



Apurement de l'enquête AMP sans frontières

Marine Bondon et Géraldine Charrance

Octobre 2023

Apurement de l'enquête AMP sans frontières

Marine Bondon et Géraldine Charrance

Octobre 2023

Table des matières

Les bases issues de ce travail d'apurement.....	2
1. Suppression des observations issus des tests	4
2. Vérification des filtres et contrôle de cohérence.....	4
3. Création de variables de type par données	5
4. Suppression des champs ouverts sensibles	5
5. Recodage des modalités « Vous ne souhaitez pas répondre » et « Vous ne savez pas »	6
6. Recodage des variables « Pays ».....	7
7. Recodage des variables mentionnant des centres et agences d'AMP	9
8. Recodage des variables [A0_1_X] (X allant de 1 à 15)	11
9. Corrections de variables pour améliorer la cohérence.....	11
10. Création de la base Quadriptyques.....	13
11. Harmonisation des libellés des centres	16
12. Réintégration des variables issues de la base quadriptyques à la base individus .	16
13. Recodage de champs ouverts	18
14. Création de variables relatives au questionnaire	20
15. Assignation de labels et formats aux variables des bases Individus & Quadriptyques	22

Les bases issues de ce travail d'apurement

Base Individus - compte 9697 observations et 1517 variables.

D:\CloudINED\AMP\16 - Apurement\Bases\Bases Individus\Base finale

Au format SAS : ampsf_individus.sas7bdat

à utiliser avec le programme de formats AMPsf_Individus_formats.sas (disponible dans le même répertoire).

Au format CSV (codes) : AMPsf_Individus_codes.csv

Au format CSV (libellés) : AMPsf_Individus_libelles.csv

Base Champs ouverts

D:\CloudINED\AMP\16 - Apurement\Bases\Bases Individus\Base finale

Au format xlsx : AMPsf_champs_ouverts

Base Quadriptyques - compte 12 308 observations et 25 variables.

D:\CloudINED\AMP\16 - Apurement\Bases\Bases Quadriptyques

Au format SAS : ampsf_quadriptyques.sas7bdat

à utiliser avec le programme de formats AMPsf_quadriptyques_formats.sas (disponible dans le même répertoire).

Au format CSV (codes) : AMPsf_Quadriptyques_codes.csv

Au format CSV (libellés) : AMPsf_Quadriptyques_libelles.csv

Les tris à plat des bases sont disponibles dans le même répertoire que les bases.

NB : précision de Géraldine Charrance par mail du 25 octobre 2023 sur l'ordre des programmes

le programme 1b ne modifie en rien les bases. Il s'agit d'une étape de contrôle qui met en lumière les choses à corriger.

Pour produire la base apurée : j'ai fait tourner le programme 1b après avoir fait tourner les 31 premières lignes du programme 1a.

J'ai ensuite écrit toutes les lignes de corrections dans le programme 1a (35 à 450), puis pour vérifier que mon code permettait de corriger tous les retours en arrière, j'ai refait tourner sur la nouvelle base le programme 1b.

Pour répondre plus pratiquement à ta question : Pour passer de la base brute à la base apurée, il faut donc faire tourner les programmes comme suit :

() programmes SAS 1a - 2,*

() programmes R 3 - 4 - 5*

() programmes SAS 6a - 6b*

1. Suppression des observations issus des tests

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 1a_der_var_remplie_corr_retours_arriere.sas

Lignes : 29 à 31

Type d'opération : suppression d'observations

Afin de donner la possibilité à l'équipe d'accéder au questionnaire en ligne (en amont et durant le terrain) pour notamment réaliser des tests, un lien spécifique a été créé par le prestataire. Les observations résultant des connexions réalisées via ce lien sont identifiables dans la base de données.

En pratique, nous avons donc exclu toutes les observations pour lesquelles la variable [SRC] vaut 909090.

2. Vérification des filtres et contrôle de cohérence

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programmes :

- 1b_ctrl_filtres.sas pour les contrôles de cohérence
- 1a_der_var_remplie_corr_retours_arriere.sas pour les corrections (lignes 35 à 450).

Type d'opération : Corrections de variables existantes (les variables d'origine sont modifiées, aucune nouvelle variable n'est créée).

Après examen de la base, nous avons identifié des incohérences dans les données, résultant des retours en arrière opérés par les répondants. Pour illustrer le propos, plaçons-nous dans un cas réel : dans la base, nous avons une personne pour qui A1= « Vous ne souhaitez pas répondre » et A1_1 = « FtM ». Dans le questionnaire, il est prévu que seules les personnes ayant répondu « Transgenre » à la question A1 se voient poser la question A1_1. Il y a donc une incohérence dans les données, très certainement due au fait que la personne a renseigné « Transgenre » en première intention à la question A1, puis a répondu à la question A1_1, puis est revenue en arrière, et a changé sa réponse à la question A1 en « Vous ne souhaitez pas répondre ». La personne ne passe donc plus dans la question A1_1, et sa réponse à cette question n'est pas supprimée (alors qu'elle devrait l'être).

L'opération de contrôles de cohérence implique donc de reprogrammer chaque filtre du questionnaire. Dès lors que l'on observe une réponse à une question X pour un individu ne vérifiant pas les conditions d'entrée, i.e. le filtre de ladite question, on supprime sa réponse (pour assurer une cohérence des réponses au sein de chaque questionnaire, mais également de veiller au respect du droit de rectification des données des individus).

