA, Torrent M, Ferrer C, Ribas-Fitó N, Sunyer J. Maternal smoking very early in
499 pregnancy is related to child overweight at age 5-7 y. *The American Journal of Clinical*
500 *Nutrition*. 2008;87:1906–13.
- 501 23. DiFranza JR, Aligne CA, Weitzman M. Prenatal and postnatal environmental tobacco
502 smoke exposure and children’s health. *Pediatrics*. 2004;113:1007–15.

- 503 24. Toschke AM, Montgomery SM, Pfeiffer U, von Kries R. Early intrauterine exposure to
504 tobacco-inhaled products and obesity. *American Journal of Epidemiology*. 2003;158:1068–74.
- 505 25. Gilman SE, Gardener H, Buka SL. Maternal smoking during pregnancy and children’s
506 cognitive and physical development: a causal risk factor? *American Journal of Epidemiology*.
507 2008;168:522–31.
- 508 26. Riedel C, Schönberger K, Yang S, Koshy G, Chen Y-C, Gopinath B, et al. Parental
509 smoking and childhood obesity: higher effect estimates for maternal smoking in pregnancy
510 compared with paternal smoking--a meta-analysis. *International Journal of Epidemiology*.
511 2014;43:1593–606.
- 512 27. Iliadou AN, Koupil I, Villamor E, Altman D, Hultman C, Långström N, et al. Familial
513 factors confound the association between maternal smoking during pregnancy and young
514 adult offspring overweight. *International Journal of Epidemiology*. 2010;39:1193–202.
- 515 28. von Kries R, Bolte G, Baghi L, Toschke AM, GME Study Group. Parental smoking and
516 childhood obesity--is maternal smoking in pregnancy the critical exposure? *International*
517 *Journal of Epidemiology*. 2008;37:210–6.
- 518 29. Kwok MK, Schooling CM, Lam TH, Leung GM. Paternal smoking and childhood
519 overweight: evidence from the Hong Kong “Children of 1997.” *Pediatrics*. 2010;126:e46–56.
- 520 30. Leary SD, Smith GD, Rogers IS, Reilly JJ, Wells JCK, Ness AR. Smoking during
521 pregnancy and offspring fat and lean mass in childhood. *Obesity (Silver Spring)*.
522 2006;14:2284–93.
- 523 31. Florath I, Kohler M, Weck MN, Brandt S, Rothenbacher D, Schöttker B, et al. Association
524 of pre- and post-natal parental smoking with offspring body mass index: an 8-year follow-up

525 of a birth cohort. *Pediatric Obesity*. 2014;9:121–34.

526 32. Chen H, Morris MJ. Maternal smoking-A contributor to the obesity epidemic? *Obesity*

527 *Research & Clinical Practice*. 2007;1:I – II.

528

Supporting Information

S1 Table. Number of Measurements available for Modeling the Height and Weight growth of children (0-5 years) from the EDEN study, France, 2003-2012

S2 File. Indirect Body Mass Index Modeling

S3 File. Model Adequacies

S4 File. Dose Effect Analysis

S5 File. Sensitivity Analyses

Supporting Information S1 Number of Height and Weight Measures

Table S1 Number of Measurements available for Modeling Height and Weight growth of Children (0-5 years) from the EDEN study, France, 2003-2012

Age range	No. of Subjects ^a			No. of Measures			No. of Measures per Subject ^b		
	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total
Weight ^c									
<6 months	864	801	1,665	5,198	4,864	10,062	6(1-9)	6(1-9)	6(1-9)
6 months - 1 year	814	754	1,568	3,263	2,987	6,250	4(1-9)	4(1-8)	4(1-9)
1 - 2 years	788	721	1,509	2,838	2,686	5,524	4(1-9)	4(1-9)	4(1-9)
2 - 4 years	750	694	1,444	3,055	2,820	5,875	4(1-9)	4(1-10)	4(1-10)
>= 4 years	635	573	1,208	1,496	1,308	2,804	2(1-5)	2(1-4)	2(1-5)
Whole Study Period	865	801	1,666	15,850	14,665	30,515	20(2-30)	19(2-32)	19(2-32)
Height									
<6 months	863	801	1,664	4,765	4,449	9,214	6(1-9)	6(1-8)	6(1-9)
6 months - 1 year	808	753	1,561	3,045	2,756	5,801	4(1-9)	4(1-8)	4(1-9)
1 - 2 years	784	715	1,499	2,697	2,555	5,252	3(1-9)	4(1-9)	4(1-9)
2 - 4 years	748	691	1,439	2,819	2,576	5,395	4(1-9)	4(1-10)	4(1-10)
>= 4 years	633	570	1,203	1,459	1,260	2,719	2(1-5)	2(1-4)	2(1-5)
Whole Study Period	865	801	1,666	14,785	13,596	28,381	18(2-30)	18(2-31)	18(2-31)

^aSubjects with More Than One Measure

^bMedian (min-max)

^cExcluding Birth Weight and including the Minimal Weight Measure in the First Four Days

Supporting Information S2 Indirect BMI modeling

Expected height (cm) or weight (kg), $Y_{i,j}$ at measurement time $t_{i,j}$ (days) for the child i were expressed as follows:

$$\begin{pmatrix} A_i \\ B_i \\ C_i \\ D_i \end{pmatrix} = \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} + \begin{pmatrix} u_{A_i} \\ u_{B_i} \\ u_{C_i} \\ u_{D_i} \end{pmatrix}$$

$$\boldsymbol{\phi}_i = \boldsymbol{\phi} + \mathbf{u}_i$$

Equation 1.

$$Y_{i,j} = \exp(A_i) + \exp(B_i) * t_{i,j} + \exp(C_i) * (1 - \exp(-\exp(D_i) * t_{i,j})) + e_{i,j}$$

The vector of fixed effects $\boldsymbol{\phi}$, is the average value of the individual parameters $\boldsymbol{\phi}_i$ of the children included.

The vector of random effects \mathbf{u}_i represents the deviation of $\boldsymbol{\phi}_i$ from their population average ($\boldsymbol{\phi}$) corresponding to between-individual variability, and $e_{i,j}$ are the residuals corresponding to within-individual variability. A combined error model was used to describe the residual error of height and weight. The general form of $Y_{i,j}$ can be expressed as

$$Y_{i,j} = f(\mathbf{x}_{i,j}, \boldsymbol{\phi}_i) + g(\mathbf{x}_{i,j}, \boldsymbol{\phi}_i, \boldsymbol{\xi})\varepsilon_{i,j}$$

where $\mathbf{x}_{i,j}$ are the regression variables or design variables and $\varepsilon_{i,j}$ are the residual errors. The residual error model is defined by the function g and the parameters $\boldsymbol{\xi}$. The combined error model assumes that $g = a + bf$ and $\boldsymbol{\xi} = (a, b)$.

Random effects have to be normally distributed with mean $\mathbf{0}$ and a symmetric variance-covariance matrix denoted $\boldsymbol{\Omega}_u$, as are the residuals.

$$\mathbf{u}_i \sim N(\mathbf{0}, \boldsymbol{\Omega}_u)$$

For a better interpretation of the parameters, Figure S1 illustrates the influence of each of them on the global growth shape. Weight trajectories 1--4 were fitted, respectively, for each decile of A_i , B_i , C_i , and D_i (Table S2) while fixing the other 3 parameters at the value of the

population average. For illustration purposes, independency of parameters was assumed which is more straightforward although not realistic. In reality, the non-null covariance between parameters can result in a larger panel of shapes.

Table S2 Distributions of Individual Weight Growth Model among Boys and Girls (0-5 years) of the EDEN study, France, 2003-2012

Parameter	Deciles									Mean
	1 st	2 nd	3 rd	4 th	5 th	6 th	7 th	8 th	9 th	
A_{wi} Birth length Extrapolation of birth weight	0.89	0.98	1.02	1.06	1.10	1.13	1.17	1.21	1.26	1.08
B_{wi} Growth Velocity -Childhood-	-5.41	-5.32	-5.26	-5.21	-5.17	-5.12	-5.07	-5.00	-4.92	-5.16
C_{wi} Spurt of Growth -First Months-	1.27	1.38	1.46	1.53	1.59	1.65	1.71	1.79	1.91	1.59
D_{wi} Curvature Degree -First Months-	-5.41	-5.25	-5.12	-5.04	-4.96	-4.87	-4.80	-4.70	-4.54	-4.97

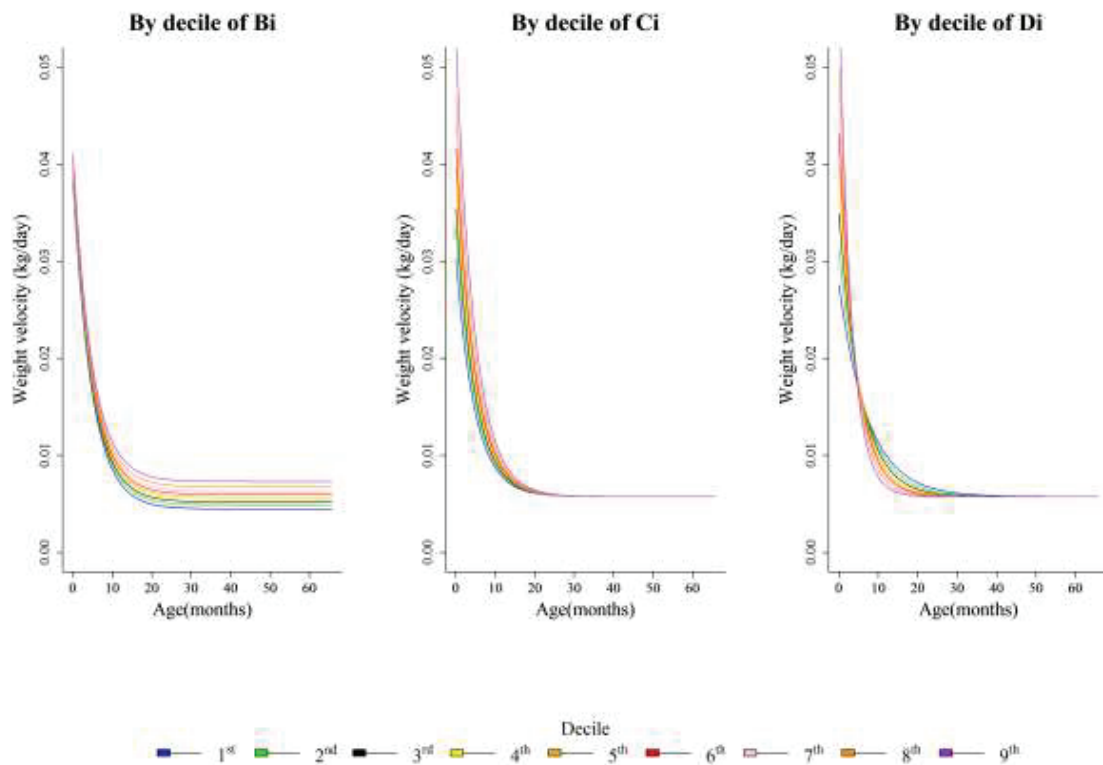


Figure S1 Weight Growth Velocities from birth to 5 years by decile of weight growth parameters (A_i , B_i , C_i and D_i) in Children of the EDEN Study, France, 2003-2012

The A component describes the length at birth or the approximation of the minimal weight in the first days of life. The B component describes the asymptote for the older ages or growth velocity beyond two years. The C component describes the spurt of growth in the first months of life, the difference between birth measure and the value of the later ages asymptote in the Y-axis. The D component describes the curvature in the first period of growth.

The distributions of individual height and weight growth parameters obtained through Jenss modeling in our sample are detailed in the Table S3.

Table S3 Distributions of Individual height and weight growth model parameters in children (0-5 years) of the EDEN study, France, 2003-2012

Model Parameters Mean (SD)	Gender	
	Boys	Girls
Height		
A _H	3.92 (0.04)	3.90 (0.04)
B _H	-3.89 (0.10)	-3.90 (0.12)
C _H	3.14 (0.13)	3.11 (0.15)
D _H	-5.37 (0.26)	-5.45 (0.28)
Weight		
A _W	1.10 (0.17)	1.06 (0.18)
B _W	-5.16 (0.20)	-5.16 (0.21)
C _W	1.64 (0.23)	1.52 (0.27)
D _W	-4.94 (0.34)	-5.00 (0.35)

Abbreviation: SD, standard deviation

Body Mass Index trajectories were deduced using height and weight equations.

Equation 3.

$$Weight_{i,j} = \exp(A_{W_i}) + \exp(B_{W_i}) * t_{i,j} + \exp(C_{W_i}) * (1 - \exp(-\exp(D_{W_i}) * t_{i,j})) + e_{i,j}$$

Equation 4.

$$Height_{i,j} = \exp(A_{H_i}) + \exp(B_{H_i}) * t_{i,j} + \exp(C_{H_i}) * (1 - \exp(-\exp(D_{H_i}) * t_{i,j})) + e_{i,j}$$

$$\begin{pmatrix} A_{H_i} \\ B_{H_i} \\ C_{H_i} \\ D_{H_i} \end{pmatrix} = \begin{pmatrix} A_H \\ B_H \\ C_H \\ D_H \end{pmatrix} + \begin{pmatrix} u_{A_{H_i}} \\ u_{B_{H_i}} \\ u_{C_{H_i}} \\ u_{D_{H_i}} \end{pmatrix} \text{ and } \begin{pmatrix} A_{W_i} \\ B_{W_i} \\ C_{W_i} \\ D_{W_i} \end{pmatrix} = \begin{pmatrix} A_W \\ B_W \\ C_W \\ D_W \end{pmatrix} + \begin{pmatrix} u_{A_{W_i}} \\ u_{B_{W_i}} \\ u_{C_{W_i}} \\ u_{D_{W_i}} \end{pmatrix}$$

With $\{A_{H_i}, B_{H_i}, C_{H_i}, D_{H_i}\}$ and $\{A_{W_i}, B_{W_i}, C_{W_i}, D_{W_i}\}$ the individual height and weight parameters respectively. $\{A_H, B_H, C_H, D_H\}$ and $\{A_W, B_W, C_W, D_W\}$ the height and weight population parameters.

