

## CHAPITRE 7

# TRAITEMENT ET ANALYSES DES DONNEES

*Ce chapitre s'adresse aux coordinateurs de l'enquête, aux spécialistes en traitement et analyse des données et aux personnes-ressources techniques. Il vous aidera à :*

- Préparer le traitement des données ;
- Mettre en place un système de gestion du traitement et de l'analyse des données ;
- Effectuer la saisie des données ;
- Contrôler les données et créer un fichier de données "propre" pour l'analyse ;
- Produire les tableaux contenant les indicateurs ;
- Préparer les fichiers de données et la documentation pour l'analyse au-delà du rapport principal, et distribuer les données.

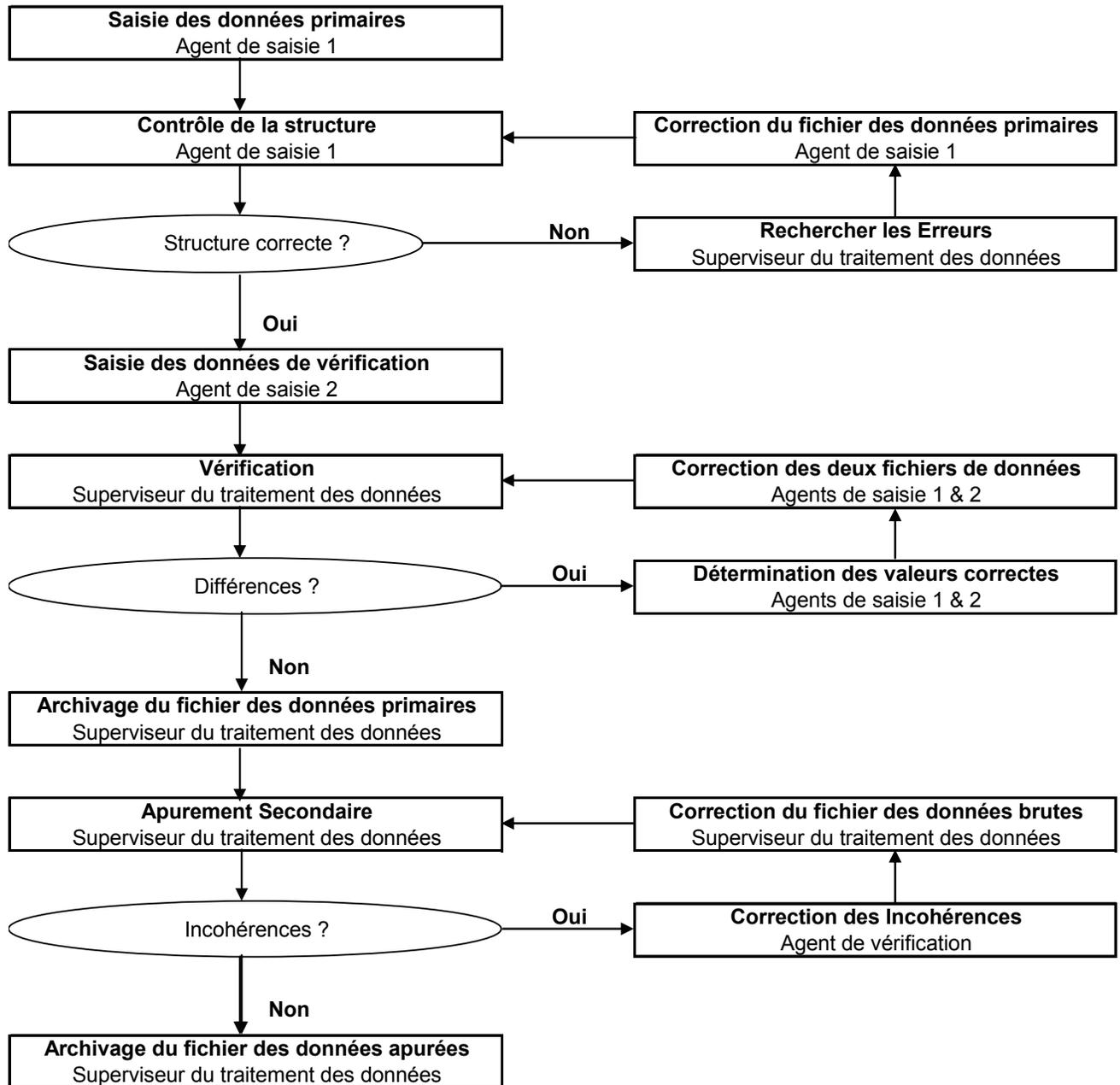
Le système de traitement des données de la MICS3 est conçu pour produire à temps les premiers résultats de l'enquête quelques semaines seulement après la fin des travaux de terrain. Ce chapitre contient des informations qui vous aideront à mettre en place un bon planning et une bonne préparation à l'avance, qui feront de cet objectif une réalité. Ce chapitre commence par une vue d'ensemble du système de traitement des données de la MICS3. Ensuite, il présente de manière détaillée chacune des composantes, en donnant des références pour des sources d'information additionnelles à chaque fois que c'est nécessaire. Il se termine par une série de trois listes de contrôle qui vous aideront à réaliser avec succès le traitement des données de votre enquête.