3. Création de variables de type paradonnées

a. Variable [der_var_remplie]

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 1a_der_var_remplie_corr_retours_arriere.sas

Lignes : 452 à 1005

Type d'opération : création de variable

La base de données destinée aux analyses comporte des questionnaires complets et incomplets. Aucun questionnaire n'est écarté car jugé trop peu complété. C'est pour cette raison que nous avons souhaité identifier pour chaque questionnaire débuté le moment de l'abandon (si abandon il y a), afin de pouvoir éventuellement filtrer les observations selon le moment de l'abandon.

La variable [der_var_remplie] créée renseigne pour chaque questionnaire la dernière question répondue par l'enquêté. Elle vaut « 00 Quest non débuté » si le répondant n'a répondu à aucune question.

b. Variable [module_der_var]

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 1a_der_var_remplie_corr_retours_arriere.sas

Lignes : 1013 et 1014

Type d'opération : création de variable

A partir de la variable [der_var_remplie], nous avons créé la variable [module_der_var] permettant d'avoir une information plus synthétique sur le moment de l'abandon, puisqu'ici, on enregistre le module durant lequel l'abandon a eu lieu.

c. Variable [statut]

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 1a_der_var_remplie_corr_retours_arriere.sas

Lignes : 1010 et 1012

Type d'opération : création de variable

A partir de la variable [der_var_remplie], nous avons construit la variable [statut] permettant de distinguer les questionnaires non commencés, les questionnaires partiels et les questionnaires complets. Pour cela, nous considérons comme complet les questionnaires pour lesquels la dernière question complétée est E9 ou E10.

4. Suppression des champs ouverts sensibles

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 2_apurement_divers.sas

Lignes : 17 à 40

Type d'opération : Suppression de variables

Informations complémentaires : liste des variables écartées disponible dans le fichier excel D:\CloudINED\AMP\16 - Apurement\AMPsf_champs_ouverts.xlsx.

Les réponses aux questions ouvertes des modules B, C, D et E ont été dissociées du reste des réponses, car elles comportent dans un grand nombre de cas des éléments trop identifiants. Dissocier les réponses permet de sécuriser les données.

Certains champs ouverts ont été conservés dans la base :

- Tous les champs ouverts du module A
- Les champs ouverts relatifs aux centres et agences d'AMP, banques de sperme, d'ovocytes ou d'embryons (B8_X_Centre, B8_X_Centrebis, C23A_DER, C23B_DER_OTH, C36A_DER, C36B_DER_OTH, C49_DER).

Ces variables feront l'objet ensuite de recodages/créations de variables homogénéisées.

Les variables écartées sont disponibles dans un fichier Excel. Compte tenu de la longueur de certaines réponses, l'import de ces variables dans des logiciels d'analyses statistiques tels que SAS ou R peut provoquer des troncatures.

5. Recodage des modalités « Vous ne souhaitez pas répondre » et « Vous ne savez pas »

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 2_apurement_divers.sas

Lignes : 44 à 1249

Type d'opération : Recodage de variables existantes (les variables d'origine sont modifiées, aucune nouvelle variable n'est créée).

Dans le cas des questions à choix multiple, on dispose dans la base d'autant de variables que de modalités proposées (y compris les modalités « vous ne souhaitez pas répondre » et « vous ne savez pas »).

Cependant, cette configuration ne nous semble pas optimale pour l'analyse des réponses. D'une part, parce qu'elle augmente le nombre de variables à mobiliser pour l'étude d'une question, d'autre part, car le fait que le répondant n'ait pas souhaité ou n'ait pas su répondre à la question doit être visible dans chaque variable relative à ladite question. Sans cette opération, on ne peut distinguer un « non » d'un « ne sait pas » sur une modalité, ce qui nous semble être une source de confusion dans l'analyse.

Une réintégration des modalités « vous ne souhaitez pas répondre » ou « vous ne savez pas » à la variable initiale a été réalisée sur les champs ouverts (pour lesquels

une case à cocher était proposée). Cela permet de ramener l'analyse d'une question à une variable au lieu de deux.

6. Recodage des variables « Pays »

L'objectif de ces opérations de recodage est de fournir une seule et même nomenclature pour toutes les variables « pays » afin de simplifier la mobilisation de ces variables.

Chaque variable « pays » repose sur la nomenclature des pays ISO 3166-1 alpha-3 en trois lettres. Deux codes supplémentaires ont été créés :

- FOM : France d'outre-mer (DROM et COM)
- TRF : Territoire de la république française (Métropole, DROM et COM) ;

Pour certaines variables, il est possible de distinguer FRA (France métropolitaine) et FOM (France d'Outre-mer). Dans le cas contraire, le code TRF sera mobilisé.

a. Variable [A0_3_pays]

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 2_apurement_divers.sas

Lignes : 1257 à 1275

Type d'opération : Recodage de variable, une variable « corrigée » a été créée [A0_3_pays_rec]

Informations complémentaires : les règles de recodage sont explicitées dans le fichier excel D:\CloudINED\AMP\16 - Apurement\AMPsf_Recode_A0_3.xlsx.

