

# Traitement des données sur la diversification alimentaire

Données mises en forme dans le cadre du projet **ANR-12-DSSA-0001** SOFI

Sophie Nicklaus, INRA CSGA, [sophie.nicklaus@inra.fr](mailto:sophie.nicklaus@inra.fr)

## Table des matières

0. Objectifs et définitions.....	1
I. Les aliments.....	2
Données initiales .....	2
Données disponibles .....	4
Procédure d'imputation .....	4
II. Regroupement des aliments.....	8
Construction des groupes d'aliments.....	8
Construction des variables d'introduction globale .....	9
III. Diversification alimentaire .....	9
Age de diversification .....	9
Age d'introduction des groupes d'aliments .....	10
IV. Pondération spécifique .....	12

## **0. Objectifs et définitions**

L'objectif de ce document est de décrire les différentes étapes de la construction des variables de diversification alimentaire.

1. Pour chaque aliment : état de la diversification à chaque point du suivi et calcul de l'âge d'introduction.
2. Regroupement des aliments : pour chaque groupe, état de la diversification et calcul de l'âge d'introduction.
3. Diversification alimentaire : définition, état à chaque point du suivi et calcul de l'âge de diversification.

Chacune de ces variables a été construite d'une part suivant la définition OMS, qui considère qu'un aliment goûté une fois est déclaré introduit, et d'autre part suivant la définition « régulière », qui considère qu'un aliment doit être consommé plusieurs fois pour être déclaré introduit.





# I. Les aliments

## Données initiales

Les données de diversification ont été recueillies auprès des mères par le biais d'un fréquentiel de consommation à compléter chaque mois du 3<sup>ème</sup> au 10<sup>ème</sup> mois de l'enfant. Nous nous intéressons ici à 2 parties du fréquentiel : l'une indiquant les fréquences de consommation d'une liste fermée d'aliments ou de groupes d'aliments et l'autre les fréquences d'ajout d'ingrédients, tel que le sel, dans l'alimentation de l'enfant.

Concernant les aliments, la question se présentait comme suit :

2. Indiquez la fréquence à laquelle votre bébé a consommé les aliments ou boissons suivants (seuls ou mélangés) en cochant la case appropriée à votre réponse : 0 = pas encore introduit ; 1 = une seule fois ; 2 = plusieurs fois ; 3 = souvent ; 4 = tous les jours ou presque.

		0	1	2	3	4
		entre 2 et 3 mois				
	Pommes de terre	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
	Haricots verts	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
	Carottes ou potiron	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
	Petits pois	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4

Ainsi, les 29 aliments suivants étaient présentés:

- Jus de fruits
- Autres boissons sucrées (tisanes pour bébé, sirops,...)
- Eau en bouteille (hors biberon de lait en poudre)
- Eau du robinet (hors biberon de lait en poudre)
- Lait de vache entier
- Lait de vache écrémé ou demi-écrémé
- Autres laits
- Céréales infantiles
- Pomme de terre
- Haricots verts
- Carottes ou potiron
- Petits pois
- Artichauts ou épinards
- Autres légumes
- Pâtes ou riz
- Pain
- Viande ou jambon
- Poisson
- Jaune d'œuf

- Blanc d'œuf
- Fromage
- Yaourts ou petits suisses
- Pommes ou poires
- Bananes
- Fraises ou framboises
- Pêches ou abricots
- Autres fruits (à partir du 7<sup>ème</sup> mois)
- Biscuits (à partir du 7<sup>ème</sup> mois)
- Autres desserts sucrés (à partir du 7<sup>ème</sup> mois)

Concernant les « ingrédients », la question était la suivante :

**3. Indiquez si vous avez ajouté ou utilisé dans l'alimentation de votre bébé les ingrédients ou aliments suivants en cochant la case appropriée à votre réponse : 0 = jamais ; 1 = parfois ; 2 = souvent ; 3 = toujours ou presque toujours ; 4 = ne sais pas.**

	jamaïs	parfois	souvent	toujours ou presque toujours	ne sais pas
	0	1	2	3	4
	<b>entre 2 et 3 mois</b>				
Du sucre	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
Du sel	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
Du beurre ou de la crème	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4