Thus, individual BMI trajectories were fitted for each child i according to equation 3 where weight is expressed in kg and height in cm:

Equation 3.

$$BMI_{i,j} = \frac{Weight_{i,j}}{(Height_{i,j}/100)^2} = \frac{\exp(A_{W_i}) + \exp(B_{W_i}) t_{i,j} + \exp(C_{W_i}) (1 - \exp(-\exp(D_{W_i}) t_{i,j}))}{[(\exp(A_{H_i}) + \exp(B_{H_i}) t_{i,j} + \exp(C_{H_i}) (1 - \exp(-\exp(D_{H_i}) t_{i,j}))) / 100]^2}$$

Supporting Information S3 Models Adequacies

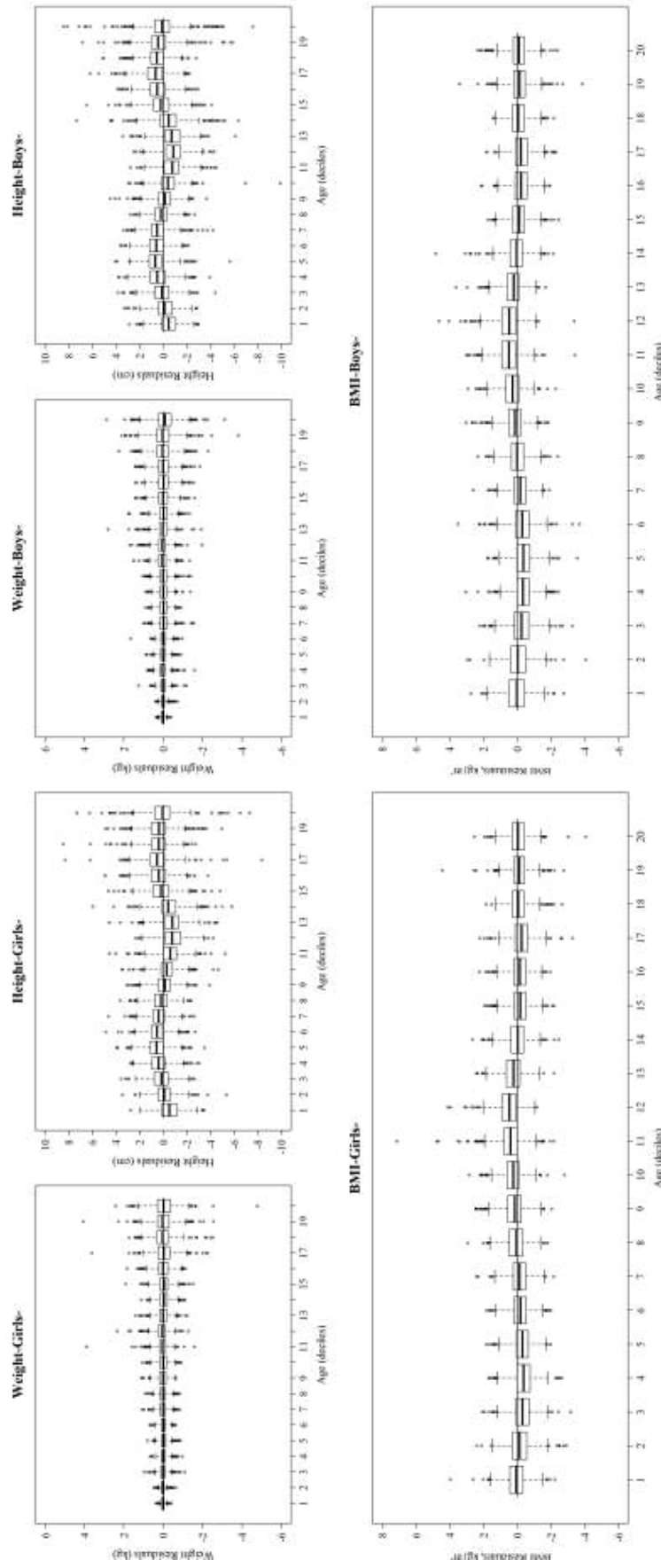


Figure S3 Distribution over time of residuals of height and weight model and resulting difference between observed and calculated values of BMI by gender in the EDEN study (0-5 years)

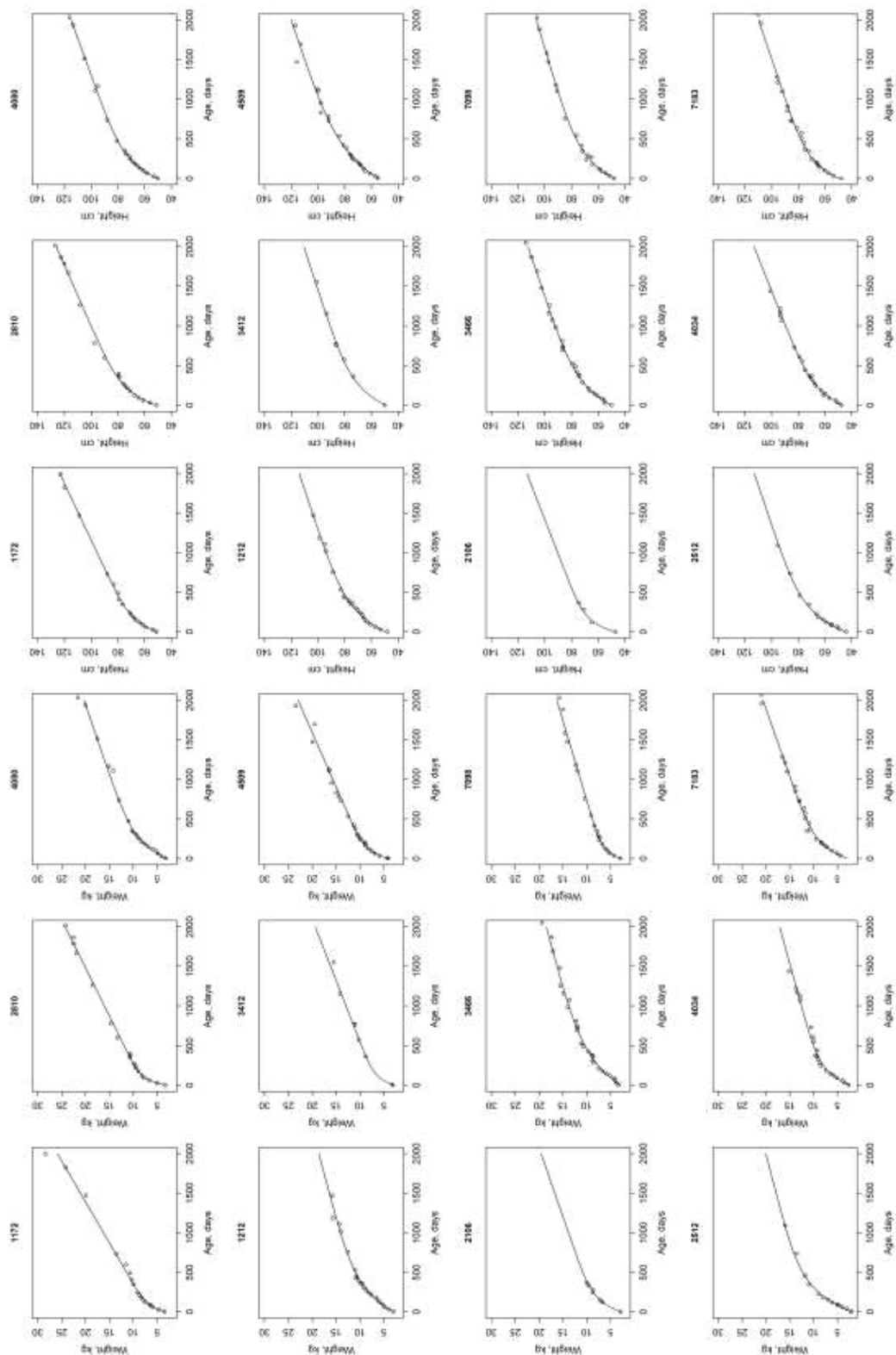


Figure S4 Height/Weight measures (circles) and Fitted Trajectories (black lines) of Height (cm)/Weight (kg) for 12 randomly selected individuals from the EDEN study (0-5 years)

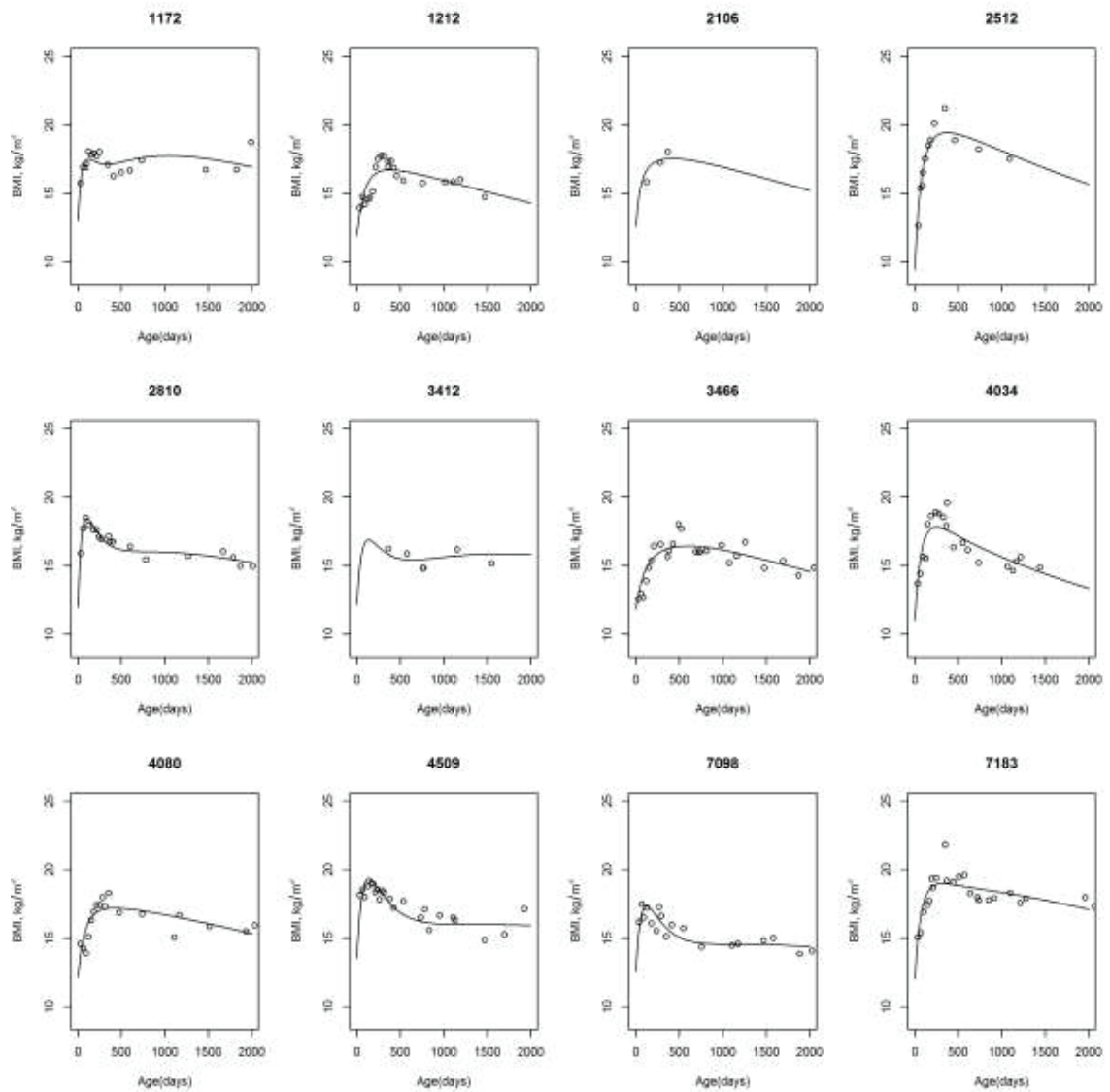


Figure S5 BMI measures (circles) and Fitted Trajectories (black lines) of BMI (kg/m^2) for 12 randomly selected individuals from the EDEN study (0-5 years)

Supporting Information S4 Dose Effect Analysis

Table S4 Mean differences in Height and Weight Model Parameters for a 5 cigarettes per day increase in children whose mothers smoked throughout pregnancy, the EDEN study, France, 2003-2012

Outcomes	A		B		C		D	
	Extrapolation of Birth weight	Birth length	Growth Velocity -Childhood-	Growth Velocity	Spurt of Growth -First Months-	Spurt of Growth	Curvature Degree -First Months-	Curvature Degree
	β_A	CI 95%	β_B	CI 95%	β_C	CI 95%	β_D	CI 95%
Height	-0.009*	[-0.013,-0.005]	0.009	[-0.006,0.025]	-0.001	[-0.019,0.017]	0.027	[-0.005,0.059]
Weight	-0.026*	[-0.042,-0.011]	0.018	[-0.005,0.042]	0.008	[-0.017,0.034]	0.014	[-0.025,0.052]

Abbreviations: BMI, body mass index, CI, confidence interval.

^aAdjusted for maternal education level, BMI and age at delivery; breast-feeding duration; recruitment center; child's gender.

*P<0.05.

Table S5 Mean differences in Predicted Anthropometric Outcomes from Birth to 5 Years for a 5 cigarettes/day increase of the consumption of tobacco during the second and third trimesters of pregnancy^a (Two-steps method)

Outcomes	Birth		3 months		6 months		1 year		2 years		4 years		5 years	
	β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI
Length/Height, cm	-0.39*	[-0.58,-0.20]	-0.26*	[-0.44,-0.08]	-0.21*	[-0.40,-0.02]	-0.19	[-0.39,0.01]	-0.21	[-0.46,0.03]	-0.19	[-0.51,0.13]	-0.16	[-0.52,0.19]
Weight, g	-82*	[-125,-39]	-28	[-83,27]	-2	[-67,64]	37	[-45,118]	86	[-19,191]	155	[-1,312]	188	[0,376]
BMI, kg/m ²	-0.12	[-0.25,0.01]	0.06	[-0.04,0.17]	0.10	[-0.01,0.21]	0.15*	[0.04,0.26]	0.19*	[0.09,0.30]	0.20*	[0.10,0.30]	0.19*	[0.09,0.30]
Height velocities, cm/month			0.028*	[0.005,0.050]	0.010	[-0.004,0.024]	-0.001	[-0.013,0.010]	-0.001	[-0.007,0.005]	0.002	[-0.004,0.007]	0.002	[-0.003,0.008]
Weight velocities, kg/month			0.010*	[0.001,0.019]	0.008*	[0.001,0.015]	0.005*	[0.001,0.010]	0.003*	[0.000,0.006]	0.003	[-0.001,0.006]	0.003	[-0.001,0.006]
BMI velocities, kg/m ² /month			0.024	[-0.008,0.057]	0.010	[-0.001,0.021]	0.006*	[0.001,0.012]	0.002	[-0.001,0.005]	0.000	[-0.002,0.002]	-0.001	[-0.002,0.001]
Age at BMI peak ^b , days					6.79	[-2.57,16.14]								
BMI at BMI peak ^b , kg/m ²					0.12*	[0.00,0.23]								

Abbreviations: BMI, body mass index, CI, confidence interval.

^aAdjusted for maternal education level, BMI and age at delivery; breast-feeding duration; recruitment center; child's gender.

^bAnalyses restricted to those with valid data for age and BMI at BMI peak (identifiable BMI peak and BMI peak before 732 days). The number of children available for analysis was 1490.

* $P < 0.05$.