### VUE D'ENSEMBLE

La raison pour laquelle le système de traitement et d'analyse des données de la MICS3 permet de réaliser un tel gain de temps, c'est le fait que les données soient traitées en même temps que les activités de terrain. Les données de chaque grappe sont stockées dans un fichier de données séparé, et sont traitées dès que les questionnaires sont ramenés du terrain. Cette approche subdivise l'opération de traitement et d'analyse des données en plusieurs étapes, lui permettant d'avancer parallèlement à l'évolution de la collecte des données sur le terrain. Ainsi, au moment où les derniers questionnaires seront achevés et ramenés au Bureau Central, la plupart des données aura déjà été traitée.

Le traitement et l'analyse des données par grappes n'est pas une tâche difficile, mais elle exige une organisation très méticuleuse. Le système de traitement et d'analyse des données peut être divisé en trois phases : la préparation, le traitement primaire des données et le traitement secondaire des données. Chacune de ces phases est résumée dans les sections qui suivent, et chacune d'elles a sa propre liste de contrôle à la fin de ce chapitre.

**Tableau 7.1**  
**Le Système de Traitement et d'Analyse des Données de la MICS3**



### **PREPARATION DE LA SAISIE DES DONNEES**

La préparation de la phase de saisie des données a pour finalité de permettre le démarrage de la saisie des données quelque temps seulement après le début des travaux de terrain. Les principales étapes de cette phase sont les suivantes :

- Acquisition de l'équipement informatique et préparation de la salle ou de l'espace pour le traitement des données ;
- Identification et recrutement du personnel approprié ;
- Adaptation des programmes informatiques en fonction des spécificités des questionnaires du pays ;
- Mise en place d'un système de gestion des questionnaires et des fichiers des données.

### **TRAITEMENT PRIMAIRE DES DONNEES**

Le but du traitement primaire des données est de produire des fichiers de données propres et apurés. Le traitement primaire des données implique les étapes suivantes :

- Saisie de tous les questionnaires d'une grappe dans un fichier de données ;
- Contrôle de la structure du fichier de données ;
- Deuxième saisie des données et vérification du fichier de données ;
- Sauvegarde du fichier de données contrôlées et vérifiées ;
- Apurement secondaire du fichier des données ;
- Sauvegarde du fichier de données apurées, ou fichier final.

Le processus de traitement primaire des données est résumé dans le schéma de la page précédente. Il est important de noter que le contrôle de la structure, la vérification de la saisie des données et l'apurement secondaire sont des procédures itératives, qui doivent être répétées autant de fois que nécessaire, jusqu'à la résolution de tous les problèmes détectés, ou jusqu'à ce que ces derniers soient considérés comme acceptables.

### **TRAITEMENT SECONDAIRE DES DONNEES**

Le but du traitement secondaire des données est de produire des fichiers d'analyse des données et de créer les tableaux standards de la MICS3. Le traitement secondaire des données implique les étapes suivantes :

- Concaténation de tous les fichiers de données de grappes en un fichier de données ;
- Exportation des données vers le logiciel SPSS ;
- Calcul des pondérations ;

- Création de l'indice de richesse<sup>1</sup> ;
- Recodification des variables pour simplifier l'analyse ;
- Préparation et sortie des tableaux requis pour l'analyse des données ;
- Archivage et distribution des fichiers de données.