Les « ingrédients », au nombre de 9, étaient:

- Sucre
- Sel
- Beurre ou crème
- Huile ou margarine
- Aliments sans gluten
- Produits « bio »
- Légumes ou fruits écrasés ou en petits morceaux (non mixés)
- Viande en petits morceaux (non mixés) (à partir du 7<sup>ème</sup> mois)
- Aliments pour bébé du commerce (en dehors du lait et autre boisson)

Les variables étaient codées de la façon suivante : **nomaliment<sub>i</sub>m**, où **i** indiquait le mois du suivi et variait de 2 à 9. Il est important pour la suite de noter que la valeur 2 correspond au 3<sup>ème</sup> mois de l'enfant et ainsi de suite jusqu'à 9 qui correspond au 10<sup>ème</sup> mois de l'enfant.

## Données disponibles

Chaque aliment est représenté par 8 variables correspondant aux 8 mois du suivi, sauf les 3 aliments n'apparaissant qu'au 7<sup>ème</sup> mois qui ne sont représentés que par 4 variables. La variable Autres Laits ne sera pas traitée avec les autres variables car très peu d'enfants sont concernés.

Une variable mesurant la qualité de réponse aux questions sur les 28 aliments a été construite, elle varie entre 0 et 212, ces valeurs correspondant au nombre de valeurs manquantes de chaque individu. Voici sa distribution dans la population totale (n=18329) :

- 6037 individus ont 212 valeurs manquantes
- 5 individus ont entre 200 et 211 valeurs manquantes
- 721 individus ont entre 150 et 199 valeurs manquantes
- 1732 individus ont entre 100 et 149 valeurs manquantes
- 1325 individus ont entre 50 et 99 valeurs manquantes
- 3137 individus ont entre 1 et 49 valeurs manquantes
- 5372 individus ont 0 valeur manquante

Nous avons développé une procédure d'imputation afin de corriger au maximum ces valeurs manquantes. On considère que les individus en ayant plus de 149 ne portent pas assez d'information pour permettre une imputation correcte, ainsi la procédure décrite ci-après ne sera appliquée qu'aux individus ayant au plus 149 valeurs manquantes, soit à un total de 6194 individus.

Des variables comptant le nombre de valeurs manquantes pour chaque aliment et chaque individu ont également été construites (notées nvm par la suite).

## Procédure d'imputation

Pour ce qui suit, on s'intéresse uniquement aux valeurs encadrant directement les valeurs manquantes. De plus, les lettres X et Y correspondent aux modalités de réponse, c'est-à-dire 0 (pas encore introduit), 1 (une seule fois), 2 (plusieurs fois), 3 (souvent) ou 4 (tous les jours ou presque).

- Remplacement par la valeur précédant la valeur manquante

Si, pour un aliment donné, un individu a au plus 3 valeurs manquantes (non consécutives et non placées à 2 mois), alors on impute la/les valeur(s) manquante(s) par la valeur du mois précédent.

La liste suivante présente les cas possibles avec 1 ou 2 valeurs manquantes (les cas à 3 valeurs manquantes sont une combinaison de ces cas, cf. exemple en dernière ligne).

2m	3m	4m	5m	6m	7m	8m	9m
X	.	Y					
X	.	.	Y				
	X	.	Y				
	X	.	.	Y			
		X	.	Y			
		X	.	.	Y		
			X	.	Y		
			X	.	.		
				X	.	Y	
				X	.	Y	
					X	.	Y
					X	.	.
						X	.

---

(0 1 2 3 ou 4)	.	(0 1 2 3 ou 4)	(0 1 2 3 ou 4)	.	(0 1 2 3 ou 4)	.	(0 1 2 3 ou 4)
-------------------	---	-------------------	-------------------	---	-------------------	---	-------------------

- Imputation par la médiane, procédure pour un aliment donné :

1. Identification des cas de valeurs manquantes de façon longitudinale (n) parmi la population formée par les individus ayant entre 1 et 149 valeurs manquantes.