Le pays du centre d'AMP par lequel l'enquêté a eu connaissance de l'enquête a été recueilli en clair. Une opération d'homogénéisation des réponses a été réalisée et un code pays selon la nomenclature ISO 3166-1 alpha-3 a été affecté à chaque réponse.

Pour les centres en France, afin de distinguer « France métropolitaine » et « France d'outre-mer », nous avons également mobilisé la variable A0_3_ville (ville du centre d'AMP par lequel l'enquêté a eu connaissance de l'enquête). Dans le cas où la ville et le nom du centre sont manquants, nous avons fait l'hypothèse que si la personne réside en métropole, alors le centre est également en France métropolitaine.

b. Variable [A14_4]

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 2_apurement_divers.sas

Lignes : 1277 à 1283

Type d'opération : Recodage de variable (correction intégrée directement à la variable d'origine)

La variable A14_4 était déjà enregistrée sous la forme ISO 3166-1 alpha-3. Cependant, pour deux pays, le code associé était erroné. Le code associé au Kosovo, anciennement « KOS » devient « XXK », et le code associé au Soudan du sud, anciennement « SDS » devient « SSD ».

c. Variables [A15_1_XXX]

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 2_apurement_divers.sas

Lignes : 1285 à 1498

Type d'opération : Recodage de variable (les variables créées sont A15_1_X (X allant de 1 à 5, 5 étant le nombre maximal de pays cités à la question A15_1 par un répondant) ; les variables d'origine ne sont pas conservées dans la base d'exploitation

Dans la base originale, la question A15_1 (Dans quel(s) pays l'enquêté (et son conjoint éventuel) ont déjà vécu pendant au moins 6 mois) se traduit par autant de variables que de pays proposés dans la nomenclature (198 pays + la modalité « Vous ne souhaitez pas répondre »). Afin de faciliter l'exploitation, nous avons choisi de modifier l'approche. Au vu de la structure des réponses (plus d'un tiers des personnes ont renseigné un seul pays), nous avons préféré l'approche « premier pays étranger mentionné », « deuxième pays étranger mentionné ». Pour les personnes n'ayant pas souhaité répondre à cette question, seule la première variable sera renseignée, avec le code « Ne souhaite pas répondre ».

Tout comme précédemment, les variables sont conformes à la nomenclature ISO 3166-1 alpha-3.

d. Variables [B8_X_PAYS] (X de 1 à 30)

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 2_apurement_divers.sas

Lignes : 1501 à 1514

Type d'opération : Recodage de variable (les corrections ont été intégrées directement aux variables originales, aucune variable n'a été créée)

Les modalités des variables B8_X_PAYS sont conformes à la nomenclature ISO 3166-1 alpha-3, hormis les deux cas cités précédemment (Kosovo, anciennement « KOS » devenu « XXK », et Soudan du sud, anciennement « SDS » devenu « SSD »).

e. Variable [PAYS_DER]

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 2_apurement_divers.sas

Lignes : 1517 à 1523

Type d'opération : Recodage de variable (les corrections ont été intégrées directement à la variable originale, aucune variable n'a été créée)

Les modalités de la variable sont conformes à la nomenclature ISO 3166-1 alpha-3, hormis les deux cas cités précédemment (Kosovo, anciennement « KOS » devenu « XXK », et Soudan du sud, anciennement « SDS » devenu « SSD »).

f. Variables [B8_4_X_PAYS] (X de 1 à 30)

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 2_apurement_divers.sas

Lignes : 1525 à 1538

Type d'opération : Recodage de variable (les corrections ont été intégrées directement aux variables originales, aucune variable n'a été créée)

Les modalités des variables B8_4_X_PAYS sont conformes à la nomenclature ISO 3166-1 alpha-3, hormis :

- les deux cas cités précédemment (Kosovo, anciennement « KOS » devenu « XXK », et Soudan du sud, anciennement « SDS » devenu « SSD »).
- Le code associé à la France : la distinction « France métropolitaine » et « France d'outremer » n'est pas demandée au niveau pays. On doit donc utiliser le code « TRF ».

g. Variables [C22_DER / C35_DER / C48_DER]

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 2_apurement_divers.sas

Lignes : 1540 à 1552

Type d'opération : Recodage de variable (les corrections ont été intégrées directement aux variables originales, aucune variable n'a été créée)

Mêmes opérations que celles réalisées sur les variables B8_4_X_pays (Kosovo, Soudan du Sud, et France entière).

h. Variables [D3_X] (X de 1 à 12)

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 2_apurement_divers.sas

Lignes : 1554 à 1567

Type d'opération : Recodage de variable (les corrections ont été intégrées directement aux variables originales, aucune variable n'a été créée)

Mêmes opérations que celles réalisées sur les variables B8_4_X_pays (Kosovo, Soudan du Sud, et France entière).

7. Recodage des variables mentionnant des centres et agences d'AMP

Afin de faciliter l'exploitation des données, un travail d'harmonisation des dénominations des centres/agences ou banques (de sperme, d'ovocytes ou

d'embryons), recueillies en clair, a été réalisé. En plus de l'harmonisation, nous avons intégré dans la dénomination la ville dans lequel le centre se situe.

a. Variable [A0_3_nom]

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 2_apurement_divers.sas

Lignes : 1569 à 1605

Type d'opération : Recodage de variable, la nouvelle variable est [A0_3_nom_rec]

Informations complémentaires : fichier excel ayant permis de réaliser les recodages : D:\CloudINED\AMP\16 - Apurement\AMPsf_Recode_A0_3.xlsx

b. Question « C23_DER »

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 2_apurement_divers.sas

Lignes : 1607 à 1642

Type d'opération : Recodage de variable, la nouvelle variable est [C23_DER_REC].