Supporting Information S5 Sensitivity Analyses

Table S6 Mean Differences in Predicted Anthropometric Outcomes from Birth to 5 Years in each Maternal Smoking Category and the Reference Category (No smoking)^a in full-term children

Outcomes by maternal smoking status	Birth		3 months		6 months		1 year		2 years		4 years		5 years	
	β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI
Length/Height, cm														
Exclusively the 1 st trimester	-0.43*	[-0.78,-0.08]	-0.05	[-0.39,0.29]	0.06	[-0.31,0.43]	0.06	[-0.36,0.47]	-0.03	[-0.54,0.48]	0.08	[-0.59,0.75]	0.20	[-0.56,0.96]
Late pregnancy	-0.69*	[-0.95,-0.42]	-0.48*	[-0.74,-0.23]	-0.35*	[-0.64,-0.07]	-0.26	[-0.57,0.06]	-0.20	[-0.58,0.19]	-0.07	[-0.58,0.43]	-0.01	[-0.58,0.57]
Weight, g														
Exclusively the 1 st trimester	-17	[-95,60]	-31	[-139,76]	-8	[-142,127]	30	[-141,200]	53	[-168,274]	58	[-274,391]	59	[-342,460]
Late pregnancy	-140*	[-199,-81]	-58	[-139,24]	-17	[-119,85]	48	[-81,177]	140	[-27,308]	266*	[14,518]	322*	[19,626]
BMI, kg/m ²														
Exclusively the 1 st trimester	0.15	[-0.08,0.39]	-0.08	[-0.30,0.14]	-0.06	[-0.29,0.17]	0.01	[-0.21,0.23]	0.07	[-0.15,0.28]	0.01	[-0.20,0.23]	-0.03	[-0.25,0.20]
Late pregnancy	-0.20*	[-0.38,-0.02]	0.11	[-0.06,0.28]	0.15	[-0.03,0.32]	0.19*	[0.03,0.36]	0.25*	[0.09,0.41]	0.27*	[0.11,0.43]	0.26*	[0.09,0.43]
Height velocities, cm/month														
Exclusively the 1 st trimester	0.061*	[0.015,0.106]	0.061*	[0.015,0.106]	0.017	[-0.01,0.045]	-0.009	[-0.032,0.015]	-0.003	[-0.015,0.009]	0.009	[-0.002,0.021]	0.011	[-0.001,0.023]
Late pregnancy	0.059*	[0.025,0.094]	0.059*	[0.025,0.094]	0.029*	[0.008,0.050]	0.008	[-0.01,0.026]	0.005	[-0.005,0.014]	0.006	[-0.003,0.014]	0.005	[-0.004,0.015]
Weight velocities, kg/month														
Exclusively the 1 st trimester	0.006	[-0.012,0.025]	0.006	[-0.012,0.025]	0.008	[-0.006,0.022]	0.004	[-0.005,0.013]	0.001	[-0.006,0.007]	0.000	[-0.007,0.007]	0.000	[-0.007,0.007]
Late pregnancy	0.015*	[0.001,0.029]	0.015*	[0.001,0.029]	0.012*	[0.002,0.023]	0.010*	[0.003,0.016]	0.006*	[0.001,0.011]	0.005	[0.000,0.010]	0.005	[-0.001,0.010]
BMI velocities, kg/m ² /month														
Exclusively the 1 st trimester	-0.059	[-0.127,0.008]	-0.059	[-0.127,0.008]	0.011	[-0.011,0.034]	0.010	[-0.001,0.022]	0.001	[-0.005,0.007]	-0.003	[-0.007,0.001]	-0.003	[-0.006,0.000]
Late pregnancy	0.030	[-0.021,0.081]	0.030	[-0.021,0.081]	0.009	[-0.008,0.026]	0.008	[-0.001,0.016]	0.003	[-0.002,0.007]	0.000	[-0.003,0.003]	-0.001	[-0.003,0.002]
Age at BMI peak ^b , days														
Exclusively the 1 st trimester					6.49	[-12.82,25.79]								
Late pregnancy					4.22	[-10.37,18.81]								
BMI at BMI peak ^b , kg/m ²														
Exclusively the 1 st trimester					-0.06	[-0.30,0.17]								
Late pregnancy					0.15	[-0.02,0.33]								

Abbreviations: BMI, body mass index, CI, confidence interval.

^aAdjusted for maternal education level, BMI and age at delivery; breast-feeding duration; recruitment center, child's gender.

^bAnalyses restricted to those with valid data for age and BMI at BMI peak (identifiable BMI peak and BMI peak before 732 days). The number of children available for analysis was 1532. P<0.05.

Table S7 Mean Differences in Predicted Anthropometric Outcomes from 3 months to 5 Years in Postnatally exposed children versus non-exposed children^a (Two-steps method)

Outcomes by maternal smoking status	3 months		6 months		1 year		2 years		4 years		5 years	
	β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI
Length/Height, cm												
Postnatal Smoking Exposure	-0.07	[-0.35,0.21]	-0.13	[-0.42,0.16]	-0.18	[-0.49,0.13]	-0.21	[-0.58,0.17]	-0.25	[-0.73,0.24]	-0.27	[-0.82,0.28]
Weight, g												
Postnatal Smoking Exposure	-8	[-92,77]	-23	[-124,79]	-45	[-170,80]	-56	[-216,104]	-39	[-274,197]	-26	[-308,256]
BMI, kg/m ²												
Postnatal Smoking Exposure	0.03	[-0.14,0.19]	0.02	[-0.15,0.19]	0.00	[-0.16,0.17]	0.00	[-0.15,0.16]	0.04	[-0.11,0.18]	0.05	[-0.10,0.20]

Abbreviations: BMI, body mass index, CI, confidence interval.

^aAdjusted for maternal education level, BMI and age at delivery; breast-feeding duration; recruitment center; child's gender.

P<0.05.

Table S8 Mean Differences in Predicted BMI from 3 months to 5 Years in each Maternal Smoking Category and the Reference Category (No Smoking)^a (Two-steps method) using FP, Splines and IM in the EDEN study

Outcomes by maternal smoking status and modeling method	3 months			6 months			1 year			2 years			4 years			5 years			
	β	95% CI	SD	β	95% CI	SD	β	95% CI	SD	β	95% CI	SD	β	95% CI	SD	β	95% CI	SD	
BMI, kg/m ²																			
Exclusively the 1 st trimester																			
Indirect Modeling	-0.05	[-0.27,0.17]	0.11	-0.04	[-0.27,0.19]	0.12	0.03	[-0.19,0.26]	0.11	0.10	[-0.12,0.31]	0.11	0.04	[-0.17,0.25]	0.11	0.00	[-0.22,0.22]	0.11	
Fractional Polynomials	-0.03	[-0.23,0.17]	0.10	0.01	[-0.20,0.22]	0.11	0.04	[-0.17,0.25]	0.11	0.05	[-0.15,0.24]	0.10	0.05	[-0.14,0.24]	0.10	0.04	[-0.15,0.24]	0.10	
Splines	-0.04	[-0.24,0.15]	0.10	0.03	[-0.20,0.25]	0.11	0.07	[-0.16,0.29]	0.12	0.06	[-0.12,0.25]	0.10	0.07	[-0.11,0.26]	0.10	0.09	[-0.11,0.28]	0.10	
Late pregnancy																			
Indirect Modeling	0.08	[-0.09,0.24]	0.09	0.11	[-0.07,0.28]	0.09	0.16	[-0.01,0.33]	0.09	0.23*	[0.07,0.39]	0.08	0.25*	[0.09,0.41]	0.08	0.24*	[0.07,0.41]	0.08	
Fractional Polynomials	0.08	[-0.07,0.23]	0.08	0.11	[-0.05,0.27]	0.08	0.14	[-0.02,0.30]	0.08	0.18*	[0.03,0.32]	0.07	0.21*	[0.07,0.35]	0.07	0.22*	[0.07,0.37]	0.07	
Splines	0.07	[-0.08,0.21]	0.07	0.14	[-0.02,0.31]	0.09	0.19*	[0.02,0.35]	0.09	0.19*	[0.05,0.33]	0.07	0.22*	[0.08,0.36]	0.07	0.24*	[0.10,0.39]	0.07	

Abbreviations: BMI, Body Mass Index; CI, confidence interval; SD, Standard Deviation

^aAdjusted for maternal education level, BMI and age at delivery; breast-feeding duration; recruitment center; child's gender

* $P < 0.05$

Table S9 Mean Differences in Height, weight and BMI obtained from clinical visits (1, 3 and 5 years) and by modeling in each Maternal Smoking Category and the Reference Category (No Smoking)^a (Two-steps method)

Outcomes by maternal smoking status and by types of data	1 year		3 years		5 years	
	β (n)	95% CI	β (n)	95% CI	β (n)	95% CI
Length/Height, cm						
Exclusively the 1 st trimester						
Data from Clinical visits	0.04 (119)	[-0.43,0.51]	0.08 (99)	[-0.66,0.81]	-0.15 (82)	[-1.19,0.89]
Model-predicted Data (subsample with clinical visits data)	0.08 (119)	[-0.35,0.52]	0.06 (99)	[-0.62,0.74]	-0.15 (82)	[-1.11,0.81]
Model-predicted Data (all children included in the main analyses)	0.18	[-0.24,0.60]	0.09	[-0.50,0.67]	0.35	[-0.40,1.11]
Late pregnancy						
Data from Clinical visits	-0.14 (228)	[-0.50,0.22]	-0.17 (159)	[-0.78,0.44]	-0.25 (142)	[-1.09,0.58]
Model-predicted Data (subsample with clinical visits data)	-0.24 (228)	[-0.55,0.08]	-0.17 (159)	[-0.74,0.39]	-0.26 (142)	[-1.03,0.51]
Model-predicted Data (all children included in the main analyses)	-0.23	[-0.55,0.08]	-0.13	[-0.57,0.31]	0.02	[-0.55,0.58]
Weight, g						
Exclusively the 1 st trimester						
Data from Clinical visits	5 (118)	[-194,204]	-62 (99)	[-396,272]	-21 (82)	[-612,571]
Model-predicted Data (subsample with clinical visits data)	29 (118)	[-151,208]	72 (99)	[-242,387]	-20 (82)	[-537,497]
Model-predicted Data (all children included in the main analyses)	75	[-95,245]	113	[-158,383]	137	[-261,535]
Late pregnancy						
Data from Clinical visits	79 (228)	[-74,232]	125 (160)	[-152,402]	358 (141)	[-117,833]
Model-predicted Data (subsample with clinical visits data)	71 (228)	[-67,210]	151 (160)	[-110,413]	324 (141)	[-90,739]
Model-predicted Data (all children included in the main analyses)	33	[-95,161]	189	[-14,392]	307*	[8,606]
BMI, kg/m²						
Exclusively the 1 st trimester						
Data from Clinical visits	-0.01 (118)	[-0.29,0.27]	-0.09 (98)	[-0.33,0.15]	0.01 (82)	[-0.29,0.30]
Model-predicted Data (subsample with clinical visits data)	0.01 (118)	[-0.23,0.24]	0.05 (98)	[-0.19,0.28]	0.01 (82)	[-0.27,0.29]
Model-predicted Data (all children included in the main analyses)	0.03	[-0.19,0.26]	0.08	[-0.13,0.28]	0.00	[-0.22,0.22]
Late pregnancy						
Data from Clinical visits	0.2 (228)	[-0.01,0.42]	0.18 (158)	[-0.02,0.37]	0.36* (141)	[0.12,0.60]
Model-predicted Data (subsample with clinical visits data)	0.23* (228)	[0.05,0.41]	0.21* (158)	[0.02,0.41]	0.33* (141)	[0.11,0.55]
Model-predicted Data (all children included in the main analyses)	0.16	[-0.01,0.33]	0.25*	[0.09,0.40]	0.24*	[0.07,0.41]

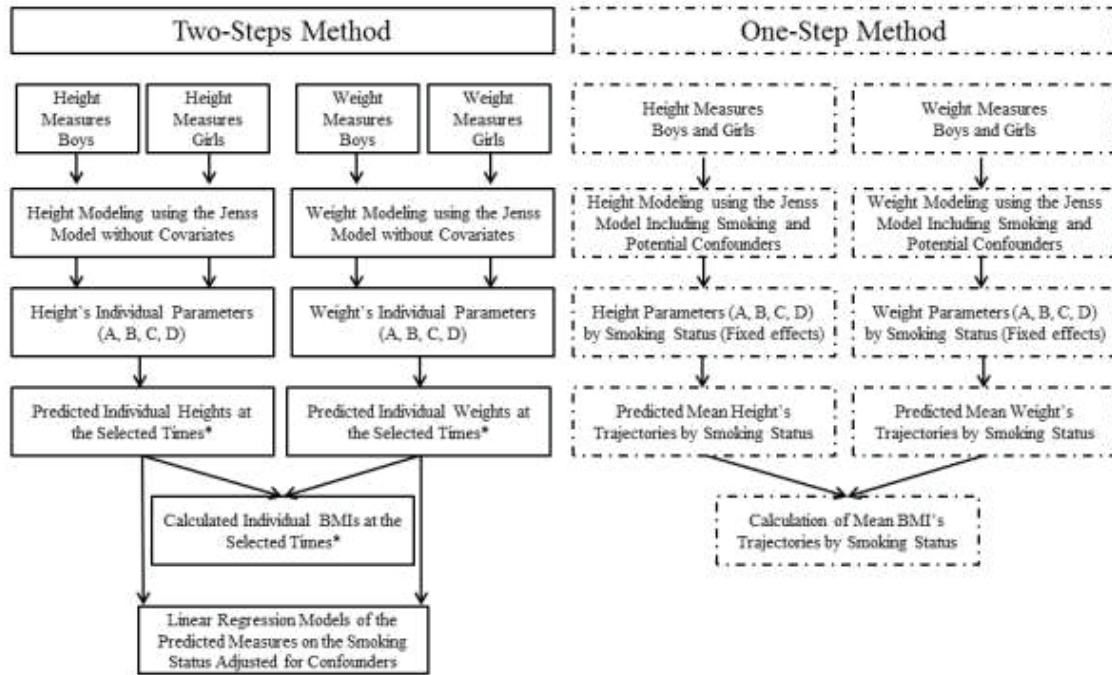
^aAdjusted for maternal education level, BMI and age at delivery; breast-feeding duration; recruitment center; child's gender.

Abbreviations: BMI, body mass index, CI, confidence interval.

* $P < 0.05$.

FIGURES

Figure 1.



*Selected times are 3 and 6 months and 1, 2, 4, 5 years

Figure 2.

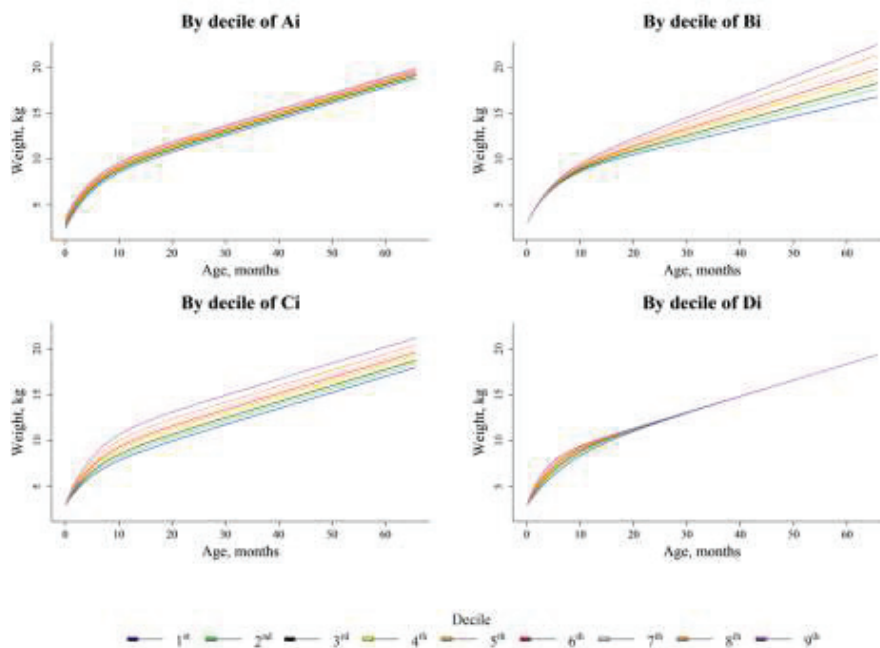
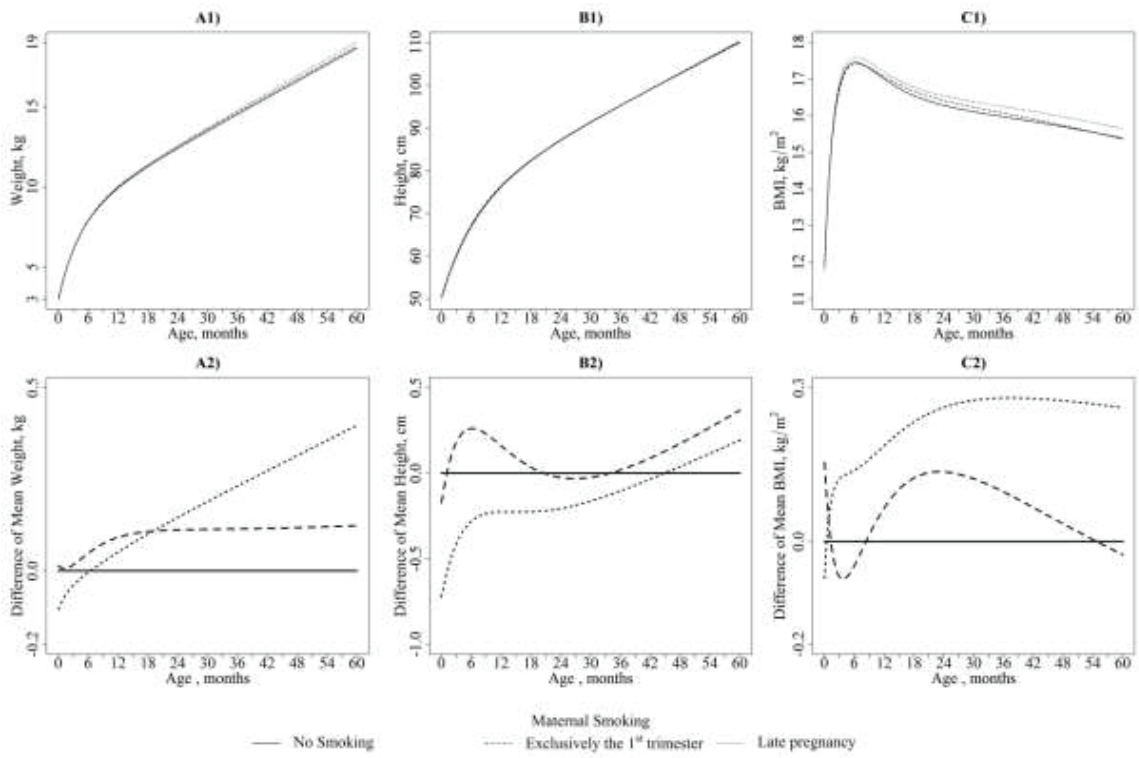


Figure 3.