	3m	4m	5m	6m	7m
Cas valeur manquante	X	.	.	.	Y

2. Identification des individus ayant le même profil (N) parmi la population formée par les individus ayant entre 0 et 149 valeurs manquantes.

	3m	4m	5m	6m	7m
Calcul médiane	X	(0 1 2 3 ou 4)	(0 1 2 3 ou 4)	(0 1 2 3 ou 4)	Y

3. Si  $n < 10\% \text{ de } N$  et  $n_{vm} < 6$ , alors calcul de la médiane de la population N
4. Imputation des valeurs manquantes par la médiane

- Cas particuliers

Contrairement à la méthode développée ci-dessus, on s'intéresse ici au profil longitudinal complet des individus (i.e. à toutes les observations du mois 2 au mois 9). On distingue 2 cas particuliers.

○ Cas 1

<b>2m</b>	<b>3m</b>	<b>4m</b>	<b>5m</b>	<b>6m</b>	<b>7m</b>	<b>8m</b>	<b>9m</b>
.	.	.	.	0	0	0	0
.	.	.	.	.	0	0	0

Les valeurs manquantes signalées par «.» sont remplacées par 0.

○ Cas 2

<b>2m</b>	<b>3m</b>	<b>4m</b>	<b>5m</b>	<b>6m</b>	<b>7m</b>	<b>8m</b>	<b>9m</b>
0	0	0	0	.	.	.	.
0	0	0	.	.	.	.	.

Les valeurs manquantes sont remplacées par les valeurs médianes de la population d'individus ayant le même profil.

Après application de cette procédure, la distribution de la variable de qualité de réponse est la suivante (en italique : données non modifiées par l'imputation):

- *6037 individus ont 212 valeurs manquantes*
- *5 individus ont entre 200 et 211 valeurs manquantes*
- *721 individus ont entre 150 et 199 valeurs manquantes*
- *1043 individus ont entre 100 et 149 valeurs manquantes (1732 avant imp.)*
- *29 individus ont entre 50 et 99 valeurs manquantes (1325 avant imp.)*
- *1782 individus ont entre 1 et 49 valeurs manquantes (3137 avant imp.)*
- *8712 individus ont 0 valeur manquante (5372 avant imp.)*

Toute cette procédure est valable pour les variables représentant les aliments. En ce qui concerne les variables « ingrédients », les modalités de réponse sont différentes :

- 0 : jamais
- 1 : parfois
- 2 : souvent
- 3 : toujours ou presque toujours
- 4 : ne sais pas

La procédure doit donc être adaptée : les valeurs manquantes pouvant être imputées sont uniquement celles encadrées par des valeurs différentes de la modalité 4. En revanche, la variable de qualité de réponse est la même que celle utilisée pour les aliments.

Les variables après imputation des données manquantes selon la procédure décrite ci-dessus ont le suffixe \_IMP.

## **II. Regroupement des aliments**

Les variables suivantes ont été créées pour la population formée par les individus ayant entre 0 et 199 valeurs manquantes. L'ajout des individus ayant plus de 149 valeurs manquantes est décidé dans le but d'éviter de perdre de l'information.

Toutes les variables décrites ci-après ont été construites suivant les définitions OMS et régulière dont la différence de codage est explicitée ci-dessous.

On rappelle les modalités de réponse aux questions:

- 0 : pas encore introduit
- 1 : une seule fois
- 2 : plusieurs fois
- 3 : souvent
- 4 : tous les jours ou presque

Dans le cas de la définition OMS, la non-introduction correspond à la modalité 0 et l'introduction correspond aux modalités 1, 2, 3 ou 4. Pour la définition régulière, la non-introduction correspond aux modalités 0 ou 1 et l'introduction aux modalités 2, 3 ou 4.