Informations complémentaires : Fichier excel sur lequel s'appuient les recodages : D:\CloudINED\AMP\16 - Apurement\AMPsf_Recode_C23_DER.xlsx

Pour créer la nouvelle variable C23_DER_rec, nous sommes appuyés sur les variables C23A_DER (banque de sperme, recueillie en clair, si C22_DER différent de Danemark et Etats-Unis), C23B_DER (liste pré-établie de banques de sperme au Danemark et aux Etats-Unis) et C23B_DER_OTH (possibilité de renseigner une autre banque de sperme si non présente dans la liste).

A l'occasion de ce recodage, nous avons constaté que des corrections devaient également être réalisées sur la variable C22_DER. Une nouvelle variable corrigée C22_DER_REC a également été créée à cette étape.

c. Question « C36_DER »

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 2_apurement_divers.sas

Lignes : 1645 à 1681

Type d'opération : Recodage de variable, la nouvelle variable est [C36_DER_REC].

Informations complémentaires : Fichier excel sur lequel s'appuient les recodages : D:\CloudINED\AMP\16 - Apurement\AMPsf_Recode_C36_DER.xlsx

Le recueil de l'information sur les banques d'ovocytes est fait sur le même modèle que celui sur les banques de sperme : la C36A_DER (banque d'ovocytes, recueillie en clair, si C35_DER différent de Danemark et Etats-Unis), C36B_DER (liste pré-établie de banques d'ovocytes au Danemark et aux Etats-Unis) et C23B_DER_OTH (possibilité de renseigner une autre banque de sperme si non présente dans la liste).

d. Question « C49_DER »

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 2_apurement_divers.sas

Lignes : 1683 à 1700

Type d'opération : Recodage de variable, la nouvelle variable est [C49_DER_REC].

L'harmonisation des réponses à la question C49_DER ne nécessite pas la fusion de plusieurs variables, il s'agit seulement d'une homogénéisation des écritures. L'opération est directement réalisée dans le programme SAS.

8. Recodage des variables [A0_1_X] (X allant de 1 à 15)

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 2_apurement_divers.sas

Lignes : 1701 à 1821

Type d'opération : Recodage de variables, les variables initiales ont été écrasées ou supprimées. A l'issue de cette opération, les réponses à la question A0_1 repose sur les variables A0_1_1 à A0_1_3.

Comme exposé précédemment, les questions à choix multiples se traduisent dans la base de données originale par autant de variables qu'il existe de modalités de réponse, ce qui rend l'exploitation de la question complexe. Cependant, lorsque les répondants ont majoritairement choisi une seule modalité, l'approche « première modalité choisie », deuxième modalité choisie » (comme présentée pour la question A15_1) semble plus judicieuse. Les variables relatives à la question A0_1 ont donc été retravaillées dans ce sens. La variable A0_1_1 comporte désormais la première association citée par l'enquêté, A0_1_2 comporte la seconde association citée et A0_1_3 comporte la troisième association citée, selon l'ordre de présentation des associations dans le questionnaire.

9. Corrections de variables pour améliorer la cohérence

a. Variables [Support_X] (X de 1 à 3)

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 2_apurement_divers.sas

Lignes : 1824 à 1831

Type d'opération : Renommage de variables

Afin de faciliter l'exploitation, nous avons renommé les variables renseignant le support mobilisé pour le remplissage du questionnaire.

Ancien nom	Nouveau nom
Support_1	Support_desktop
Support_2	Support_tablet
Support_3	Support_mobile

b. Variables [B13_2] et [B14_2_X] (X de 1 à 5)

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 2_apurement_divers.sas

Lignes : 1832 à 1866

Type d'opération : Recodages de variables. De nouvelles variables ont été créés à cette occasion, nom identique à la variable initiale + suffixe « _rec »

Pour les variables [B13_2] et [B14_2_1] à [B14_2_5], la prolongation du terrain a conduit à ajouter deux modalités de réponse : 2022 et 2023, rendant la nomenclature des réponses incohérente.

Nous avons donc retravaillé cette nomenclature sur ces 6 variables et créé des variables recodées (nom de la variable initiale suffixée par « _rec ») :

Libellé de la modalité	Code de la variable initiale	Code de la variable recodée
2023	23	1
2022	22	2
2021	1	3
2020	2	4
2019	3	5
2018	4	6
2017	5	7
2016	6	8
Au moins une fois entre 2005 et 2015	7	9
Au moine une fois avant 2005	8	10
Vous ne souhaitez pas répondre	10	88
Vous ne savez pas	9	99

c. Variables [B15_2_X_Y]

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes SAS.

Programme : 2_apurement_divers.sas

Lignes : 1870 à 1897

Type d'opération : Renommage de variables

La prolongation du terrain de l'enquête a nécessité l'ajout des modalités 2022 et 2023 à certaines variables, ce qui a nécessité certaines fois, de retravailler les variables.

