

## **Note Méthodologique concernant le codage des professions et secteurs d'activités des mères de la cohorte ELFE**

Equipe Santé Travail Environnement, ISPED Centre Inserm U897, Université de Bordeaux

Responsable projet : Ghislaine Bouvier

Référent pour le codage : Xavier Schwall

### **-INTRODUCTION**

L'objectif de ce travail était, pour chacune des mères de la cohorte, d'attribuer des codes permettant de définir et de catégoriser les professions ainsi que les secteurs d'activité dans lesquels elles ont travaillé durant la grossesse de l'enfant Elfe. Ce travail de codage est une étape importante préliminaire pour pouvoir étudier par la suite l'impact des expositions professionnelles des mères sur la santé et/ou le développement de l'enfant Elfe.

Les nomenclatures utilisées pour ce travail sont :

pour le codage des secteurs d'activité :

- NAF (Nomenclature d'Activités Française) 2003
- CITI (Classification Internationale Type des Industries) 1975 (en anglais ISIC 75)

pour le codage des professions :

- PCS (Professions et Catégories socio-professionnelles) 2003
- CITP (Classification Internationale Type des Professions) 1968 (en anglais ISCO 68)

Nous avons choisi ces nomenclatures et plus particulièrement d'en utiliser deux (une récente et une plus ancienne) à chaque fois, car elles sont complémentaires et permettent ainsi de traiter un plus grand nombre d'informations même lorsque celle-ci sont imprécises, mais aussi car elles permettent d'utiliser des matrices emploi/exposition, françaises ou étrangères.

La principale difficulté pour traiter le fichier de données est sa taille et le nombre important de variables qu'il convient de prendre en compte pour coder le plus précisément possible les professions et les secteurs d'activité.

On retrouve également parfois pour un même sujet des informations discordantes selon les variables / les questionnaires (maternité / deux mois).

Afin de pouvoir traiter le fichier de façon homogène et fiable il a donc fallu hiérarchiser les variables afin de privilégier une information plutôt qu'une autre, particulièrement lorsque l'on est en présence de discordances.

#### -PRINCIPE GENERAL DE CODAGE

Le fichier de travail, qui comprend 18 312 lignes correspondant aux sujets de la cohorte ELFE, a été traité de façon manuelle, aucun logiciel de codage n'a été utilisé.

Les outils utilisés ont été les nomenclatures « papier », ainsi que l'outil internet « CAPS » (<https://ssl3.isped.u-bordeaux2.fr/CAPS-FR/>), outil développé par l'InVS et le Centre de Recherche et de Développement en Informatique Médicale (CREDIM) de l'Université de Bordeaux, qui permet une recherche informatique au sein de 3 des 4 nomenclatures utilisées.

Dans un premier temps nous avons séparé les informations du fichier de travail en deux :

-les informations recueillies à la maternité lors de l'accouchement

-les informations recueillies lors de l'enquête téléphonique deux mois plus tard.

Après concertation avec l'équipe Elfe, nous avons décidé, dans le cas de discordance entre ces deux sources d'information, de privilégier celles obtenues lors de l'enquête à deux mois.

En effet, les données obtenues au cours de l'enquête à deux mois étaient a priori de meilleure qualité de par les conditions de l'entretien (mère à son domicile vs à l'hôpital après l'accouchement) et la personne qui mène l'entretien (enquêteur formé vs personnel hospitalier).

Par ailleurs, la structure des nomenclatures utilisées nous permet en cas d'information incomplète ou incertaine de proposer un codage partiel, donnant des indications sur le statut voire sur le type d'emploi du salarié.

#### -METHODE : LES VARIABLES UTILISEES

##### **- Les variables du questionnaire « maternité »**

\* NAISGEM : Cette variable permet l'identification de jumeaux et donc la suppression de 288 « doublons » concernant l'emploi des mères (NAISGEM : oui/ N=576)

\*PROFMAUCUN : Cette variable nous signale lorsque aucun emploi n'a été exercé au cours de la vie du sujet (PROFMAUCUN : oui/N=2701, pourtant pour 1688 de ces sujets il a tout de même été possible d'attribuer un codage emploi)

\*PROFESSP : dernière profession déclarée (N=12 462)

\*SITUG : Information sur la situation du sujet lors de la grossesse

### **-Les variables du questionnaire à deux mois**

\*ACTIVANTE3 : variable qui nous informe si le sujet a déjà exercé une profession au cours de sa vie (ACTIVANTE3 : NON/N=514 mais 75 ont tout de même été codés)

\*SITUA 3 et SITUAE 3 : nous informent sur les activités exercées (salarié, étudiant, chômage....)

\*PROFI5 3, PROFI6 3 : libellé de l'emploi exercé (ou du dernier connu)

\*PROFI7 3 et PROFI7 3B : statut du salarié (cadre, ouvrier qualifié, non qualifié...)