### **Construction des groupes d'aliments**

Dans un premier temps, les variables regroupant les différents groupes d'aliments ont été créées pour chaque âge du suivi:

- Groupe Légumes (**DIV\_legumeimOMS** et **DIV\_legumeimREG**):
  - Haricot vert, carotte, petit pois, artichaut, autres légumes.
- Groupe Fruits (**DIV\_fruitimOMS** et **DIV\_fruitimREG**) :
  - Pomme, banane, fraise, pêche, autres fruits.
- Groupe Protéines animales\* (**DIV\_protanimimOMS** et **DIV\_protanimimREG**) :
  - Viande (**DIV\_viandeimOMS** et **DIV\_viandeimREG**), poisson (**DIV\_poissonimOMS** et **DIV\_poissonimREG**)
- Groupe Œufs (**DIV\_oeufsimOMS** et **DIV\_oeufsimREG**):
  - Jaune d'œuf, blanc d'œuf.
- Groupe Jus de fruit (**DIV\_jusfruitimOMS** et **DIV\_jusfruitimREG**) :
  - Jus de fruit.
- Groupe Autres boissons sucrées (**DIV\_boissucrimOMS** et **DIV\_boissucrimREG**):
  - Autres boissons sucrées.
- Groupe Eau (**DIV\_eauimOMS** et **DIV\_eauimREG**):
  - Eau en bouteille, eau du robinet.
- Groupe desserts (**DIV\_dessertimOMS** et **DIV\_dessertimREG**) :
  - Biscuits, autres desserts sucrés.
- Groupe Pomme de terre (**DIV\_pommeterreimOMS** et **DIV\_pommeterreimREG**):
  - Pomme de terre.
- Groupe Pain et pâtes (**DIV\_painpateimOMS** et **DIV\_painpateimREG**):
  - Pain, pâtes.
- Groupe Lait\* (**DIV\_laitimOMS** et **DIV\_laitimREG**):



- Lait entier (**DIV\_laitentier<sub>im</sub>OMS** et **DIV\_laitentier<sub>im</sub>REG**), lait écrémé (**DIV\_laitecreme<sub>im</sub>OMS** et **DIV\_laitecreme<sub>im</sub>REG**)
- Groupe Produits laitiers (**DIV\_prodlaitier<sub>im</sub>OMS** et **DIV\_prodlaitier<sub>im</sub>REG**):
  - Fromage, yaourts.
- Groupe Céréales infantiles (**DIV\_cerealeinf<sub>im</sub>OMS** et **DIV\_cerealeinf<sub>im</sub>REG**):
  - Céréales infantiles.

\*Cas où nous avons étudié l'évolution de certains aliments à la fois individuellement et au sein d'un groupe. Par exemple, les données du groupes Protéines animales sont étudiées en parallèle des données de l'aliment viande et de l'aliment poisson. C'est pourquoi ces aliments ont aussi été recodés de la même manière que le groupe Protéines animales.

### Construction des variables d'introduction globale

Dans un second temps, des variables d'introduction globale ont été construites selon 2 cas de figure pour chaque définition et pour chaque âge du suivi (OMS et régulière) :

- 1 Prise en compte de tous les groupes excepté l'eau (**DIV\_ConsoAlim\_SE<sub>im</sub>OMS**, **DIV\_ConsoAlim\_SE<sub>im</sub>REG**)
- 2 Prise en compte de tous les groupes excepté l'eau et les autres boissons sucrées (**DIV\_ConsoAlim\_SE\_SBSUC<sub>im</sub>OMS**, **DIV\_ConsoAlim\_SE\_SBSUC<sub>im</sub>REG**)

Au mois *i*, pour *i* variant de 2 à 9, les variables construites dans les 2 paragraphes précédents valent :

- 0 si AUCUN des aliments du groupe/groupes n'a été introduit
- 1 si AU MOINS UN des aliments du groupe/groupes a été introduit

A la fin de ces étapes, 264 variables représentant les groupes d'aliments et 32 variables représentant l'introduction globale ont été construites.

## **III. Diversification alimentaire**

### Age de diversification

Nous avons construit les variables âge de diversification, indiquant l'âge auquel a eu lieu la première introduction d'au moins un groupe d'aliments, de la façon suivante : les valeurs allant de 2m à 9m représentent les variables **DIV\_ConsoAlim** construites précédemment.