Les variables B15_2_X_Y portent sur les années d'adhésion aux associations auxquelles l'enquêté a pu adhérer pour se faire aider dans ses démarches. Au début de l'enquête, la question comportait 9 modalités (+ les deux modalités « vous ne souhaitez pas répondre » et « vous ne savez pas »). Avec la prolongation, deux modalités ont été ajoutées : 2022 et 2023. S'agissant d'une question à choix multiple, il existe autant de variables que de modalités proposées. Compte tenu de la dénomination des variables relatives aux modalités initiales de la variable, il a

été difficile pour GIDE de maintenir une cohérence avec l'ajout des nouvelles modalités.

Dans le tableau ci-dessous est explicitée la façon dont les variables ont été renommées pour rendre le tout cohérent :

	Avant modifications	Après modification
2023	B15_2_X_23	B15_2_X_23
2022	B15_2_X_22	B15_2_X_22
2021	B15_2_X_1	B15_2_X_21
2020	B15_2_X_2	B15_2_X_20
2019	B15_2_X_3	B15_2_X_19
2018	B15_2_X_4	B15_2_X_18
2017	B15_2_X_5	B15_2_X_17
2016	B15_2_X_6	B15_2_X_16
Au moins une fois entre 2005 et 2015	B15_2_X_7	B15_2_X_0515
Au moins une fois avant 2005	B15_2_X_8	B15_2_X_av05
Vous avez toujours été adhérente	B15_2_X_9	B15_2_X_tjrs

Ces opérations ont été réalisées pour les 15 associations proposés (X variant donc de 1 à 15).

10. Création de la base Quadriptyques

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes R

Programme : 3_Création_base_quadriptyques.R

Type d'opération : Création d'une nouvelle base de données (autre unité d'observations)

Les recours à l'AMP étant décrits dans la base des réponses au niveau individuel, nous avons constitué une base au niveau quadriptyque pour faciliter l'analyse des parcours. L'unité d'observation de cette base n'est plus le répondant mais le quadriptyque (1 ligne = 1 quadriptyque).

A ce stade, de probables incohérences liées à des retours en arrière qui n'avaient pas pu être identifiés avec la base au niveau individu sont apparues : pour 7 répondants, on dispose d'une valeur pour B8_4_X_pays (pays de résidence au moment du quadriptyque X) alors qu'aucun quadriptyque X n'est renseigné. Ces lignes de quadriptyques ont donc été supprimées.

La création de cette base a donné lieu à la construction de nouvelles variables propres aux quadriptyques.

a. Ordre chronologique des quadriptyques [ordre_chrono]

Cette variable vise à reconstituer l'ordre chronologique des quadriptyques, qui étaient initialement ordonnés dans la base de niveau Individus par ordre de déclaration. Selon les configurations, il n'a pas toujours été possible de construire la variable [ordre chrono] :

- On dispose des variables B8_3_X¹ (X allant de 1 à 30) qui indiquent l'ordre chronologique des quadriptyques. Ces variables sont selon les cas :
 - o renseignées par le répondant, lorsqu'il y a une ambiguïté sur l'ordre chronologique de son parcours (parcours à plusieurs quadriptyques dont plusieurs réalisés la même année ou avec au moins un quadriptyque dont l'année n'est pas spécifiée (refus, ne sait pas)). On lui demande alors d'ordonner ses quadriptyques ;
 - o renseignées automatiquement quand il n'y a pas d'ambiguïté quant à la chronologie du parcours.
- On ne dispose pas des variables B8_3_X car le répondant a cessé de répondre entre la saisie des quadriptyques et le passage dans la question B8_3_X (ou son remplissage automatique). Il y a alors deux possibilités :
 - o On peut reconstituer l'ordre chronologique lorsqu'il n'y a qu'un seul quadriptyque renseigné ou plusieurs mais tous réalisés des années distinctes (et également toujours renseignées) ;
 - o On ne peut pas reconstituer l'ordre chronologique lorsque le parcours comprend plusieurs quadriptyques dont plusieurs ont été réalisés la même année/période où il y a au moins un quadriptyque dont l'année n'est pas spécifiée (refus, ne sait pas, non réponse). La variable [ordre_chrono] est alors laissée vide pour l'ensemble des quadriptyques du répondant.

b. Ordre déclaratif des quadriptyques [ordre_decla]

Cette variable indique pour chaque quadriptyque le rang auquel il a été renseigné. Elle correspond, dans la base au niveau individus, à la valeur X de chaque variable B8_X renseignée (B8_X_methode, B8_X_centre, etc.).

c. Appartenance du quadriptyque au champ de l'enquête [champ_quadri]

Code	Libellé	Construction
0	Quadriptyque hors champ	Si le pays de résidence lors de la réalisation du quadriptyque

¹ Attention à l'utilisation contre-intuitive de cette variable : B8_3_X = 2 signifie que le quadriptyque rempli en 2^e (le recours identifié par le numéro 2) est chronologiquement le x^{ème} quadriptyque (et pas que le quadriptyque x est en 2^e position).