## PERSONNEL ET INFRASTRUCTURE

### PERSONNEL

Pour la MICS3, l'équipe de traitement et d'analyse des données comprend quatre types de personnel : un archiviste, des agents de saisie, des opérateurs en charge de l'apurement secondaire des données et un superviseur du traitement et de l'analyse des données. Des responsabilités distinctes sont assignées à chacune de ces fonctions, et le fait de les combiner pourrait compromettre la qualité de vos données.

L'**archiviste** a pour tâche de contrôler et de classer les questionnaires qui rentrent du terrain. Lorsque les documents d'une grappe arrivent au niveau du bureau de traitement des données, il/elle vérifie que tous les questionnaires sont effectivement présents et prêts à être saisis. Si des questionnaires sont manquants, il/elle devra résoudre ce problème avec l'aide de l'équipe de terrain (les étapes précises que l'archiviste doit suivre sont détaillées plus bas).

Les **agents de saisie** ont pour tâche la saisie des données. Ils doivent aussi avoir une expérience dans le domaine de saisie et une bonne connaissance des questionnaires de l'enquête. Une manière d'y parvenir est de faire participer les agents de saisie à la formation des enquêteurs. Avant le démarrage de la saisie, une session de formation spécifique de 2-3 jours devrait être organisée, afin de permettre aux agents de saisie de se familiariser avec les programmes et les masques de saisie, et pour leur permettre d'avoir un temps de pratique. Avant la fin de la saisie, les agents de saisie devraient être à l'aise avec le programme de saisie et être conscients de leurs responsabilités quotidiennes. Le nombre d'agents de saisie nécessaires dépend du nombre d'ordinateurs disponibles, et ce point est traité en détail dans les sections qui suivent.

Les **agents de vérification** ont pour tâche de rechercher et de traiter les incohérences complexes relevées par le programme d'apurement secondaire. Ils doivent avoir une très bonne connaissance des questionnaires et des objectifs de l'enquête. Les procédures d'apurement sont fournies dans l'Annexe Sept en vue de les aider dans le processus d'apurement secondaire des données. Habituellement, un ou deux agents de vérification seront en mesure de faire tout le travail.

Le **superviseur du traitement des données** est un membre particulièrement important de l'équipe de traitement et d'analyse des données de l'enquête. Il a en charge l'adaptation des pro-

---

<sup>1</sup> Encore appelé "Indice de Bien-être Economique"

grammes modèles aux spécificités des questionnaires utilisés dans le pays, ainsi que la supervision de toutes les activités de traitement et d'analyse. Le superviseur du traitement des données doit avoir une bonne expérience dans la gestion des opérations de traitement informatique d'enquêtes de grande envergure ou de recensements, une parfaite maîtrise des questionnaires et des compétences avérées dans l'utilisation et la programmation avec les logiciels CPro et SPSS. Le superviseur du traitement des données doit être disponible à plein temps durant la période de saisie, d'apurement et de tabulation.

Le superviseur du traitement des données doit être identifié suffisamment à l'avance pour être impliqué dans la révision des questionnaires de la MICS3. Il doit être consulté, afin de s'assurer que les codes utilisés dans les questionnaires sont appropriés et sans ambiguïté et que toutes les informations relatives à l'identification dont on a besoin en font partie. Le superviseur du traitement des données sera également en mesure de contribuer à la révision finale des questionnaires, sur la base de l'expérience acquise lors de la saisie des questionnaires du pré-test.

#### **MATERIEL INFORMATIQUE ET AUTRES EQUIPEMENTS**

Les équipements nécessaires pour le traitement et l'analyse des données sont listés ci-dessous :

- Ordinateurs pour la saisie des données ;
- Ordinateur du superviseur du traitement des données ;
- Un périphérique secondaire de sauvegarde des données (un port USB ou un lecteur de CD-RW) ;
- Des disquettes (ou tout autre moyen permettant aux agents de saisie de transférer les fichiers de données vers l'ordinateur du superviseur du traitement des données, par exemple par réseau) ;
- Une imprimante ;
- Du papier ;
- Des cartouches d'encre/rubans d'imprimante ;
- Des interrupteurs électriques de sécurité ;
- Des onduleurs (UPS) ;
- Des bics verts.