Joint Bayesian weight and height post-natal growth model to study the effects of maternal smoking during pregnancy

Sophie Carles, Marie-Aline Charles, Barbara Heude, Ismaïl Ahmed, Jérémie Botton and the EDEN mother-child study group

Abstract

Growth models used for describing the dynamics of body weight and height generally consider each trait independently. We proposed to model height and weight trajectories jointly with a non-linear heteroscedastic mixed model based on the Jenss-Bayley growth function with correlated individual random effects and using Bayesian inference techniques. Simulations first showed that our model provides good estimates of the growth parameters. Then, we illustrated how it can be used to assess the associations between maternal smoking during pregnancy, an early-life factor potentially involved in prenatal programming of obesity, and children's growth from birth to five years of age. We used real data from the EDEN study, a large French mother-child cohort study with a high number of height and weight measurements (approximately 30,000 measurements each in 1,666 children). Our results supported the existence of a relationship between maternal smoking during pregnancy and growth from birth to five years of age. Children from mothers who smoked throughout pregnancy were shown to display a higher BMI from the first few months of life onwards compared to children from non-smokers. At five years of age, their mean BMI was 0.21 kg/m² higher than unexposed children. It was mainly explained by the fact that those children tended to be smaller at birth but rapidly exceeded the weights of children from non-smokers.

Keywords: Bayesian modeling; multivariate; joint model, postnatal growth trajectories; non-linear modeling.

1 Introduction

Growth of children is used as an indicator of health and development and is influenced by genetic and environmental factors. In epidemiology, growth is studied by different approaches that generally aim to identify critical periods when growth is associated with later health or diseases or as an outcome to study its determinants. Among them, structural modeling is attractive to provide parameters with meaningful

information on key growth features [1]. Most growth models are generally implemented to assess patterns of height and weight growth separately despite interest in jointly studying their evolution given the high individual correlation between these two traits. Appropriately modeling children's growth during the first few years of life creates statistical challenges due to the complexity of growth patterns. Repeated collections of height and weight measures are needed to study growth and must account for the non-independence of measures of a same individual across time. Mixed-effects models provide a flexible and powerful tool to analyze these data [2].

Growth patterns have been fitted using non-linear models based on the structural Jenss-Bayley equation [3] that adequately fit growth data from birth to eight years and describe individual growth trajectories using four subject-specific parameters with clinical interpretations [1]. The likelihood of non-linear mixed models (NLMMs) does not have a closed-form expression due to the integration of random effects, making inference about parameters somewhat tedious [4]. Stochastic versions of the EM (Expectation Maximization) algorithms as Monte Carlo EM (MCEM) have been proposed [5]. The "stochastic approximation EM algorithm" (SAEM) is one that is more efficient computationally than MCEM due to smoothing on the E step based on recycling simulations from between iterations. This algorithm is now implemented in R [6] in the package SAEMIX [7].

Bayesian approaches to joint NLMMs have been proposed especially for HIV dynamics and shown to be a powerful way to analyze such complex data. Indeed, specifying prior distributions for the model parameters allows to gain insight from the results of previous studies but also to handle some parameter identifiability problems. The Bayesian approach also offers a flexible way to fit complex models, especially hierarchical or multi-level ones [8]. One major interest, inherent to Bayesian reasoning, is that the full posterior distribution of parameters is available. Thus, this approach is suitable for predictions that account for uncertainty in the parameters [9].

Herein, we propose a joint Bayesian model of height and weight growth and illustrate how it can be used to assess the associations between maternal smoking during pregnancy, an early-life factor potentially involved in prenatal programming of obesity, and children's growth from birth to five years of age. We used real data from the EDEN study, a large French mother-child cohort study with a high number of height and

weight measurements (approximately 30,000 measurements each in 1,666 children). We hypothesized that the Bayesian approach would offer computational facilities especially when the number of factors to consider for their association with the growth trajectory is high. Our approach should allow to simultaneously assess the associations between the determinant of interest and growth in weight and height and infer the association with Body Mass Index (BMI).

In section 2, we introduce the bivariate model based on the Jenss-Bayley growth function. In section 3, we apply this model to simulated data to assess its ability to adequately fit growth data by providing accurate parameter estimates. In section 4, we apply the model to assess the adjusted association between maternal smoking and height, weight and BMI growth on a set of growth data resulting from the follow-up of children from the EDEN mother-child cohort study. Section 5 provides some elements of discussion.

2 Model for postnatal growth during childhood

2.1 A general 3 stage hierarchical model

The joint model consists of the following two non-linear mixed effects sub-models: one for height and the other for weight. The model can be displayed as a general 3-stage hierarchical Bayes model along the same line as in Davidian and Giltinan [10].

2.1.1 Stage 1: intra-individual variability

We start by writing the NLMM under its usual frequentist expression, as follows:

$$y_{ijk} = f(\boldsymbol{\phi}_{ik}, t_{ijk}) + e_{ijk} \quad (1)$$

$k = (H, W)$ denotes the two outcomes of interest i.e., height and weight. $\mathbf{y}_{ik} = (y_{i1k}, \dots, y_{in_{ik}k})^T$ denotes a n_{ik} vector of observed measures of height or weight for individual i at time t_{ijk} , $i=1, \dots, N$, $j=1, \dots, n_{ik}$. f is a non-linear function of a m -vector of individual specific parameters $\boldsymbol{\phi}_{ik}$. $\mathbf{e}_{ik} = (e_{i1k}, \dots, e_{in_{ik}k})^T$ is a n_{ik} vector of residual errors of height or weight that are assumed to be independently and normally distributed with a null mean and variance $\sigma_{e_k}^2$ so that

$$\mathbf{e}_{ik} = (e_{ijk}) \sim N(0, \sigma_{e_k}^2 I_{n_{ik}}) \quad (2a)$$

where $I_{n_{ik}}$ is the identity matrix.

For the simulation study (see Section 3), we considered that the residual variances were homogeneous with respect to age for both height and weight.

$$\log(\sigma_{e_{ijk}}^2) = \delta_{0k}$$

For the application (see Section 4), we used a general variance function involving residual variations with age to adjust for heteroscedasticity, as described by Foulley et al. and Duval et al. [11,12]. Two functions were considered to model the residual variance for both height and weight sub-models, separately within the joint model ($\log(\sigma_{e_{ijk}}^2) = \delta_{0k} + \delta_{1k}t_{ijk}$ and $\log(\sigma_{e_{ijk}}^2) = \delta_{0k} + \delta_{1k}t_{ijk} + \delta_{2k}t_{ijk}^2$). The combination of functions for weight and height sub-models that minimized the fit parameters was selected.

Finally, we considered the following equation for height:

$$\log(\sigma_{e_{ijH}}^2) = \delta_{0H} + \delta_{1H}t_{ijH}$$

and the following equation for weight:

$$\log(\sigma_{e_{ijW}}^2) = \delta_{0W} + \delta_{1W}t_{ijW} + \delta_{2W}t_{ijW}^2$$

Formula (1) and (2a) are equivalent in the Bayesian hierarchical models to write the conditional distribution of the data given the individual parameters ϕ_{ik} as

$$\mathbf{y}_{ik} | \phi_{ik} \sim N(f_{ik}, \sigma_{e_k}^2 I_{n_{ik}}) \quad (2b)$$

where $\mathbf{y}_{ik} = (y_{ijk})$ for $j=1, \dots, n_{ik}$ is a n_{ik} vector and similarly f_{ik} is defined as $f_{ik} = (f(\phi_{ik}, t_{ijk}))$ for $j=1, \dots, n_{ik}$.

For the weight and height modeling, f is a modification of the initial non-linear Jenss-Bayley equation that improves the interpretation of parameters and accounts for mathematical constraints [13]. This growth function allows the expression of expectations of height in centimeters (cm) and weight in kilograms (kg) for the individual i at time t_{ijk} , as follows:

$$f(\boldsymbol{\phi}_{ik}, t_{ijk}) = \exp(A_{ik}) + \exp(B_{ik}) t_{ijk} + \exp(C_{ik}) (1 - \exp(-\exp(D_{ik}) t_{ijk})) \quad (3)$$

with $k = (H, W)$ and $\boldsymbol{\phi}_{ik} = (A_{ik}, B_{ik}, C_{ik}, D_{ik})^T$.

2.1.2 Stage 2: inter-individual variability

Individual effects are treated as random with the following multinormal distribution:

$$\boldsymbol{\phi}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Omega}) \quad (4)$$

where $\boldsymbol{\phi}_i = (\boldsymbol{\phi}_{iH} \ \boldsymbol{\phi}_{iW})^T$ with height $\boldsymbol{\phi}_{iH}$ and weight $\boldsymbol{\phi}_{iW}$ vector components having mean vector $\boldsymbol{\mu}_i = (\boldsymbol{\mu}_{iH} \ \boldsymbol{\mu}_{iW})^T = (\mu_{iH}^A, \dots, \mu_{iH}^D, \mu_{iW}^A, \dots, \mu_{iW}^D)^T$ and variance covariance matrix $\boldsymbol{\Omega} = \begin{bmatrix} \Omega_{HH} & \Omega_{HW} \\ \Omega_{WH} & \Omega_{WW} \end{bmatrix}$ of dimension 2 m x 2 m.

In formula (3), $\exp(A_{iH})$ describes the length at birth of subject i . For weight modeling, to bypass the neonatal weight loss not correctly fitted by the model, birth weight measures were excluded and the minimum weight in the first four days of life was selected for each child. Consequently, in the weight model of subject i , $\exp(A_{iW})$ represents the extrapolated weight at time 0 without accounting for neonatal weight loss; for any trait k , $\exp(B_{ik})$ is the growth velocity of height or weight beyond two years of age; $\exp(C_{ik})$ is the spurt of height/weight growth in the first 2 years of life and $\exp(D_{ik})$ is the trajectory curvature in early life (combination of growth velocity and deceleration rate in the beginning of life). Means $\boldsymbol{\mu}_{ik}$ of the distribution shown in (4) are related to a matrix of explanatory variables \mathbf{X}_i of dimension $m \times p$ (intercept and discrete or continuous covariates) via a classical linear model involving a p dimensional vector of parameters $\boldsymbol{\beta}_k$.

$$\boldsymbol{\mu}_{ik} = \mathbf{X}_i \boldsymbol{\beta}_k \quad (5)$$

In models without covariates, the vector of “fixed” effects $\boldsymbol{\beta}_k$ is restricted to intercept elements $(\beta_{0Ak}, \beta_{0Bk}, \beta_{0Ck}, \beta_{0Dk})^T$ corresponding to the population means of growth parameters. In models with covariates on growth parameters (cf. section 4), $\boldsymbol{\beta}_k$ includes both a vector of m parameters intercepts $\beta_{0Ak}, \beta_{0Bk}, \beta_{0Ck}, \beta_{0Dk}$ and a vector of

$m p_{cov}$ regression coefficients $\beta_{1Ak} \dots \beta_{p_{covAk}}, \dots, \beta_{1Dk} \dots \beta_{p_{covDk}}$ pertaining to the p_{cov} covariates of interest.

2.1.3 Stage 3: hyperprior distributions

We specified standard prior distributions usually employed for Bayesian hierarchical models [9,14] and parameter values that take advantage of published studies in the field of growth, including ours [13].

The prior distributions are described in Table I (in rjags notations). Briefly, for the weight sub-model, a mean value of 1 for the A parameter specifies the prior expectation of the extrapolation of the average minimal weight after birth to be 2.7 kg and a precision (inverse of variance) of 16 (95% probability in the range [1.70, 4.40]). We defined a mean B parameter value of -5, a mean C parameter of 1.5 and a mean D parameter of -5. A precision of 16 was also considered for these three parameters. The same approach was applied for the height sub-model. The priors for the height model parameters (see Table I for prior distribution) were adapted to provide plausible birth lengths and height trajectories.

In the presence of covariates, to keep the values generating realistic growth trajectories, we assumed the prior distribution for all the coefficients of the β_k vector except the intercept terms to be normally distributed with a mean zero and a precision of 16, (i.e. a variance of 0.0625 or a standard deviation of 0.25), i.e., in the range [-0.49, 0.49]) with a 95% probability for $k = (H, W)$.

For the residual precision and the matrix of precision, we used prior distributions conveniently fitted to Bayesian hierarchical models (e.g., de la Cruz et al. and Donnet et al. for nonlinear mixed models) [9,14]. For the simulation study, taking advantage of conjugacy properties, the residual precision was assumed to follow a Gamma distribution $\sigma_{e_k}^{-2} \sim \gamma(a, b)$. Here, we used $a = b = \varepsilon$ with ε set equal to a small value appropriately chosen (i.e., here $\varepsilon=10^{-4}$) to approximate a non-informative Jeffreys prior density. For the application, normal distributions with a mean of zero and precision of 16 a priori distribution was specified on the parameters of the residual variance function δ_{0H} and δ_{1H} for height and δ_{0W}, δ_{1W} and δ_{2W} for weight. The precision matrix $\Gamma = \Omega^{-1}$ of the 8 individual parameters (A, B, C, D for height and weight) was supposed

to follow a Wishart distribution $W(R, \nu)$, a convenient distribution for positive-definite matrices due to conjugacy properties. Γ is thus a priori centered on νR and has ν degrees of freedom [15]. As advocated by Gelman and Hill, ν was set to $2m+1=9$ [16]. The specification of the prior mean, was based on previous estimations of the variance-covariance between the height and weight model parameters (see supplementary material I for detailed values).

2.2 Bayesian implementation

2.2.1 Algorithm and software

The full conditional posterior distribution of all model parameters was calculated using Markov Chain Monte Carlo (MCMC) methods within the rjags package (version 3-15) in R software (version 3.1.0) [17]. In this package, MCMC are simulated using the Gibbs sampler. Within this iterative process, after choosing arbitrary starting values, each parameter of the model is sampled for each iteration t from the conditional distribution given all the others parameters at iteration $t-1$. In other words, for each iteration s , each component θ_j of the whole vector of parameters θ will be sampled conditional on the most recent values of the other components of θ [18,19].