		(B8_4_PAYS) est renseigné mais n'est pas la France
1	Quadriptyque dans le champ	Si le pays de résidence lors de la réalisation du quadriptyque est la France, et a) le pays de réalisation du quadriptyque (B8_PAYS_rec) est renseigné mais n'est pas la France ou b) la méthode réalisée (B8_METHODE) est une GPA/PPA ou une insémination artisanale
99	Appartenance au champ indéterminée	Si le pays de résidence lors de la réalisation du quadriptyque (B8_4_PAYS) est manquant ou que le pays de réalisation du quadriptyque (B8_PAYS_rec) vaut Refus ou Ne sait pas

d. Disponibilité des informations clés du quadriptyque [quadri_comp]

Indicatrice du niveau d'informations fournies dans le quadriptyque : dispose-t-on au moins des 4 informations « clés » (méthode, pays, centre, année) sur le quadriptyque ?

Code	Libellé	Construction
0	Quadriptyque incomplet	Si au moins une des variables parmi la méthode (B8_METHODE), le pays (B8_PAYS_rec), le centre (B8_CENTRE_rec) et l'année (B8_ANNEE) est non renseignée pour le quadriptyque
1	Quadriptyque complet	Si on dispose pour le quadriptyque de la méthode (B8_METHODE), le pays (B8_PAYS_rec), le centre (B8_CENTRE_rec) et l'année (B8_ANNEE)

e. Disponibilité des informations du parcours [parcours_comp]

Indicatrice du niveau d'informations fournies dans le quadriptyque : le quadriptyque est-il rempli jusqu'à la question du lieu de résidence au moment de sa réalisation ? Par définition, cette variable distingue les quadriptyques pour lesquels on peut déterminer l'appartenance au champ (champ_quadri = 0 ou 1 pour les parcours_comp = 1) de ceux pour lequel elle reste indéterminée (champ_quadri = 99 pour les parcours_comp = 0).

Code	Libellé	Construction
0	Parcours incomplet	Si le pays de résidence au moment de la réalisation du quadriptyque est manquant (B8_3_PAYS non renseignée)
1	Parcours complet	Si on dispose pour le quadriptyque de l'information sur le pays de résidence au moment de sa réalisation (B8_3_PAYS renseignée).

11. Harmonisation des libellés des centres

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes R

Programme : 3_Création_base_quadriptyques.R

Type d'opération : recodage de variables, nouvelles variables créées

Informations complémentaires : Fichier excel sur lequel s'appuient les recodages : D:\CloudINED\AMP\16 - Apurement\AMPsf_Recode_B8_CENTRE.xlsx

Les noms des centres médicaux et des pays associés ont été harmonisés donnant lieu à la création des variables recodées B8_CENTRE_REC, B8_CENTREBIS_REC, B8_PAYS_REC et B8_PAYSBIS_REC².

12. Réintégration des variables issues de la base quadriptyques à la base individus

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes R

Programme : 4_Quadriptyques_ordonnes_base_individus.R

Une fois les opérations réalisées sur la base Quadriptyques, nous avons réintégré à la base Individus les variables concernant les quadriptyques, afin notamment de

² Cette variable n'existait pas dans la base initiale où la possibilité pour le 2^e centre cité (CENTREBIS) d'un même recours qu'il soit dans un autre pays que le 1^{er} centre mentionné n'était pas envisagée, nous l'avons donc créée pour prendre en compte cette configuration.

prendre en compte le recodage des noms de centres et de pays. Quelques nouvelles opérations ont été réalisées à cette occasion.

a. Renommage des variables portant sur les quadriptyques

Toutes les variables où apparaissait l'index du quadriptyque (rang de déclaration) ont été renommées de sorte que l'index se trouve à la fin du nom de la variable et non au milieu comme c'était le cas initialement. Par exemple, B8_X_methode devient B8_methode_X³. Voici l'ensemble des variables concernées :

Avant	Après
B8_X_methode	B8_methode_X
B8_X_pays	B8_pays_X
B8_X_centre	B8_centre_X
B8_X_centrebis	B8_centrebis_X
B8_X_annee	B8_annee_X
B8_X_nb	B8_nb_X
B8_X_centrebisyn	B8_centrebisyn_X
B8_4_X_pays	B8_4_pays_X
B8_4_X_dept	B8_4_dept_X
B9_X_who	B9_who_X
B9_X_gender	B9_gender_X
B10_X_success	B10_success_X

Nous avons également intégré à la base Individus les variables « ordre_chrono_X ». Elles indiquent pour chaque quadriptyque X son rang (par ordre chronologique) dans le parcours du répondant, lorsque l'on a l'information sur l'ordre chronologique de l'ensemble du parcours (« ordre_chrono » non vide).

Ces variables remplacent les B8_3_X dont l'utilisation n'était pas aisée (cf. note de bas de page ¹ ci-dessus) et les complètent (cas des abandons avant B8_3_X pour lesquels on a pu tout de même reconstituer la chronologie, cf. 1.a ci-dessus).

b. Informations sur le dernier quadriptyque réalisé (ordre chronologique)

Pour chaque répondant pour qui nous disposons de l'ordre chronologique des quadriptyques ([ordre_chrono] non vide), nous avons intégré à la base Individus les informations sur le dernier quadriptyque réalisé.

³ La valeur X qui prenait dans la base individus initiale des valeurs entre 1 et 30 est maintenant limitée à 22 : c'est le nombre maximal de quadriptyques renseignés.