Il est recommandé que les ordinateurs dont on se servira pour la saisie des données aient un processeur de type Pentium, utilisent Windows 95 ou plus, aient au minimum 32 mégaoctets de mémoire (RAM) et au moins un disque dur d'un gigaoctet, ainsi qu'un lecteur de disquettes de 3,5 pouces (ou soit mis en réseau). Le nombre d'ordinateurs dont on a besoin pour l'exploitation de l'enquête dépendra de la taille de l'échantillon, du nombre d'heures de saisie hebdomadaire, du local disponible et du calendrier de l'enquête. Pour avoir le nombre d'ordinateurs dont on a besoin pour la saisie des données, il faut estimer le temps moyen de saisie des questionnaires d'un ménage par agent, et le multiplier par le nombre de ménages attendus en fonction de la taille de l'échantillon. Si vous n'êtes pas en mesure d'estimer le temps nécessaire pour saisir les questionnaires, utilisez 20 à 30 minutes comme une moyenne générale de saisie par ménage ; ce-

la dépendra du nombre attendu de femmes et d'enfants par ménage. Il faut multiplier ce temps estimé par le nombre de ménages pour obtenir le nombre d'heures nécessaires pour la saisie des données. Il faut maintenant diviser cette dernière estimation par le nombre d'heures de saisie hebdomadaire par agent ; puis il faut encore diviser par le nombre de semaines dont vous aurez besoin pour achever la saisie des données (si vous comptez finir cette saisie une semaine après avoir reçu les derniers questionnaires du terrain).

Si, par exemple, la taille de l'échantillon de l'enquête est de 6.000 ménages, et si chaque ménage prend en moyenne 20 minutes de saisie, le temps nécessaire pour la saisie de l'ensemble des ménages est de 2.000 heures. Si vous avez 8 semaines pour achever la saisie des données, 250 heures de travail hebdomadaire seront nécessaires. Si chaque agent de saisie travaille 40 heures par semaine, vous aurez besoin de 7 ordinateurs et de 7 agents de saisie. Il arrive parfois qu'une équipe de deux agents de saisie travaillent sur le même ordinateur. Chaque agent travaillera par exemple pendant 6 heures, de telle sorte qu'un ordinateur soit employé pendant 12 heures par jour.

Le superviseur du traitement des données devrait disposer d'un ordinateur plus puissant, avec un processeur plus rapide, Windows 98 ou plus, au moins 64 mégaoctets de RAM, un disque dur d'un gigaoctet ou plus, ainsi qu'un lecteur de disquettes de 3,5 pouces (ou soit mis en réseau avec les ordinateurs des agents de saisie), et un second périphérique de stockage des données pour les copies de sauvegarde des fichiers de données.

Des interrupteurs électriques de sécurité et des onduleurs (UPS) sont nécessaires dans les pays affectés par de fréquentes coupures de courant électrique. Les bics verts doivent être utilisés à chaque fois qu'un membre de l'équipe de traitement des données modifie les données d'un questionnaire. L'encre de couleur verte permet de distinguer ces changements des données d'origine enregistrées par les agents enquêteurs (avec de l'encre bleue) et toute modification effectuée par l'équipe de terrain (avec de l'encre rouge).

## **LOGICIELS**

Les programmes standards pour le traitement et l'analyse des enquêtes MICS3 ont été développés dans CPro 2.6 et SPSS. CPro qui a été utilisé par plusieurs services de statistiques à travers le monde pour le traitement des recensements et des enquêtes, a été développé en partenariat par United States Census Bureau, ORC Macro International et SerPro Ltd. Il peut être téléchargé gratuitement à partir du site Web de l'US Bureau of Census<sup>2</sup>. SPSS est un logiciel commercial qui est disponible au siège de l'UNICEF et auprès de plusieurs fournisseurs de logiciels.