2.2.2 Computational aspects

Models were fitted using 40,000 iterations for the burn-in (warm-up) phase that were not used for posterior inference. After checking for chain mixing, 20,000 additional iterations were performed. Chains were thinned to keep parameter estimates at intervals of 20.

2.2.3 Convergence assessment

Convergence assessments were based on visual inspections of the simulation trace plot for each parameter and on the Brooks-Gelman-Rubin criteria [19–21]. This criteria compares between and within chain variability using the ratio of the total to the within chain variations. A criterion close to 1 indicates a good convergence.

2.3 Model selection and evaluation

In the application (see Section 4), to test the effects of smoking covariates on height and weight trajectories, models with and without those covariates were compared using the

Deviance Information Criterion (DIC) [22] and the Watanabe Akaike Information Criterion (WAIC) [23] (see supplementary material II for details on both criterions).

To validate the selected model, posterior predictive checks derived by Gelman et al. were done [19]. At each simulations (retained iterations) s ($s = 1, \dots, S$), simulated replicated data $y^{\text{rep } s}$ were generated under the model from the predictive distribution of θ^s at the corresponding simulation. At each simulation, using a Chi-squared discrepancy measure, we assessed, for height and weight separately, the test quantity for replicated datasets $T(y_k^{\text{rep } s}, \theta^s)$ and for observed data $T(y_k, \theta^s)$. We calculated the posterior predictive p-value for each trait k (p_{B_k}) that corresponds to the proportion of simulations S for which :

$$T(y_k^{\text{rep } s}, \theta^s) \geq T(y_k, \theta^s).$$

$$\text{where } T(y_k, \theta^s) = \sum_{i=1}^N \sum_{j=1}^n \frac{(y_{ijk} - E(y_{ijk} | \theta^s))^2}{\text{Var}(y_{ijk} | \theta^s)} \text{ and } (y_k^{\text{rep } s}, \theta^s) = \sum_{i=1}^N \sum_{j=1}^n \frac{(y_{ijk}^{\text{rep } s} - E(y_{ijk} | \theta^s))^2}{\text{Var}(y_{ijk} | \theta^s)}.$$

3 Simulation study

3.1 Simulated growth data

We performed a simulation study to assess and illustrate the ability of the model to adequately estimate the key growth features with respect to classical properties of estimators (bias, standard and mean square error). Datasets were simulated mimicking height and weight growth of typical children. A total of 100 datasets were generated each containing the anthropometric data of 500 children with 6 measurements at time points ranging from day 5 to day 1825: days 5, 30.44 (1 month), 91.32 (3 months), 182.64 (6 months), 273.96 (9 months), 365.25 (1 year), 730.5 (2 years), 1095.75 (3 years), 1461 (4 years), 1826.25 (5 years). Population parameters values were guided by the results obtained from a previous study through a frequentist SAEM algorithm.

- 1) Each individual parameter was sampled from a multinormal distribution with the mean parameter values as described in Table II (see true values). The 8-dimensional variance covariance matrix between the individual height and weight parameters was empirically chosen according to previous research (see values of Ω_0 in supplementary material I).

- 2) Heights and weights were then predicted at each time point (days 5-1825) for each child in the simulated dataset by applying the Jenss-Bayley equation. We considered these values as the true height and weight measurements free from measurement error.
- 3) Consequently, we introduced some measurement error to generate data that would be closer to those reported in practice. For each individual in the simulated dataset, the observed data at the measured time were obtained by sampling one value from a normal distribution with the mean being the predicted value (2) and a standard error of 0.40 both for height (in cm) and weight (in kg).

We ran the Bayesian model on the 100 simulated datasets. We fixed a burn-in period of 30,000 iterations for 3 chains and ran the chains for an additional 20,000 iterations, collecting samples at intervals of 20 iterations.

The bias (i.e., difference between the true value of the parameters (step 1) and the mean estimate obtained), the relative bias (i.e., ratio of the bias over the true value), the RMSE (i.e., root mean square error) and the probability coverage (i.e., the proportion of 95% credible interval containing the true value) were additionally computed. Estimates from the Bayesian models used for computations were the expectation of the posterior distribution.

3.2 *Simulation results*

Estimates obtained for the eight main parameters (A , B , C and D for height and weight parameters) were quite satisfactory from a frequentist point of view with absolute and relative biases closed to 0 and coverage probabilities ranging from 90 to 99% and mainly between 93 and 97% (Table II summarizes the results for the main parameters and their variance; results for the covariance between parameters are described in supplementary material III). Most other parameters (i.e., variances and covariances) were also adequately estimated by the model, except a rather lower ability to estimate the covariance between the A parameter of weight and the D parameter of height and weight (relative bias, 0.41 and -0.28 respectively).

Figure I and Figure II illustrate the simulated and fitted (mean trajectories and 95% credible interval of the trajectories) weight and height growth trajectories for three randomly selected children and provide a visual validation of the model adequacy.

4 Application to the analysis of associations between maternal smoking during pregnancy and growth

4.1 *The EDEN study dataset*

We applied the joint Bayesian model to the data from the EDEN study, a birth cohort designed to assess pre and early postnatal determinants of child health and development [24]. We analyzed weight and height records (30,515 and 28,381, respectively) of 1,666 children from birth to five years of age. Our aim was to assess whether height, weight and BMI growth patterns differ between children exposed to different degrees of maternal smoking, namely: exposure during only the first trimester of pregnancy, throughout pregnancy, and unexposed children. To reduce confounding, we adjusted our analyses for the following variables: maternal education level (high school or less; 2-year university degree; 3-year university degree), maternal pre-pregnancy BMI (overweight/obese or not), maternal age at delivery (continuous), breastfeeding duration (≥ 3 months or < 3 months), recruitment center, and child's gender. For computational reasons, maternal age was centered on its population mean.

4.2 *Associations between Maternal Smoking and Height and Weight Growth Parameters*

We obtained a DIC of 121013 for the model without the smoking covariates. The DIC value was smaller for the model that included smoking covariates (120997). This was also true with WAIC (121646 *versus* 121669 for the models with and without smoking covariates). We interpreted these differences as an association between maternal smoking and growth.

Furthermore, our model with smoking covariates had a global posterior predictive p-value of 0.48 for height and 0.60 for weight. Graphical plots of $T(y^{\text{rep}}, \theta)$ against $T(y, \theta)$ also indicated good predictive ability of the model (Joint distribution of simulation draws of $T(y, \theta)$ and its replication $T(y^{\text{rep}}, \theta)$ are illustrated in supplementary material IV).

Figure III and Figure IV describe the collected weight and height observations for 6 randomly selected children with a mean growth trajectory predicted by the model (i.e., mean of the posterior distribution) and its 95% credibility interval. This finding provides a visual indication that the model adequately fits the height and weight data.

Table III describes the association between maternal smoking and the main growth parameters of height and weight. Women who smoked throughout pregnancy delivered children with lower birth lengths (A_H parameter regression coefficient = -0.015, 95% credibility interval: -0.021, -0.008), compared with non-smoking women. The A_W parameters were also lower for these mother-child pairs. Maternal smoking throughout pregnancy was also associated with higher B_W parameters in children, indicating a higher growth velocity of weight from 2 years of age onwards. Maternal smoking exclusively during the first trimester was associated with a higher D_H parameter, indicating a higher curvature degree of the height growth curve in the first months of life.

More moderate associations were observed for some parameters (Figure V). The posterior distribution of the regression coefficients for both maternal smoking exclusively during the first trimester and maternal smoking throughout pregnancy shifted toward positive values for height B_H . The regression coefficient for the height D_H parameter also shifted toward positive values for maternal smoking throughout pregnancy. The posterior distribution of the regression coefficient for C_H shifted toward negative values for maternal smoking during the first trimester exclusively. This finding provides evidence of a greater height growth velocity after 2 years of age in exposed children from both groups (for B_H), a greater height spurt in the first 2 years of life in children exposed in early pregnancy (for C_H) and a greater curvature of height in children exposed throughout pregnancy (for D_H).

To quantify the differences in height, weight and BMI between the children exposed during only the first trimester or children exposed throughout pregnancy and the reference category, we simulated height, weight and BMI at day 5, 15, 25, 30 and then at each month from 30 days to 5 years for fictitious children in each smoking group (10000 children per group), setting fixed values for covariates (male gender, breastfeeding less than 3 months, birth in Poitiers, and having mothers with a normal BMI, a high school level or less and an age of 30 years at delivery). The mean height, weight and BMI values at each time point were computed among subgroups of children. The differences in the mean values between children whose mothers smoked exclusively during the first trimester or throughout pregnancy and the reference category (i.e., children whose

mothers did not smoke) were then computed. Table IV and Figure VI illustrate the mean posterior estimates of these differences and their 95% credibility intervals.

Maternal smoking during only the first trimester was not associated with any difference in BMI, height or weight at the selected ages. Children whose mothers smoked throughout pregnancy were born shorter (mean difference at 5 days: -0.71 cm, 95% credibility interval: -1.03, -0.38) and lighter. These babies had a 0.05 kg/m² lower BMI at 5 days (95% credibility interval: -0.23, 0.12). On average, those children reached the weight and height trajectories of the reference category between 4 and 6 months and approximately 4 years later, respectively. Children exposed to maternal smoking throughout pregnancy tended to weigh more than children of non-smoking mothers at the end of the study period : at 4 and 5 years of age, they weighed 231 and 301 grams more (credibility intervals: -36, 505 and -24, 639, respectively). Children whose mothers smoked throughout pregnancy tended to be smaller and heavier than those from non-smokers. Consequently, maternal smoking throughout pregnancy was positively associated with BMI in those children. Between 570 and 600 days (19-20 months), the BMI of children whose mothers smoked throughout pregnancy already surpassed non-smokers, reaching 0.21 kg/m² 1825 days after birth, i.e., at 60 months (credibility interval: 0.03, 0.40).

5 Discussion

We propose a method to jointly model height and weight in children 0-5 years of age. The simulation study indicated that the joint modeling yielded unbiased estimates and a good coverage probability, reflecting its ability to adequately fit theoretical growth data.

The Bayesian framework offers a flexible technique to predict expected heights, weights and BMIs of children, allowing the study of its potential determinants, such as socio-demographical profiles or environmental exposures. In our example, the joint model allowed the assessment of growth trajectories and predict the mean differences over time based on maternal smoking and the credibility intervals, not only for weight and height but also for BMI.

Our results showed that children from mothers who smoked throughout pregnancy had higher BMIs already in their first months of life compared to children from non-smokers.

At 5 years, they had a 0.21 kg/m² higher BMI. These results support previous studies that found higher BMIs or risk of overweight in children from mothers who smoked during pregnancy [25,26]. The model provides key information about the dynamics of the association. As described elsewhere, contrarily to a direct BMI modeling approach, the different outcomes (e.g., height and weight) do not need to be measured at the same time points in joint modeling and the model does not assume that the same number of repeated measurements are available for the different outcomes [27]. Therefore, joint modeling enables the use of all available information on height and weight to predict BMI trajectory.

One of the main strengths compared to the frequentist approach is the fact we can infer BMI and obtain credibility intervals for BMI trajectories of each category of maternal smoking during pregnancy in one step using joint models of height and weight and determine the differences between the trajectories of the subgroups of interest using the posterior predictive distribution of the estimated parameters. Indeed, in the frequentist approach, the estimated height and weight parameters can be used to estimate BMI in subgroups of interest and differences across subgroups, but the uncertainty in the parameters is ignored, and consequently, the confidence intervals of the differences cannot be assessed in a single step like the Bayesian approach and require additional simulation studies. In the Bayesian one, the uncertainty of the parameters is considered in the predictions.

Another strength of our approach, related to structural models and not specific to the Bayesian framework, relies on the interpretability of the growth parameters, providing biological information on the individual growth processes compared to the main growth modeling parameters of the non-structural techniques such as splines or fractional polynomials. To our knowledge, a single recently published paper used a similar approach. In this study, the model developed by Cole et al. [28], called the "Superimposition by Translation and Rotation" (SITAR) was generalized in the Bayesian framework to jointly model multiple outcomes of embryonic growth. This multivariate model described individual growth by three subject-specific parameters with an obvious meaning [29]. This model was applied to study prenatal growth but could be applied to postnatal growth data.

We intended to make the model as simple as possible using standard assumptions in NLMM methodology. However, distribution assumptions made about the data model could have been changed or made more complex in several ways. First, the Jentsch-Bayley growth curve function was parameterized according to a modified form (equation 3) based largely on sign constraints on parameters A , B and D resulting at the individual level in lognormal *versus* normal assumptions on these parameters. Instead, we could have considered the original form:

$$f(\boldsymbol{\phi}_{ik}, t_{ijk}) = A_{ik} + B_{ik}t_{ijk} + \exp(C_{ik}) [1 - \exp(-D_{ik}t_{ijk})]$$

with the usual normality assumptions on $\boldsymbol{\phi}_{ik} = (A_{ik}, B_{ik}, C_{ik}, D_{ik})^T$.

However, due to convergence issues using this original form despite the use of truncated normal distribution for a priori distribution of the parameters, we used the modified form for the final model.

We assumed that there was no correlation $\sigma_{e_{HW}} = 0$ between the residual components of height e_{ijH} and weight e_{ijW} recorded at the same occasion j and on the same individual i . This is a reasonable assumption as long as the correlation between the weight and height measurements at the same time originates from the subject i or the factors included in the model.

We selected among two candidate functions ($\log(\sigma_{e_{ijk}}^2) = \delta_{0k} + \delta_{1k}t_{ijk}$ and $\log(\sigma_{e_{ijk}}^2) = \delta_{0k} + \delta_{1k}t_{ijk} + \delta_{2k}t_{ijk}^2$) to describe the residual variance of height and weight but did not perform a formal comparison with other possibilities. Consequently, other functions could have improved the aforementioned model.

The normality assumption of the residual errors could also have been investigated. For instance, the Student t-distribution has been proposed for heavy tailed or contaminated data with outliers or, more generally, the generalized error distribution for any type of kurtosis or skewness [30].

We took advantage of results from previous studies by specifying prior distributions of the mean and covariate effects. Regarding the inter-individual variance covariance matrix, we adopted the inverse Wishart distribution. As previously described, this

practice is by far the most common due to its conjugate properties and therefore, simplicity, but having just one parameter as the degree of freedom forces to have the same uncertainty in all elements of the variance covariance matrix. However, the influence of every prior distribution was limited in our study considering the large amount of information in this dataset (both in terms of the total number of measurements on height and weight and in the number of individuals recorded).

In conclusion, due to the complexity of the non-linear mixed effects models used for the height and weight modeling, the computational facilities offered by the Bayesian approach add to the above-described advantages [8] and is one of the key reason why we applied it. We showed that using datasets with rather large number of subjects and measurements by subjects, our model was suitable for growth modeling. The good properties of this joint model could be challenged in less favorable contexts. We assessed the associations between maternal smoking and multiple dimensions of growth, considering several potential confounders in these associations. The model we used could be extended by adding additional covariates of interest. However, the model is already complex, and the computational time is non negligible. The ability to fit more complex models could consequently be limited by practical aspects.

Acknowledgements

We thank Jean-Louis Foulley for introducing us to Bayesian modeling and providing us advices for the conduction of the analyses.

We thank Lise Giorgis-Allemand and Anne Forhan for data management, the midwife research assistants (Lorraine Douhaud, Sophie Bedel, Brigitte Lortholary, Sophie Gabriel, Muriel Rogeon, and Monique Malinbaum) for data collection, and the data entry operators (Patricia Lavoine, Josiane Sahuquillo, and Ginette Debotte).