Ces variables sont identifiables par leur suffixe « _DER » : B8_METHODE_DER, B8_PAYS_DER, B8_PAYS_REC_DER, B8_CENTRE_DER, B8_CENTRE_REC_DER, B8_CENTREBISYN_DER, B8_PAYSBIS_REC_DER, B8_CENTREBIS_DER, B8_CENTREBIS_REC_DER, B8_ANNEE_DER, B8_NB_DER, B8_4_PAYS_DER, B8_4_DEPT_DER, B9_WHO_DER, B9_GENDER_DER, B10_SUCCESS_DER, CHAMP_QUADRI_DER.

Elles sont vides lorsque l'on ne peut identifier le dernier quadritpyque réalisé.

c. Appartenance du répondant au champ de l'enquête « champ_indiv »

Synthèse de la variable [quadri_champ} au niveau individuel :

Code	Libellé	Construction
0	Hors champ	Si l'ensemble des quadriptyques du répondant sont hors champ (quadri_champ = 0 pour tous les quadriptyques)
1	Dans le champ	Si au moins 1 quadriptyque du répondant est dans le champ (au moins 1 quadri_champ = 1 pour l'ensemble des quadriptyques)
99	Appartenance au champ indéterminée	Si aucun quadriptyque du répondant est dans le champ (aucun quadri_champ = 1) et on ne peut déterminer l'appartenance au champ pour au moins un quadriptyque (au moins un quadri_champ = 99)

13. Recodage de champs ouverts

Un travail de recodage de certaines variables pour lesquelles un champ ouvert était renseigné a été effectué.

a. Moyens de connaissance de l'enquête

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes R
Programme : 5_Recodage_champs_ouverts_et_variables_construites.R
 Type d'opération : recodage de variables, nouvelles variables créées
 Informations complémentaires : Les recodages effectués sont consignés dans le fichier excel D:\CloudINED\AMP\16 – Apurement\ AMPsf_Recode_A0.xlsx, bien que la programmation ne fasse pas appel à ce fichier (le recodage est fait directement dans R).

Ce recodage a impliqué un travail à deux niveaux :

- La première question « Par quel(s) moyen(s) avez-vous eu connaissance de l'enquête ? » à choix multiple (variables dichotomiques A0_X, pour X de 1 à 6, et A0_OTH pour les champs ouverts).

- Les sous-questions suivantes filtrées selon les réponses à la question initiale, demandant des précisions sur les moyens déclarés. Celles-ci se présentent dans la base sous des formats différents et ont parfois impliqué la création de nouvelles variables/modalités.
 - A0_1_1, A0_1_2 et A0_1_3 correspondant aux **associations** citées en positions 1, 2 et 3 (selon le recodage détaillé à la partie 8 ci-dessus) et A0_1_OTH pour le champ ouvert à recoder.
Nous avons ajouté 5 nouveaux codes pour des associations qui ne figuraient pas parmi les modalités proposées, dès lors qu'elles étaient au moins citées par 2 répondants (La planning familial, Collectif Famille.s, Association PMA (Pour Mieux Accompagner), ENDOmind et Jumeaux et Plus)
 - A0_3_nom_rec, A0_3_ville_rec et A0_3_pays_rec (selon le recodage détaillé dans les parties 6a et 7a ci-dessus) correspondant aux noms des **centres médicaux**, villes et pays et A0_3_OTH pour le champ ouvert à recoder.
Nous avons élargi le concept de « centres médicaux » pour y inclure la communication émanant du milieu médical en général, notamment des professionnels de la santé.
 - A0_4_X et A0_5_X correspondant respectivement aux **médias** et **réseaux sociaux** cités (variables dichotomiques) et A0_4_OTH et A0_5_OTH pour les champs ouverts à recoder.
Pour les médias nous avons créé la variable A0_4_6_rec correspondant aux podcasts et pour les réseaux sociaux la variable A0_5_5_rec au réseau LinkedIn, ces deux réponses étant fréquemment citées dans les champs ouverts.

b. Variables [A3], [A4], [A11] et [A12]

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes R
Programme : 5_Recodage_champs_ouverts_et_variables_construites.R
 Type d'opération : recodage de variables, nouvelles variables créées avec le suffixe « _rec »
 Informations complémentaires : Les recodages ont été réalisés sur la base des fichiers excel suivants :

- D:\CloudINED\AMP\16 – Apurement\ AMPsf_Recode_A3_OTH.xlsx
- D:\CloudINED\AMP\16 – Apurement\ AMPsf_Recode_A4_OTH.xlsx
- D:\CloudINED\AMP\16 – Apurement\ AMPsf_Recode_A11_OTH.xlsx
- D:\CloudINED\AMP\16 – Apurement\ AMPsf_Recode_A12_OTH.xlsx

Ces variables correspondent au niveau de diplôme et à la situation professionnelle d'ego et de son éventuel conjoint.

c. Variables « A16_ »

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes R
Programme : 5_Recodage_champs_ouverts_et_variables_construites.R
Type d'opération : recodage de variables, nouvelles variables créées avec le suffixe « _rec »+création de variables
Informations complémentaires : Les recodages ont été réalisés sur la base du fichier excel D:\CloudINED\AMP\16 – Apurement\ AMPsf_Recode_A16_OTH.xlsx

Ces variables portent sur les enfants d'ego lorsqu'ego a déclaré être en couple (question à choix multiple avec variables dichotomiques A16_X avec X de 1 à 6 et le champ ouvert A16_OTH) et leur nombre (questions A16_2_nb, A16_3_nb et A16_4_nb filtrées selon les réponses données à la question précédente). La variable A16_6_type construite à l'issue de ce recodage permet de caractériser l'information contenue dans la variable A16_OTH.