\*VALPROFESS : permet de vérifier l'exactitude de l'emploi exercé (VALPROFESS : NON /N=194)

\*PROFESSG : Si VALPROFESS=NON alors prendre en compte l'emploi énoncé dans PROFESSG

\*TACHEG : descriptif détaillé des tâches effectuées au cours de l'emploi (en clair)

\*ACTPEG : activité principale de l'entreprise (en clair)

\*DELAICT et DELAICE : nous informent sur le délai écoulé entre la fin du dernier emploi et le début de la grossesse.

D'autres variables ont été prises en compte mais non de façon systématique, uniquement lorsque l'emploi le nécessitait.

Par exemple pour un exploitant agricole à son compte il a fallu regarder des variables telles que :

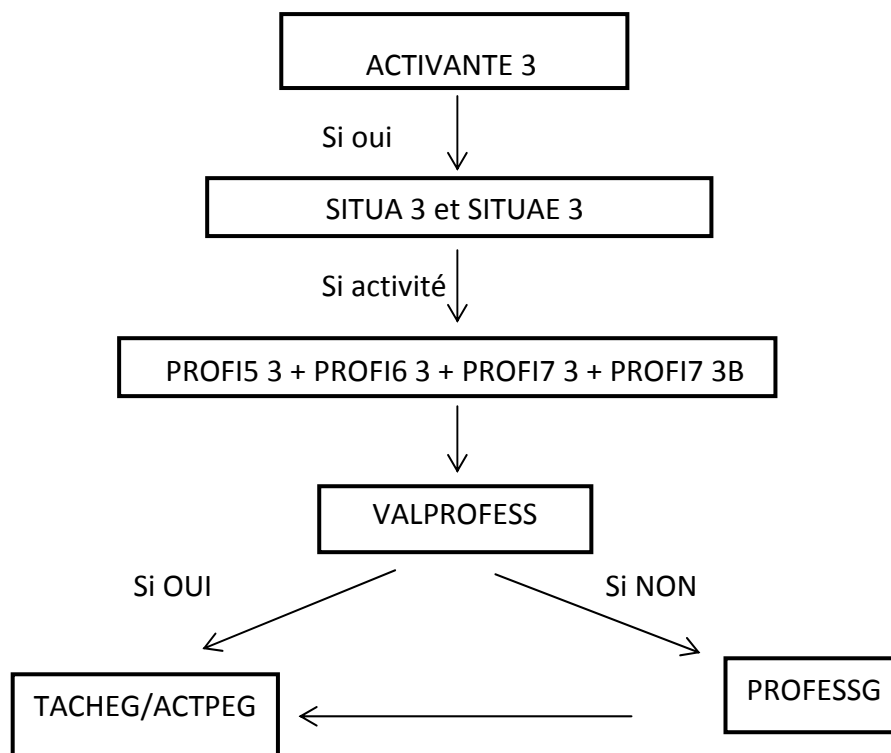
-SALARIES 3 : pour connaître la taille de l'entreprise (personnel salarié)

-SUPH 3 et SUPA 3 : pour connaître la superficie de l'exploitation

-OPA 3 : pour connaître l'orientation principale de l'exploitation agricole.

D'une façon générale les autres variables ne sont donc pertinentes que dans des cas bien précis, et bien souvent assez peu fréquemment renseignées.

## SCHEMA RECAPITULATIF :



Il s'agit d'une logique/hiérarchie de codage, mais il a fallu bien entendu regarder l'ensemble des variables pour s'assurer de la concordance des informations et de la pertinence du codage obtenu.

Les variables ACTPEG (pour le secteur d'activité) et TACHEG (pour l'emploi) priment :

- si les informations concordent avec celles apportées par les variables PROFIS/6/7 : pas de problème pour coder
- si cela ne concorde pas, les variables TACHEG et ACTPEG priment dans le choix du code. Ce choix se justifie par le fait que le contenu de ces variables est du déclaratif en clair par les personnes et donc moins sujet aux erreurs que les variables reposant sur un menu déroulant par exemple.

### -PARTICULARITES DE CODAGE DES NOMENCLATURES

Pour chacune des nomenclatures utilisées le principe de codage est le même : constitué de 4 à 5 digits (4 à 5 chiffres ou lettres), le degré de précision du code va de la gauche (code le moins précis) vers la droite (code le plus précis). Si l'information est insuffisante et que l'on

ne peut attribuer un code complet, le ou les chiffres en question sont remplacés par des « + ».

#### - NAF 2003

Les codes de cette nomenclature sont constitués de 3 chiffres et d'une lettre majuscule.

Par exemple le secteur de l'industrie correspond au chiffre 1, s'il s'agit de la seule information que l'on a on obtient un code : 1+++.

S'il s'agit d'une industrie alimentaire sans autre précision on code 15++.