References

1. Hauspie RC, Cameron N, Molinari L. *Methods in Human Growth Research*. Cambridge University Press; 2004.
2. Pinheiro JC, Bates DM. *Mixed-Effects Models in S and S-PLUS*. Springer Science & Business Media; 2000.
3. Jenss R, Bayley N. A mathematical method for studying the growth of a child. *Hum Biol.* 1937;9:556–63.
4. Samson A, Lavielle M, Mentré F. The SAEM algorithm for group comparison tests in longitudinal data analysis based on non-linear mixed-effects model. *Stat Med.* 2007;26:4860–75.
5. Levine RA, Casella G. Implementations of the Monte Carlo EM Algorithm. *Journal of Computational and Graphical Statistics.* 2001;10:422–39.
6. R Core Team (2015). R: A language and environment for statistical computing. [Internet]. R Foundation for Statistical Computing, Vienna, Austria; Available from: <https://www.R-project.org/>
7. Comets E, Lavenu A, Lavielle M. SAEMIX, an R version of the SAEM algorithm for parameter estimation in nonlinear mixed effect models [Internet]. 2012 [cited 2015 Jan 7]. Available from: http://papersjds14.sfds.asso.fr/submission_54.pdf
8. Huang Y, Liu D, Wu H. Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. *Biometrics.* 2006;62:413–23.
9. De la Cruz-Mesía R, Marshall G. Non-linear random effects models with continuous time autoregressive errors: a Bayesian approach. *Stat Med.* 2006;25:1471–84.
10. Davidian M, Giltinan DM. *Nonlinear Models for Repeated Measurement Data*. CRC Press; 1995.

11. Foulley JL, San Cristobal M, Gianola D, Im S. Marginal likelihood and Bayesian approaches to the analysis of heterogeneous residual variances in mixed linear Gaussian models. *Computational Statistics & Data Analysis*. 1992;13:291–305.
12. Duval M, Robert-Granié C, Foulley J-L. Estimation of heterogeneous variances in nonlinear mixed models via the SAEM-MCMC algorithm with applications to growth curves in poultry. *Journal de la Société Française de Statistique*. 2009;150:65–83.
13. Botton J, Scherdel P, Regnault N, Heude B, Charles M-A. Postnatal weight and height growth modeling and prediction of body mass index as a function of time for the study of growth determinants. *Ann. Nutr. Metab.* 2014;65:156–66.
14. Donnet S, Foulley J-L, Samson A. Bayesian analysis of growth curves using mixed models defined by stochastic differential equations. *Biometrics*. 2010;66:733–41.
15. Anderson T. *An Introduction to Multivariate Statistical Analysis*. second edition. Wiley; 1984.
16. Gelman A, Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press; 2006.
17. Plummer M, Stukalov A. rjags: Bayesian graphical models using MCMC [Internet]. 2014 [cited 2014 Oct 9]. Available from: <http://cran.r-project.org/web/packages/rjags/index.html>
18. Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. *The BUGS Book: A Practical Introduction to Bayesian Analysis*. CRC Press; 2012.
19. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis, Third Edition*. CRC Press; 2013.
20. Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. *Statist.Sci.* 1992;7:457–72.
21. Cowles MK, Carlin BP. Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *Journal of the American Statistical Association*. 1996;91:883.
22. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2002;64:583–639.
23. Watanabe S. Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *J. Mach. Learn. Res.* 2010;11:3571–94.
24. Heude B, Forhan A, Slama R, Douhaud L, Bedel S, Saurel-Cubizolles M-J, et al. Cohort Profile: The EDEN mother-child cohort on the prenatal and early postnatal determinants of child health and development. *Int. J. Epidemiol.* 2015;dyv151.

25. Chen A, Pennell ML, Klebanoff MA, Rogan WJ, Longnecker MP. Maternal smoking during pregnancy in relation to child overweight: follow-up to age 8 years. *Int J Epidemiol.* 2006;35:121–30.
26. Durmus B, Kruithof CJ, Gillman MH, Willemsen SP, Hofman A, Raat H, et al. Parental smoking during pregnancy, early growth, and risk of obesity in preschool children: the Generation R Study. *Am. J. Clin. Nutr.* 2011;94:164–71.
27. Fieuws S, Verbeke G. Joint modelling of multivariate longitudinal profiles: pitfalls of the random-effects approach. *Statist. Med.* 2004;23:3093–104.
28. Cole TJ, Donaldson MDC, Ben-Shlomo Y. SITAR--a useful instrument for growth curve analysis. *Int J Epidemiol.* 2010;39:1558–66.
29. Willemsen SP, Eilers PHC, Steegers-Theunissen RPM, Lesaffre E. A multivariate Bayesian model for embryonic growth. *Statist. Med.* 2015;34:1351–65.
30. Zhang Z. Bayesian growth curve models with the generalized error distribution. *Journal of Applied Statistics.* 2013;40:1779–95.

TABLES

Table I. Prior Distributions Used for Hyperparameters involved in The Bayesian Model (rjags notation).

Parameters	Weight ($k = W$)	Height ($k = H$)
β_{0Ak}	$dnorm(1; 16)$	$dnorm(4; 16)$
β_{0Bk}	$dnorm(-5; 16)$	$dnorm(-4; 16)$
β_{0Ck}	$dnorm(1.5; 16)$	$dnorm(3; 16)$
β_{0Dk}	$dnorm(-5; 16)$	$dnorm(-5; 16)$
$\beta_{1(A-D)_k}$ to $\beta_{p_{cov(A-D)_k}}$	$dnorm(0; 16)$	$dnorm(0; 16)$
Ω^{-1}	$dwish(R^{-1}, 9)$	
Residual variance		
Simulation study		
$\sigma_{e_k}^{-2}$	$dgamma(0.0001; 0.0001)$	$dgamma(0.0001; 0.0001)$
Application		
δ_{0k}	$dnorm(0; 16)$	$dnorm(0; 16)$
δ_{1k}	$dnorm(0; 16)$	$dnorm(0; 16)$
δ_{2W}	$dnorm(0; 16)$	NA

Table II. True values, bias and probability coverage obtained from the Bayesian joint model of height and weight on 100 simulated datasets.

Parameter	True Value	Bias	Relative Bias	95% Coverage Probability	RMSE
A_W	1,000	-0.00101	-0.00101	93	0.01005
B_W	-5,000	0.00074	0.00015	99	0.00743
C_W	1.5	-0.00056	-0.00037	97	0.0056
D_W	-5,000	-0.00031	-0.00006	96	0.00311
A_H	4,000	-0.00021	-0.00005	96	0.00206
B_H	-4,000	-0.00025	-0.00006	98	0.00246
C_H	3,000	0.00069	0.00023	95	0.00691
D_H	-5,000	-0.00098	-0.0002	96	0.00981
σ_{eW}	0.4	0.00094	0.00234	91	0.00937
σ_{eH}	0.4	0.00043	0.00107	96	0.00427
$A_W A_W$	0.01444	0.00004	0.00281	96	0.00041
$B_W B_W$	0.04	0.00056	0.01395	98	0.00558
$C_W C_W$	0.05444	0.00157	0.02876	95	0.01566
$D_W D_W$	0.09556	0.00108	0.01126	95	0.01076
$A_H A_H$	0.00111	0.00002	0.02037	94	0.00023
$B_H B_H$	0.01111	0.00027	0.02464	90	0.00274
$C_H C_H$	0.01778	0.00023	0.013	95	0.00231
$D_H D_H$	0.06778	0.00186	0.02748	94	0.01862

$(A - D)_k(A - D)_k$ refer to the variance of the four parameters describing the trait k corresponding to the diagonal values of the variance-covariance matrix Ω_{kk} ; where $k = H$ for height and $k = W$ for weight.

Table III. Mean differences^a in height and weight model parameters between each maternal smoking category and the reference category (no maternal smoking) and 95% credibility intervals, Children 0-5 years old from the EDEN study.

Outcomes	A		B		C		D	
	Birth length		Growth Velocity		Spurt of Growth		Curvature Degree	
	Extrapolation of Birth		-Childhood-		-First Months-		-First Months-	
	β_A	CI 95%	β_B	CI 95%	β_C	CI 95%	β_D	CI 95%
Height								
Exclusively the 1 st trimester	-0.004	-0.012, 0.005	0.027	-0.002, 0.057	-0.021	-0.057, 0.015	0.072	0.004, 0.136
Throughout pregnancy	-0.015	-0.021, -0.008	0.021	-0.002, 0.043	0.005	-0.023, 0.033	0.039	-0.012, 0.092
Weight								
Exclusively the 1 st trimester	0.006	-0.028, 0.040	-0.001	-0.050, 0.048	0.014	-0.042, 0.067	-0.014	-0.097, 0.067
Throughout pregnancy	-0.037	-0.061, -0.012	0.034	0.000, 0.068	0.009	-0.032, 0.052	0.022	-0.035, 0.080

Abbreviations: CI, credibility interval

^aAdjusted for maternal education level, BMI and age at delivery, breastfeeding duration, recruitment center, child's gender

Table IV. Mean differences in body mass index (kg/m²), weight (g) and height (cm) between children exposed to maternal smoking during only the first trimester, children exposed throughout pregnancy and the reference category (children from non-smoking mothers during pregnancy).

Variable	Maternal Smoking			
	First Trimester Exclusively		Throughout Pregnancy	
	Posterior Coefficient	95% Credibility Interval	Posterior Coefficient	95% Credibility Interval
Difference of mean weight, g				
5 days	19	-77,119	-103	-176,-29
30 days (1 month)	20	-81,124	-72	-150,6
90 days (3 months)	29	-100,160	-29	-129,69
180 days (6 months)	45	-107,201	2	-115,116
270 days (9 months)	58	-114,231	18	-112,146
365 days (1 year)	69	-122,262	31	-112,175
730 days (2 years)	88	-155,328	90	-90,269
1095 days (3 years)	94	-194,384	158	-58,376
1470 days (4 years)	97	-255,449	231	-36,505
1825 days (5 years)	98	-323,520	301	-24,639
Difference of mean height, cm				
5 days	-0.15	-0.57,0.27	-0.71	-1.03,-0.38
30 days (1 month)	-0.02	-0.42,0.37	-0.60	-0.90,-0.29
90 days (3 months)	0.16	-0.24,0.55	-0.42	-0.73,-0.11
180 days (6 months)	0.23	-0.2,0.67	-0.29	-0.61,0.04
270 days (9 months)	0.20	-0.24,0.65	-0.24	-0.56,0.09
365 days (1 year)	0.14	-0.32,0.60	-0.22	-0.56,0.12
730 days (2 years)	-0.01	-0.58,0.57	-0.20	-0.63,0.24
1095 days (3 years)	0.04	-0.60,0.69	-0.12	-0.61,0.38
1470 days (4 years)	0.19	-0.54,0.94	0	-0.56,0.58
1825 days (5 years)	0.37	-0.48,1.24	0.13	-0.52,0.82
Difference of mean BMI, kg/m²				
5 days	0.14	-0.09,0.38	-0.05	-0.23,0.12
30 days (1 month)	0.08	-0.13,0.29	0.07	-0.09,0.24
90 days (3 months)	-0.01	-0.25,0.24	0.16	-0.04,0.36
180 days (6 months)	-0.02	-0.27,0.24	0.15	-0.04,0.36
270 days (9 months)	0.01	-0.24,0.27	0.15	-0.05,0.35
365 days (1 year)	0.06	-0.20,0.31	0.15	-0.04,0.35
730 days (2 years)	0.12	-0.12,0.36	0.19	0.02,0.37
1095 days (3 years)	0.09	-0.13,0.31	0.21	0.05,0.38
1470 days (4 years)	0.03	-0.19,0.25	0.22	0.05,0.39
1825 days (5 years)	-0.02	-0.25,0.21	0.21	0.03,0.40

FIGURES

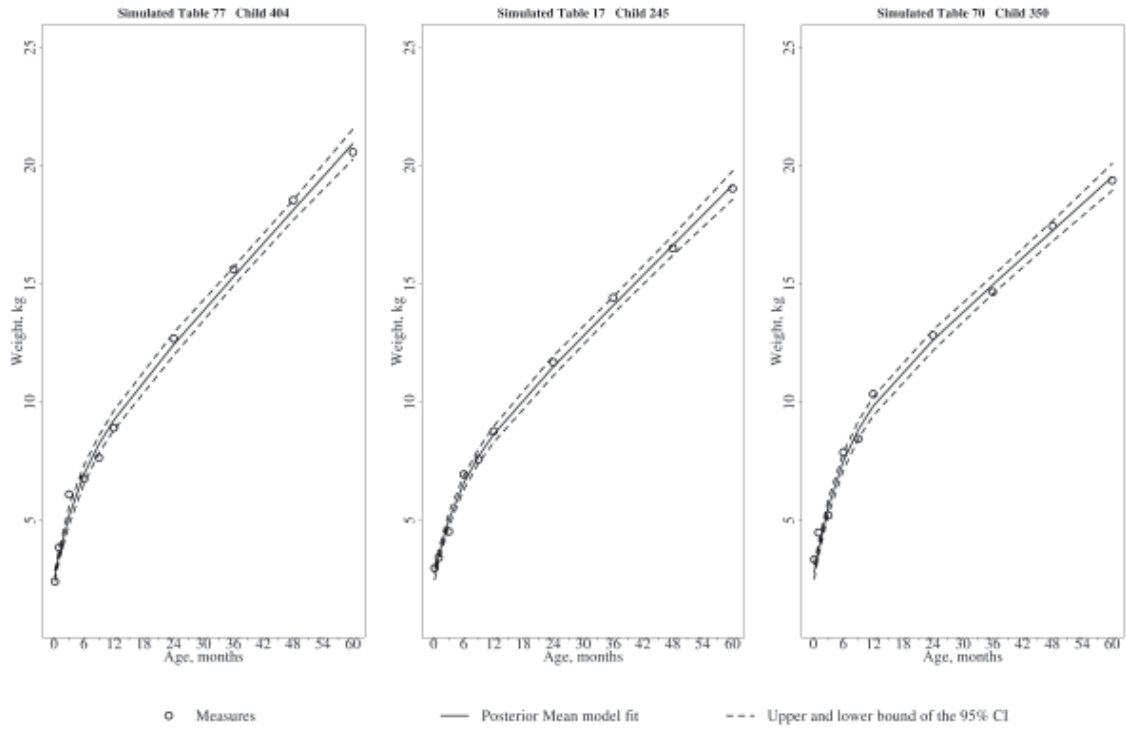


Figure I. Weight measures (simulated data) of randomly selected individuals (open circles), mean curves fitted by the joint Bayesian model and 95% credible intervals.