---

<sup>2</sup> L'adresse du site Web est : <http://www.census.gov/ipc/www/cspro/>

## ESPACE DE TRAVAIL

La saisie et l'apurement des données doivent se faire dans des salles séparées. La salle ou l'espace dans lequel les opérations de saisie seront exécutées doit avoir suffisamment de place pour que chaque agent de saisie puisse travailler à l'aise avec son ordinateur et ses questionnaires. Il doit y avoir également des bureaux ou des tables de travail et des installations électriques nécessaires en quantité suffisante. La salle doit être fraîche, bien éclairée et autant que possible elle ne doit pas avoir de poussière ou d'humidité. Si le climat est chaud, la salle doit être climatisée. Pour éviter de perdre des données en raison des coupures de courant, vous devez utiliser un onduleur (UPS) pour chaque ordinateur. Si les coupures de courant sont fréquentes ou prolongées, il sera nécessaire d'avoir une autre source de courant de secours telle que les groupes électrogènes.

La salle de vérification et d'apurement des données est réservée à l'archiviste et aux agents de vérification. Elle aussi doit être fraîche et bien éclairée, et avoir suffisamment d'espace pour permettre aux agents de vérification d'effectuer leur travail de révision des questionnaires dans de bonnes conditions. L'idéal, c'est d'avoir une salle de vérification qui dispose de suffisamment d'étagères ou d'armoires réservées pour le stockage des questionnaires qui reviennent du terrain de manière ordonnée. Si les questionnaires ne peuvent pas être stockés dans la salle de vérification, ils devraient être stockés à proximité de la salle de traitement des données, de sorte qu'ils soient facilement accessibles, puisqu'on aura besoin de les manipuler pendant toute la durée du traitement des données. Soyez prudent, et ne sous-estimez pas le volume d'espace dont vous aurez besoin pour stocker les milliers de questionnaires que vous aurez au bureau à la fin des opérations de terrain.

## ADAPTATION DES PROGRAMMES STANDARDS

Comme nous l'avons indiqué au Chapitre 3, le questionnaire modèle MIC3 doit être adapté à la situation de chaque pays. Cela signifie que les programmes modèles de saisie, d'apurement et de tabulation des données seront également modifiés pour qu'ils soient conformes aux changements faits dans le questionnaire. Plus on fera des changements dans le questionnaire modèle, plus on devra passer de temps à adapter et à tester les programmes. Par exemple, si des nouvelles questions sont ajoutées au questionnaire, des ajouts correspondants devront être faits sur les programmes de saisie, d'apurement et de tabulation des données.

Cette procédure sera beaucoup plus facile si les numéros de questions dans le questionnaire modèle sont maintenus. Si des questions sont ajoutées, une lettre devrait être ajoutée au numéro de la question (par exemple, une question ajoutée entre WS4 et WS5 devrait être numérotée WS4A). Par ailleurs, si des questions sont éliminées, les questions restantes ne doivent pas être renumérotées. De même, lorsque des catégories de codes sont ajoutées à celles du questionnaire modèle, elles devraient être ajoutées à la fin de la liste existante, de sorte que les autres codes restent inchangés. L'adaptation des programmes de saisie et d'apurement des données doit être

achevée avant le pré-test. Les questionnaires du pré-test peuvent être saisis et apurés en utilisant ces programmes. Cette procédure servira pour deux raisons. Elle identifiera les problèmes liés à la codification et aux sauts dans le questionnaire et identifiera également les erreurs dans les programmes. Une fois le pré-test achevé et le questionnaire finalisé, les changements définitifs peuvent être faits aux programmes. Les sections qui suivent donnent des directives de base pour la modification des modèles de dictionnaires des données et des modèles d'applications de CSPro. Des directives plus détaillées sur le contenu des applications de CSPro sont présentées dans d'autres documents.

Même dans le cas où vous n'ajoutez pas de questions aux questionnaires modèles, les modèles de dictionnaires des données et les modèles d'applications comportent des éléments qui doivent impérativement être mis à jour (il s'agit, par exemple, de l'étendue acceptable pour la date de l'interview, l'étendue acceptable pour le numéro de la grappe, etc.). Ces éléments sont spécifiques à chaque pays, et ne peuvent être remplis que par vous. Alors, même si votre pays utilise le questionnaire modèle, vous devriez procéder à l'adaptation des programmes standards.