2m	3m	4m	5m	6m	7m	8m	9m	DIV_AgelnitDiv (mois)
1								3
0	1							4
0	0	1						5
0	0	0	1					6
0	0	0	0	1				7

0	0	0	0	0	1				8
0	0	0	0	0	0	1			9
0	0	0	0	0	0	0	1		10
0	0	0	0	0	0	0	0		ND*

\* Non diversifié.

Ces variables sont au nombre de 4 et sont notées : DIV\_ageInitDiv\_OMS\_SE, DIV\_ageInitDiv\_REG\_SE, DIV\_ageInitDiv\_OMS\_SE\_SBSUC, DIV\_ageInitDiv\_REG\_SE\_SBSUC.

Le tableau ci-dessous indique les données disponibles pour les 4 variables « âge de diversification » avec et sans imputation.

	Sans imputation	Avec imputation
DIV_ageInitDiv_OMS_SE	9874	11 350
DIV_ageInitDiv_REG_SE	9693	11 257
DIV_ageInitDiv_OMS_SE_SBSUC	9823	11 326
DIV_ageInitDiv_REG_SE_SBSUC	9657	11 237

## Age d'introduction des groupes d'aliments

Ces variables sont construites de la même façon que les âges de diversification à la différence que l'on s'intéresse à la première introduction d'au moins un aliment du groupe considéré. Pour un groupe donné, le calcul de l'âge d'introduction est fait à partir des variables construites dans le paragraphe « Construction des groupes d'aliments ».

Exemple : pour le groupe légume, l'âge est calculé sur la base des variables **DIV\_legume** comme suit :

2m	3m	4m	5m	6m	7m	8m	9m	DIV_ageIntroLeg (mois)
1								3
0	1							4
0	0	1						5
0	0	0	1					6
0	0	0	0	1				7
0	0	0	0	0	1			8
0	0	0	0	0	0	1		9
0	0	0	0	0	0	0	1	10
0	0	0	0	0	0	0	0	NI*

\*NI=groupe non introduit à 10 mois révolus

On a finalement un total de 34 variables notées :

- DIV\_ageIntroLegOMS
- DIV\_ageIntroLegREG
- DIV\_ageIntroFruitOMS

- DIV\_ageIntroFruitREG
- DIV\_ageIntroProtAniOMS
- DIV\_ageIntroProtAniREG
- DIV\_ageIntroViandeOMS
- DIV\_ageIntroViandeREG
- DIV\_ageIntroPoissonOMS
- DIV\_ageIntroPoissonREG
- DIV\_ageIntroOeufOMS
- DIV\_ageIntroOeufREG
- DIV\_ageIntroJusfruitOMS
- DIV\_ageIntroJusfruitREG
- DIV\_ageIntroBoissucrOMS
- DIV\_ageIntroBoissucrREG
- DIV\_ageIntroEauOMS
- DIV\_ageIntroEauREG
- DIV\_ageIntroDessertOMS
- DIV\_ageIntroDessertREG
- DIV\_ageIntroProdLaitOMS
- DIV\_ageIntroProdLaitREG
- DIV\_ageIntroPommeterreOMS
- DIV\_ageIntroPommeterreREG
- DIV\_ageIntroPainpateOMS
- DIV\_ageIntroPainpateREG
- DIV\_ageIntroLaitOMS
- DIV\_ageIntroLaitREG
- DIV\_ageIntroLaitentierOMS
- DIV\_ageIntroLaitentierREG
- DIV\_ageIntroLaitecremeOMS
- DIV\_ageIntroLaitecremeREG
- DIV\_ageIntroCerealeinfOMS
- DIV\_ageIntroCerealeinfREG

#### ***IV. Pondération spécifique***

L'auto-questionnaire Sicomiam, relatif à l'alimentation du nourrisson sur la période de 3 à 10 mois, fait l'objet d'une pondération spécifique (m02f1\_pond). Les détails du calcul de cette pondération, basé sur l'ensemble des répondants au questionnaire, est disponible sur Pandora dans la partie « 2-10 mois : alimentation ».