Code	Libellé
1	Enfant décédé (déclaré ou non parmi les enfants)
2	Décès intra-utérin
3	IVG
4	Décès intra-utérin et IVG
5	Adoption (déclaré ou non parmi les enfants)
6	Projet enfant en cours
7	Précision sur enfant
8	Précision pas d'enfant
9	Précision non comprise

14. Création de variables relatives au questionnaire

Afin de faciliter l'exploitation, quelques premières variables ont été construites. Leur construction est détaillée dans les paragraphes ci-dessous.

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes R
Programme : 5_Recodage_champs_ouverts_et_variables_construites.R
Type d'opération : création de variables

a. Classe d'âge du répondant [age_cl]

Construite à partir des variables A2AGE (âge au moment du remplissage du questionnaire, calculée automatiquement dans la base initiale) et A2_codes (variable construite qui identifie les « Ne souhaite pas répondre » à A2).

Code	Libellé	Construction
1	Moins de 30 ans	A2AGE < 30
2	Entre 30 et 34 ans	30 <= A2AGE <= 34
3	Entre 35 et 39 ans	35 <= A2AGE <= 39
4	Entre 40 et 44 ans	40 <= A2AGE <= 44

5	45 ans et plus	A2AGE > 44 ET A2AGE ≠ 888
8	Ne souhaite pas répondre	A2AGE = 888

b. Type de couple [type_couple]

Construite à partir des variables A6 (situation de couple), A1 (sexe de l'enquêté) et A8 (sexe de son partenaire). Elle porte sur les personnes en couple uniquement (A6 = 1).

Code	Libellé	Construction
1	Couple hétérosexuel	(A1 = 1 & A8 = 2) OU (A1 = 2 & A8 = 1)
2	Couple de femmes	A1 = 1 & A8 = 1
3	Couple d'hommes	A1 = 2 & A8 = 2
4	Couple composé d'au moins une personne trans, intersexe ou qui se définit autrement que "Homme" ou "Femme"	A6 = 1 & (A1 = 3,4,5 OU A8 = 3,4,5)
8	Ne souhaite pas répondre à la question sur son sexe et/ou le sexe du partenaire	A6 = 1 & (A1 = 6 OU A8 = 6)

c. Variable Ego a des enfants [enfants]

Construite à partir des variables issues des réponses aux questions à choix multiple A16_X_rec (telles que recodées comme expliqué à la partie 13c ci-dessus) et A17_X (fait d'avoir des enfants pour les personnes qui ne sont pas en couple ou ont refusé de répondre à la question sur le couple). On considère que les enfants eus par le conjoint de son côté ne sont pas des enfants de l'enquêté.

Code	Libellé	Construction
1	Au moins un enfant et sans grossesse en cours	((A16_2_rec = 1 OU A16_3_rec = 1) & A16_5_rec = 0) OU (A17_1 = 1 & A17_2 = 0 & A17_3 = 0)
2	Au moins un enfant et grossesse en cours	((A16_2_rec = 1 OU A16_3_rec = 1) & A16_5_rec = 1) OU (A17_1 = 1 & A17_2 = 0 & A17_3 = 1)

3	Sans enfant et une grossesse en cours	(A16_5_rec = 1 & A16_2_rec = 0 & A16_3_rec = 0) OU A17_3 = 1 & A17_1 = 0
4	Ni enfant ni grossesse en cours	A16_1_rec = 1 OU (A16_4_rec = 1 & A16_2_rec = 0 & A16_3_rec = 0 & A16_5_rec = 0) OU (A17_2 = 1 & A17_1 = 0 & A17_3 = 0)
5	Autres configurations	A16_6_rec = 1 & A16_1_rec = 0 & A16_2_rec = 0 & A16_3_rec = 0 & A16_4_rec = 0 & A16_5_rec = 0
8	Ne souhaite pas répondre	A16_1_rec = 88 OU A17_1 = 88

15. Assignation de labels et formats aux variables des bases Individus & Quadriptyques

Répertoire : D:\CloudINED\AMP\16 - Apurement\Programmes R
Programme : 6b_Base_Individus_assign_formats_labels.sas (pour la base Individus)
6a_Base_Quadri_assign_formats_labels.sas (pour la base Quadriptyques)
Type d'opération : assignation de formats & labels

Une fois les bases constituées, la dernière étape consiste à assigner des labels (libellés des questions) et des formats (modalités de réponse en clair).

Des programmes de construction des formats sont disponibles :

- Pour la base Individus : AMPsf_Individus_formats.sas dans le dossier D:\CloudINED\AMP\16 - Apurement\Bases\Bases Individus\Base finale
- Pour la base Quadriptyques : AMPsf_quadriptyques_formats.sas dans le dossier D:\CloudINED\AMP\16 - Apurement\Bases\Bases Quadriptyques\