S'il s'agit du secteur de la production de boissons on code 159+.

Si l'information est complète, par exemple production de cidre alors le code est complet : 159J.

La NAF possède une autre caractéristique : celle de pouvoir utiliser un autre type de code dans le cas où plusieurs activités sont regroupées dans la même entreprise. Il s'agit du code NACE qui lui se compose de 4 chiffres. Ce type de codage n'est réellement utile que dans quelques cas précis (N=68 pour le fichier des emplois des mères dans Elfe.)

Par exemple le codage NACE 1591 regroupe les codes 159A (production d'eau de vie naturelle) et 159B (fabrication de spiritueux).

Dans ce fichier le code NACE le plus utilisé est le code 3530 qui correspond à la construction aéronautique et spatiale. Ce code regroupe les codes 353A (construction de moteurs pour avions), 353B (construction de cellules d'avions) et 353C (construction de lanceurs et engins spatiaux). Il était en effet très rare pour ce secteur d'avoir un degré d'information suffisant pour avoir un codage NAF complet.

#### -CITI 1975

Les codes de cette nomenclature sont composés de 4 chiffres et fonctionnent de la même façon que la NAF en ce qui concerne la précision.

Ainsi le code 1 regroupe le secteur Agriculture, chasse, sylviculture et pêche (1+++), 11 les activités agricoles (11++), 12 la sylviculture (12++), 13 la pêche (13++), etc...

#### -PCS 2003

Les codes de cette nomenclature se composent de 3 chiffres et d'une lettre minuscule, toujours sur le même principe.

Par exemple : 6 représente la catégorie « ouvrier » (6+++), 61 à 65 les ouvriers qualifiés (61++à 65++), 67 à 68 les ouvriers non qualifiés (67++à 68++).

On peut ensuite être plus précis pour chacune des catégories :

Exemple : 621 pour bâtiment et travaux publics (621+)

621a pour les chefs d'équipe

621 b pour les ouvriers qualifiés du travail du béton

etc...

-CITP 1968

Cette nomenclature fonctionne avec 5 digits et va du plus large au plus précis de la même façon que les autres de la gauche vers la droite.

Par exemple :

les codes commençant par un 4 regroupent l'ensemble des personnels du commerce. Les digits suivants nous donnant une information plus précise :

-40 pour les directeurs (40020 pour le commerce de gros, 40030 pour le commerce de détail)

-42 pour les chefs de ventes et acheteurs (42120 chef de vente commerce de gros, 42130 chef des ventes commerce de détail....)

-43 les agents commerciaux et voyageurs de commerce: 431 agents et inspecteurs commerciaux, 432 voyageurs de commerce, et dans un dernier degré de précision par exemple 43120 agent commercial technicien, 43130 inspecteur commercial technicien.

De la même façon on retrouvera :

-45120 pour les vendeurs au détail 45130 pour les vendeurs de gros, 45150 pour les démonstrateurs, ... etc.

-les agents d'assurance (44120), les agents immobiliers (44130) et les agents commerciaux (43120), ... etc.

#### -CONTRÔLE ET HOMOGENEITE DES DONNEES

Le codage dans chacune des nomenclatures s'est fait manuellement, sans outil informatique. Une fois toutes les codages effectués, divers tris ont été effectués (par code mais aussi par type de variable) afin de vérifier que pour un même emploi (et/ou secteur d'activité) on avait bien attribué les même codes.

Par ailleurs il s'est avéré que certaines mères déclaraient plusieurs emplois et il a donc fallu créer de nouvelles variables afin de pouvoir coder au maximum trois emplois par mère. Au total une mère occupe trois emplois, et 13 mères occupent deux emplois.

Enfin, pour faciliter la lecture de ce fichier, une variable supplémentaire a été créée, pour chacun des quatre codages effectués (NAF, PCS, CITI, CITP). Ces variables intitulées NAF2003QUALIT, CITI1975 QUALIT, CITP1968 QUALIT, 2003 QUALIT, donnent une information sur le degré de précision du codage.

Pour illustrer ceci voici un tableau récapitulatif de la précision obtenue dans le fichier emploi des mères de la cohorte Elfe :

	NAF	CITI	CITP	PCS	NAF 2	CITI2	CITP2	PCS2	NAF3	CITI3	CITP3	PCS3
code complet:1	12655	13945	14666	15072	10	11	13	14	1	1	1	1
code partiel:2	2158	1185	1704	1389	2	1	1	0	0	0	0	0
infos absente ou insuffisante:3	2022	1774	534	443	2	2	0	0	0	0	0	0
à revoir:4	8	8	8	8	0	0	0	0	0	0	0	0
aucun emploi déclaré:5	1112	1112	1112	1112	0	0	0	0	0	0	0	0
jumeaux à enlever, doublons:6	288	288	288	288	0	0	0	0	0	0	0	0
code NACE utilisé:10	69											