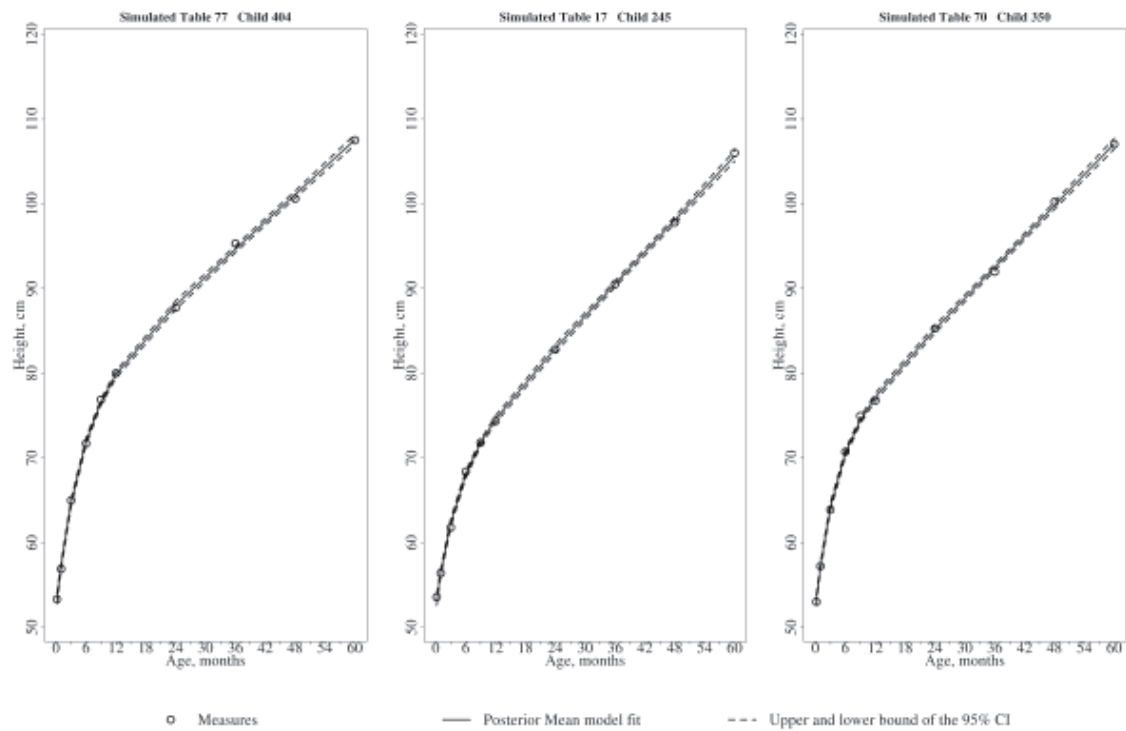


Figure II. Height measures (simulated data) of randomly selected individuals (open circles), mean curves fitted by the joint Bayesian model and 95% credible intervals.

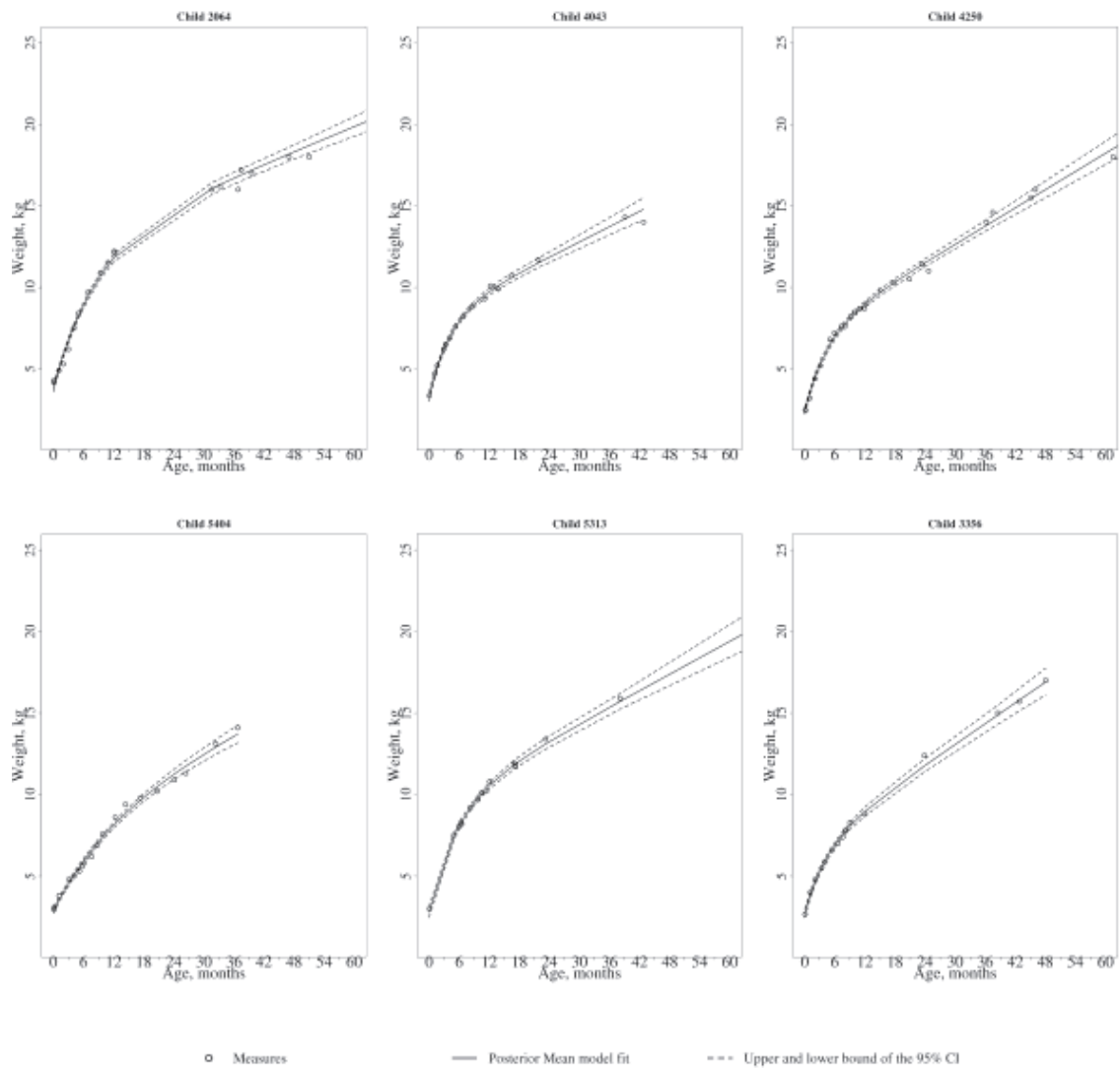


Figure III: Weight measures (open circles), mean curves fitted by the joint Bayesian model and 95% credible intervals of randomly selected children from the EDEN study.

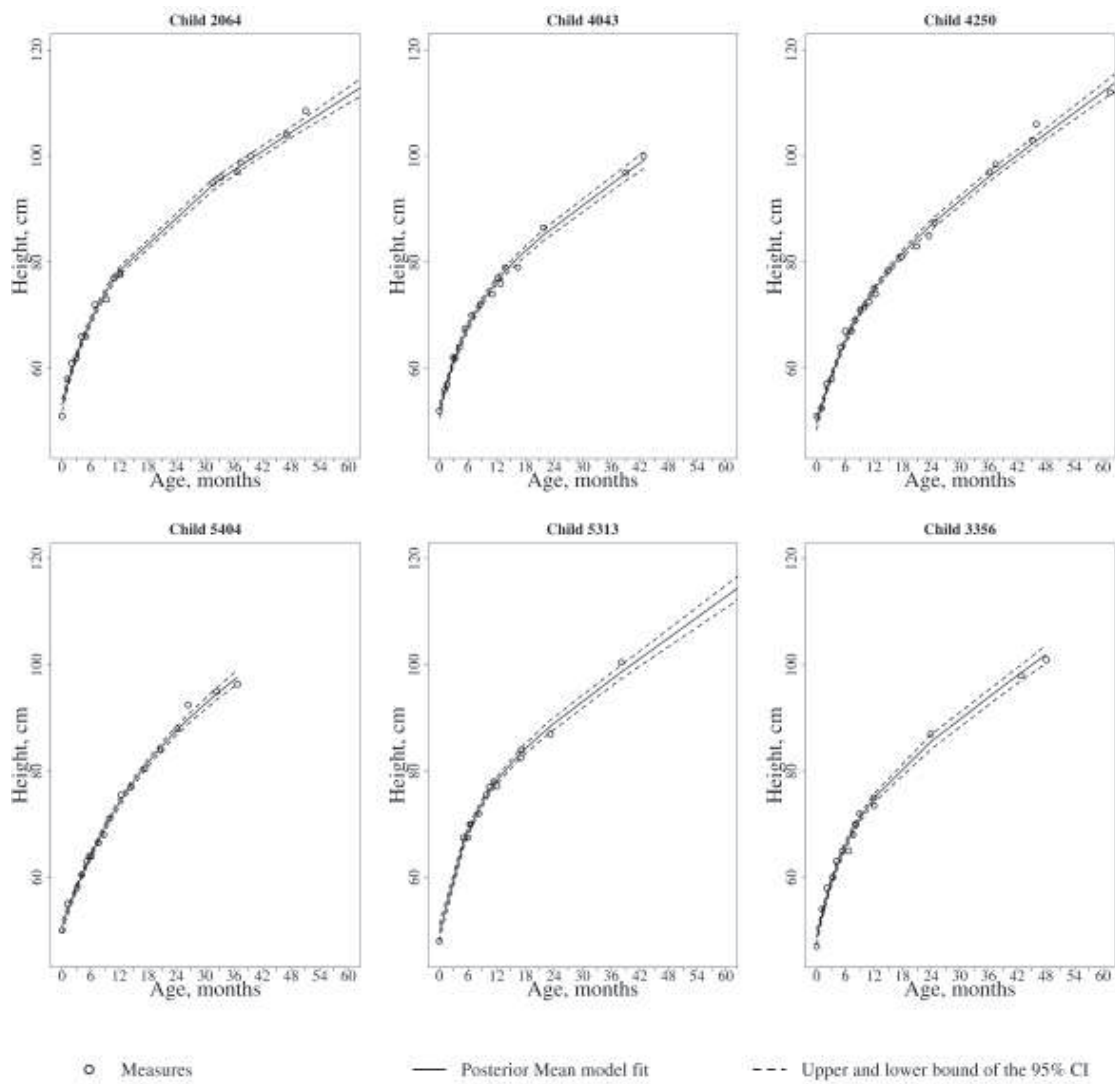


Figure IV. Height measures (open circles), mean curves fitted by the joint Bayesian model and 95% credible intervals of randomly selected children from the EDEN study.

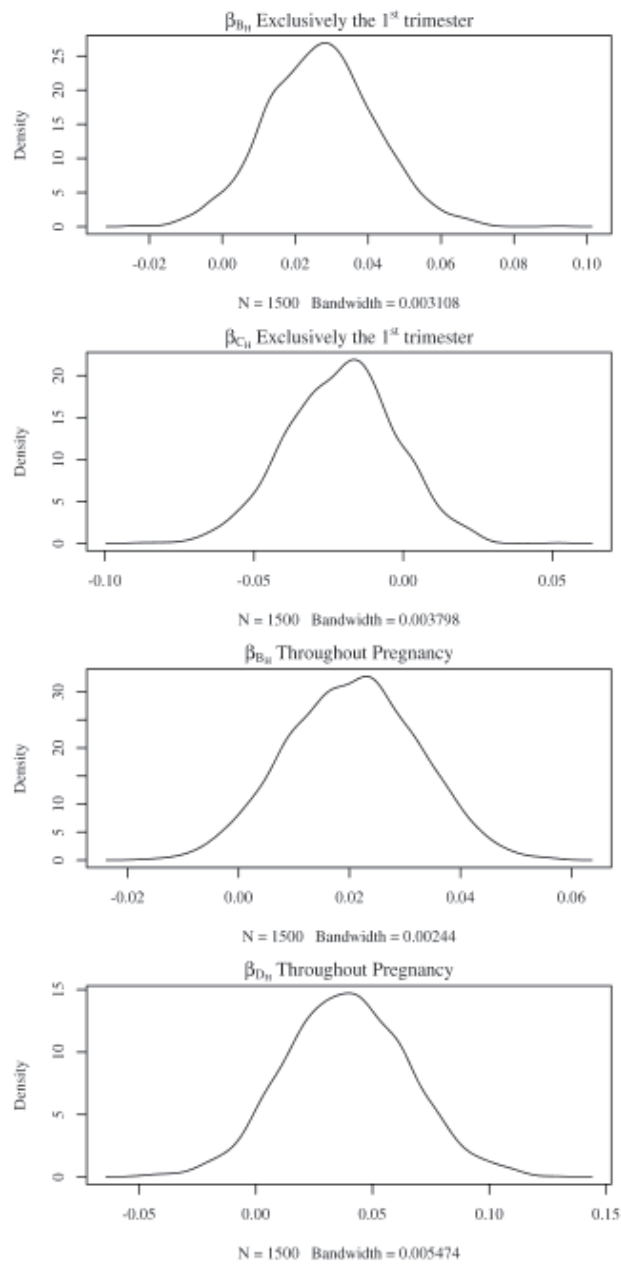


Figure V. Posterior density of regression coefficients associated with maternal smoking during pregnancy and selected height and weight parameters

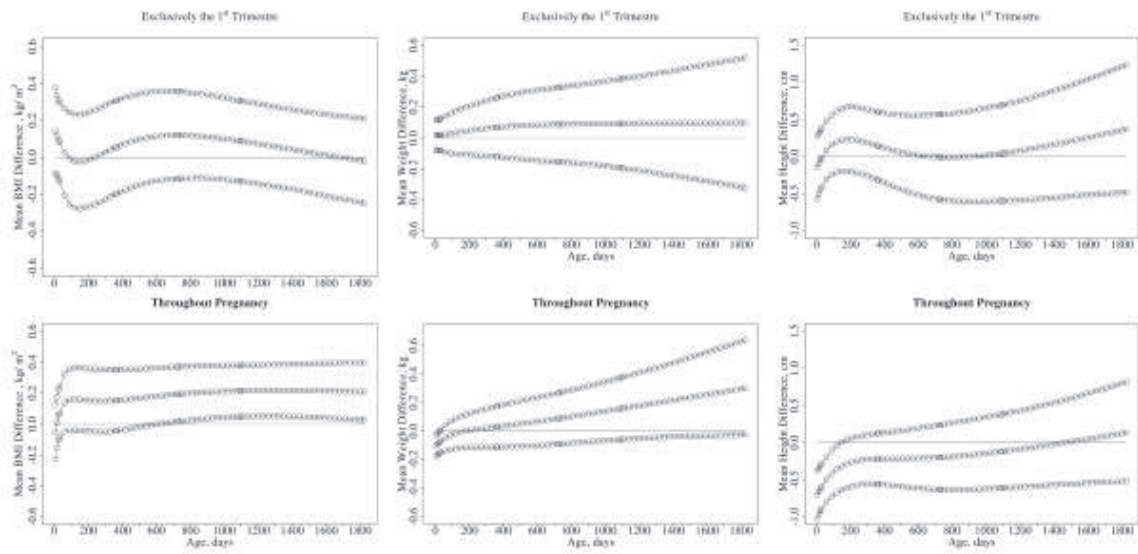


Figure VI. Mean differences and 95% credibility intervals of body mass index (kg/m^2), weight (kg) and height (cm) between children exposed to maternal smoking during only the first trimester and the reference category (children from non-smoking mothers during pregnancy) (upper part of the graph); between children exposed throughout pregnancy and the reference category (lower part of the graph) from birth to 1825 days .

SUPPLEMENTARY MATERIAL

Supplementary material I: Parametrisation of the Wishart distribution

$$\Omega_0 = \begin{matrix} & A_W & B_W & C_W & D_W & A_H & B_H & C_H & D_H \\ \begin{matrix} A_W \\ B_W \\ C_W \\ D_W \\ A_H \\ B_H \\ C_H \\ D_H \end{matrix} & \left(\begin{array}{cccccccc} 0.014 & 0.004 & 0.001 & 0.001 & 0.003 & 0.001 & -0.003 & 0.001 \\ 0.004 & 0.040 & -0.019 & 0.027 & 0.001 & 0.011 & 0.000 & 0.002 \\ 0.001 & -0.019 & 0.054 & -0.044 & 0.001 & -0.003 & 0.012 & -0.002 \\ 0.001 & 0.027 & -0.044 & 0.096 & 0.001 & 0.011 & -0.017 & 0.029 \\ 0.003 & 0.001 & 0.001 & 0.001 & 0.001 & 0.000 & 0.000 & -0.001 \\ 0.001 & 0.011 & -0.003 & 0.011 & 0.000 & 0.011 & -0.007 & 0.018 \\ -0.003 & 0.000 & 0.012 & -0.017 & 0.000 & -0.007 & 0.018 & -0.022 \\ 0.001 & 0.002 & -0.002 & 0.029 & -0.001 & 0.018 & -0.022 & 0.068 \end{array} \right) \end{matrix}$$

And

$$R^{-1} = \begin{matrix} & \left(\begin{array}{cccccccc} 0,13 & 0,04 & 0,01 & 0,01 & 0,03 & 0,02 & -0,03 & 0,01 \\ 0,04 & 0,36 & -0,17 & 0,24 & 0,01 & 0,10 & 0,00 & 0,02 \\ 0,01 & -0,17 & 0,49 & -0,40 & 0,01 & -0,03 & 0,11 & -0,02 \\ 0,01 & 0,24 & -0,40 & 0,86 & 0,01 & 0,10 & -0,15 & 0,26 \\ 0,03 & 0,01 & 0,01 & 0,01 & 0,01 & 0,00 & 0,00 & -0,01 \\ 0,01 & 0,10 & -0,03 & 0,10 & 0,00 & 0,10 & -0,06 & 0,16 \\ -0,03 & 0,00 & 0,11 & -0,15 & 0,00 & -0,06 & 0,16 & -0,20 \\ 0,01 & 0,02 & -0,02 & 0,26 & -0,01 & 0,16 & -0,20 & 0,61 \end{array} \right) \end{matrix}$$

Supplementary material II: Deviance Information Criterion (DIC) and Watanabe Akaike Information Criterion (WAIC).

Bivariate models with and without smoking covariates were compared based on the deviance information criterion (DIC) [22] defined as:

$$DIC = D(\bar{\theta}) + 2p_{DIC} = \bar{D} + p_{DIC}$$

where $D(\bar{\theta}) = -2 \sum_{ij} E\{\log[p(y_{ij}|\theta)]\}$ is the deviance function for parameter θ based on the closest stochastic parents of the observed data $y_{ij} = (y_{ijH}, y_{ijW})^T$ with $\bar{D} = E_{post}(D)$, $\bar{\theta} = E_{post}(\theta)$ and $p_{DIC} = \bar{D} - D(\bar{\theta})$ being the penalty term.

We also computed the Watanabe Akaike information criterion (WAIC) proposed more recently by Watanabe (2010) [23] and implemented in Bayesian software by Gelman et al. (2013).

$$WAIC = -2LPPD + 2p_{WAIC}$$

where $LP yD$ stands for the log pointwise predictive density equal to

$$LPPD = \sum_{ij} \log\{E_{post}[p(y_{ij}|\theta)]\}$$
 and the penalty term being expressed as $p_{WAIC} = \sum_{ij} Var_{post}[p(y_{ij}|\theta)]$.

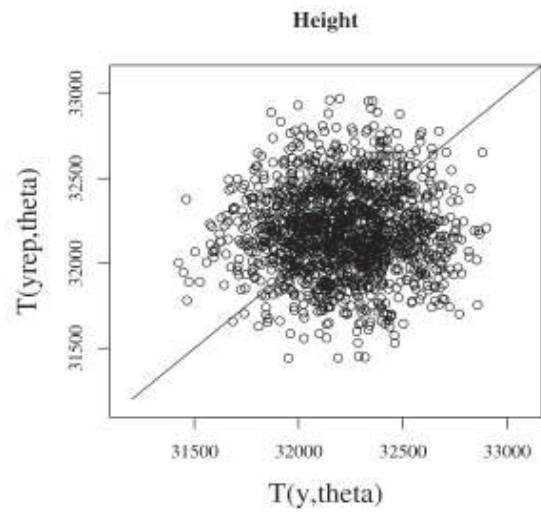
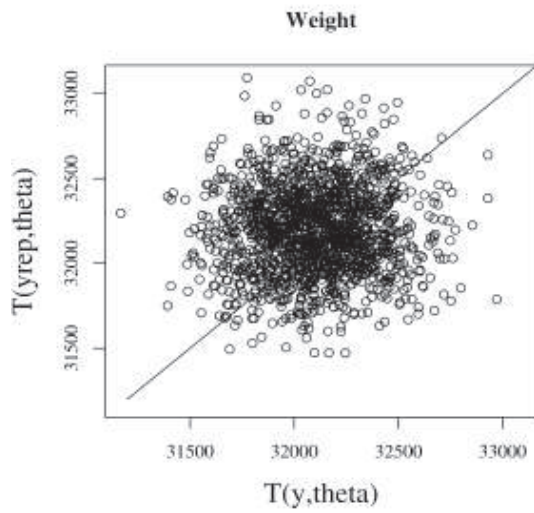
Both criteria aimed to quantify the out-of-sample predictive ability of models after correcting for the bias from evaluating predictive ability on the same data used to fit the models. Our preference is WAIC because it avoids some disadvantages of criteria based on point estimation $p(y|\theta = \bar{\theta})$, such as DIC.

Supplementary material III: True values, bias and probability coverage obtained from the Bayesian joint model of height and weight on 100 simulated datasets.

Parameter	True Value	Bias	Relative Bias	95% Coverage Probability	RMSE
A_W	1,000	-0.00101	-0.00101	93	0.01005
B_W	-5,000	0.00074	0.00015	99	0.00743
C_W	1.5	-0.00056	-0.00037	97	0.0056
D_W	-5,000	-0.00031	-0.00006	96	0.00311
A_H	4,000	-0.00021	-0.00005	96	0.00206
B_H	-4,000	-0.00025	-0.00006	98	0.00246
C_H	3,000	0.00069	0.00023	95	0.00691
D_H	-5,000	-0.00098	-0.0002	96	0.00981
σ_{eW}	0.4	0.00094	0.00234	91	0.00937
σ_{eH}	0.4	0.00043	0.00107	96	0.00427
$A_W A_W$	0.01444	0.00004	0.00281	96	0.00041
$A_W B_W$	0.00444	-0.00004	0.00897	94	0.0004
$A_W C_W$	0.00111	-0.00007	-0.06559	94	0.00073
$A_W D_W$	0.00111	0.00032	-0.2849	97	0.00317
$A_W A_H$	0.00333	0.00007	0.01989	93	0.00066
$A_W B_H$	0.00111	-0.00017	0.1511	92	0.00168
$A_W C_H$	-0.00333	0.00002	-0.00596	96	0.0002
$A_W D_H$	0.00111	-0.00045	0.40938	94	0.00455
$B_W B_W$	0.04	0.00056	0.01395	98	0.00558
$B_W C_W$	-0.01889	-0.00058	-0.03088	94	0.00583
$B_W D_W$	0.02667	0.00037	0.0138	93	0.00368
$B_W A_H$	0.00111	-0.00003	0.02553	96	0.00028
$B_W B_H$	0.01111	0.00021	0.01914	91	0.00213
$B_W C_H$	0,000	-0.00003	NA	91	0.00025
$B_W D_H$	0.00222	0.00018	0.08133	94	0.00181
$C_W C_W$	0.05444	0.00157	0.02876	95	0.01566
$C_W D_W$	-0.04444	-0.00088	-0.01976	92	0.00878
$C_W A_H$	0.00111	0.00004	0.03741	91	0.00042
$C_W B_H$	-0.00333	-0.00015	-0.0462	95	0.00154
$C_W C_H$	0.01222	0.0004	0.03301	93	0.00403
$C_W D_H$	-0.00222	-0.00025	-0.11475	95	0.00255
$D_W D_W$	0.09556	0.00108	0.01126	95	0.01076
$D_W A_H$	0.00111	-0.00004	0.03577	92	0.0004
$D_W B_H$	0.01111	0.00041	0.03695	91	0.00411
$D_W C_H$	-0.01667	-0.00078	-0.04672	93	0.00779
$D_W D_H$	0.02889	0.00171	0.05936	95	0.01715
$A_H A_H$	0.00111	0.00002	0.02037	94	0.00023
$A_H B_H$	0,000	-0.00004	NA	93	0.00036
$A_H C_H$	0,000	0.00002	NA	95	0.00016
$A_H D_H$	-0.00111	-0.00008	-0.07391	91	0.00082
$B_H B_H$	0.01111	0.00027	0.02464	90	0.00274
$B_H C_H$	-0.00667	-0.00008	-0.01135	98	0.00076
$B_H D_H$	0.01778	0.00057	0.0323	95	0.00574
$C_H C_H$	0.01778	0.00023	0.013	95	0.00231
$C_H D_H$	-0.02222	-0.00038	-0.01689	96	0.00375
$D_H D_H$	0.06778	0.00186	0.02748	94	0.01862

$(A - D)_k(A - D)_k$ refer to the variance of the four parameters describing the trait k , corresponding to the diagonal values of the variance-covariance matrix Ω_{kk} ; where $k = H$ for height and $k = W$ for weight. $(A - D)_W(A - D)_H$ refer to the covariance between the four parameters of the weight sub-model and those from the height one and correspond to the elements of the covariance matrix Ω_{WH} .

Supplementary material IV: Joint distribution of simulation draws of $T(y; \theta)$ and its replication $T(y^{\text{rep}}; \theta)$



Titre : Modélisation de la croissance postnatale des enfants pour l'étude de ses déterminants précoces

Mots clés : bayésien, cohorte mère-enfant, croissance postnatale, déterminants, tabagisme maternel, étude EDEN, équation de Jemss-Bayley, indice de masse corporelle, modèles non linéaires à effets mixtes

Résumé :

Contexte : La croissance dans les premières années de vie est associée à la santé et au risque de maladies à court et long terme. Elle est influencée par de nombreux facteurs dont certains, tels que le tabagisme maternel pendant la grossesse, sont susceptibles de l'altérer à la fois durant les périodes pré- et postnatale précoce. L'étude de la croissance postnatale nécessite une modélisation précise. Aucun consensus sur la méthode optimale n'existe à ce jour.

Objectifs : Etudier des approches de modélisation de la croissance du poids, de la taille et de l'indice de masse corporelle (IMC) de la naissance à 5 ans pour l'étude de ses déterminants précoces.

Méthodes : Les données de la cohorte mère-enfant EDEN, qui a inclus 2002 femmes enceintes dans les maternités de Nancy et Poitiers entre 2003 et 2006, ont été utilisées. Au total, 39177 mesures de poids (nombre de mesures médian par enfant : 22) et 30026 mesures de taille (nombre de mesures médian par enfant : 17) ont été collectées lors d'examens cliniques (à 1, 2 et 3 ans) et dans le carnet de santé. Le statut pondéral à 5 ans a été défini selon les seuils IOTF à partir de l'IMC. Les croissances du poids et de la taille ont été modélisées par des modèles non-linéaires à effets mixtes utilisant l'équation de Jemss-Bayley. Les associations entre exposition et croissance ont été étudiées par deux approches, en prenant l'exemple de l'exposition au tabagisme maternel pendant la grossesse. La première approche, fréquentiste, se décomposait en deux méthodes : i) méthode en 2 étapes : le poids et la taille ont été modélisés séparément sans covariables ; les associations entre les prédictions individuelles à différents âges issues de cette modélisation et le tabagisme maternel ont été analysées dans un second temps par régressions linéaires multiples ; ii) méthode en 1 étape : le tabagisme maternel et les covariables ont été inclus d'emblée dans la modélisation du poids et de la taille et les trajectoires d'IMC déduites.

Une seconde approche, bayésienne, a permis de modéliser conjointement le poids et la taille, de prédire les différences de trajectoires d'IMC en fonction du statut tabagique maternel et d'obtenir le degré d'incertitude sur celles-ci. Afin d'étudier l'amélioration de la qualité des estimateurs, les précisions des prédictions obtenues ont été comparées à celles obtenues à partir de modélisations séparées. Une analyse bayésienne exploratoire a permis l'étude de différences dans la co-évolution du poids et de la taille selon le statut pondéral à 5 ans, en modélisant la matrice de variances-covariances des effets aléatoires.

Résultats : Les approches fréquentiste et bayésienne donnaient des résultats concordants. Les enfants de femmes fumeuses tout au long de la grossesse tendaient à conserver, au cours des premières années, leur déficit de taille acquise à la naissance. Leur déficit de poids de naissance ne persistait pas au-delà des premiers mois. Ainsi, leur IMC surpassait, dès les premiers mois, l'IMC des enfants de femmes non-fumeuses. La méthode en 2 étapes était pertinente car la variance intra-sujet des prédictions était faible par rapport à la variance inter-sujets. En ce qui concerne la méthode bayésienne, l'incertitude sur les prédictions de poids et de taille était légèrement plus faible quand les profils de croissance étaient modélisés conjointement plutôt que séparément. L'analyse exploratoire montrait que, dans les tout premiers mois de vie, pour une même variation de la vitesse de croissance staturale, la variation de la vitesse de croissance pondérale était plus forte chez les enfants en surpoids ou obèses comparés à ceux de poids normal.

Conclusion : Les outils utilisés ont permis de caractériser finement les trajectoires de croissance postnatale. En permettant d'identifier les facteurs associés à des modifications de ces trajectoires très tôt dans l'enfance et de façon longitudinale, ils sont un moyen d'étudier précisément les déterminants de la croissance.

Title : Modeling postnatal growth of children to study its early determinants

Keywords : Bayesian, body mass index, determinants, EDEN study, Jenss-Bayley equation, maternal smoking, mother-child cohort, non-linear mixed effects models, postnatal growth.

Abstract :

Background: Growth in the first years of life is associated with both the short and long-term risk of diseases. Growth is influenced by many factors, among which some can affect both the early pre- and post-natal periods. This is the case of maternal smoking during pregnancy. The study of post-natal growth needs an accurate modeling. At present, there is no consensus on the optimal method.

Objectives: Modeling postnatal growth of weight, height and body mass index (BMI), from birth to 5 years, using different approaches to study its early determinants.

Methods: The work relies on data from the mother-child EDEN cohort, in which 2002 pregnant women were recruited between 2003 and 2006 in the maternities of Nancy and Poitiers. A total of 39,177 weight measurements (median by children: 22) and 30,026 height measurements (median by children: 17) has been collected during clinical visits (at 1, 2 and 3 years) and in the health booklet. The child's overweight status at 5 years was defined according to the IOTF classification based on their BMI. Weight and height growths have been modeled with non-linear mixed effects models using the Jenss-Bayley equation. Associations between exposure and growth have been studied with two approaches, considering maternal smoking during pregnancy as an example. The first approach, frequentist, included two methods: i) a two-steps method: the weight and height were modeled separately without covariates; associations between individual predictions at different ages and maternal smoking were assessed in a second step with multiple linear regressions; ii) one-step method: maternal smoking and other covariates were directly included in the weight and height modeling and BMI trajectories deduced.

This second Bayesian approach allowed to jointly model the weight and height, to predict the differences between BMI trajectories according to the maternal smoking status and to get uncertainty degrees. In order to study any improvement in the quality of the estimators, the precisions of predictions obtained were compared to those obtained with separate models of height and weight. An exploratory Bayesian analysis, allowed studying the differences in the co-evolution of weight and height according to the 5-years overweight status, by modeling the random effects variance-covariance matrix.

Results: Both the frequentist and the Bayesian approaches showed consistent results. The children from smoking mothers during the whole pregnancy tended to maintain their birth length deficit along their first years of life. The birth weight deficit did not persist after the first months. Thus, their BMI overpassed the BMI of children from non-smoking mothers as soon as the first months of life. The two-steps method was relevant because the within-subject variance of predictions was low compared to the between-subject variance. For the Bayesian method, the uncertainty on the predictions of weight and height was slightly lower when growth patterns were modeled jointly rather than separately. The exploratory analysis showed that, in the first months, for a similar variation of the statural growth velocity, the variation of the weight growth velocity was stronger in overweight or obese children compared to children with a normal weight.

Conclusion: We used tools that allowed the accurate characterization of the postnatal growth trajectories. They permit identification of factors associated with variations of these trajectories very early during childhood. Consequently, they proved helpful to study precisely the determinants of growth in early life.