### **LES DICTIONNAIRES DES DONNEES**

Dans l'Enquête par Grappes à Indicateurs Multiples, les groupes de questions couvrant le même thème (par exemple, la mortalité maternelle, l'utilisation de la contraception et la vaccination) sont regroupées dans des modules, qui eux-mêmes sont regroupés dans des questionnaires (pour le ménage, pour les femmes individuellement et pour les enfants de moins de cinq ans). Dans CSPro, le dictionnaire des données sert à décrire cette structure des données : un groupe de variables portant sur le même thème (questions) constitue un enregistrement (module), et un groupe d'enregistrements constitue un niveau (questionnaire). Ces derniers sont stockés dans un fichier dictionnaire (qui a comme extension : *dcf*). En plus du dictionnaire des données, les formulaires liés au dictionnaire sont utilisés pour la saisie des données. Habituellement, il y a un formulaire pour chaque enregistrement. Les formulaires sont stockés dans un fichier de formulaires (qui a comme extension : *fmf*). Les fichiers *dcf* et *fmf* peuvent être modifiés directement. La meilleure manière de le faire est d'ouvrir le fichier de formulaires dans CSPro. Cela vous permettra d'accéder en même temps au dictionnaire des données et aux formulaires, et de vous assurer que les deux demeurent synchronisés. Il est recommandé de conserver une copie de sauvegarde du modèle de dictionnaire des données et du fichier des formulaires pour référence.

Il existe trois types de questionnaires pour la MICS3. Le Questionnaire Individuel de la Femme et le Questionnaire des Enfants de Moins de Cinq Ans correspondent à une unité primaire d'analyse : une femme et un enfant respectivement. Le Questionnaire Ménage comporte deux unités d'analyse : le ménage et les membres du ménage. Tous les types de questionnaires sont stockés dans *mics3.dcf* et *entry.fmf*.

## VARIABLES ET NIVEAUX D'IDENTIFICATION

Dans CSPro, chaque questionnaire doit avoir une série de variables qui l'identifient de manière unique. Par exemple, un ménage est identifié par son numéro de grappe et son numéro ménage. Les variables qui permettent d'identifier un questionnaire sont connues sous le nom de variables d'identification. Le Tableau 7.2 ci-dessous liste les types de questionnaires et leurs variables d'identification.

**Tableau 7.2**  
**Types de Questionnaires et leurs Variables d'Identification**

Questionnaire	Numéro de Grappe	Numéro Ménage	Numéro de Ligne
Ménage	HH1	HH2	
Femmes	HH1	HH2	LN
Enfants	HH1	HH2	LN

Comme le montre le tableau, les femmes et les enfants ont les mêmes variables d'identification. Etant donné que chaque membre du ménage est listé sur une ligne distincte dans la Feuille d'Enregistrement du Ménage, deux femmes ou deux enfants ne sauraient avoir le même numéro de ligne, même s'ils sont dans le même ménage. Ainsi, le numéro de ligne, lorsqu'il est combiné avec le numéro de grappe et le numéro ménage, identifie de manière unique la femme ou l'enfant.

Dans un dictionnaire CSPro, un niveau est défini par un ensemble de variables d'identification. Il y a deux niveaux dans le dictionnaire de la MICS3 : les ménages et les individus (autrement dit, les femmes éligibles et les enfants éligibles). Les ménages constituent le premier niveau, alors que les femmes et les enfants composent le second niveau. Cette structure hiérarchique est naturelle, car dans le questionnaire de la MICS3, chaque femme ou chaque enfant appartient à un ménage, alors qu'un ménage donné peut avoir plusieurs femmes et enfants.

Les questionnaires individuels femmes et les questionnaires enfants sont stockés au même niveau, car ils s'appliquent tous deux aux membres du ménage. L'application de saisie des données contient une logique qui saute les formulaires appartenant aux enfants lors de la saisie du questionnaire d'une femme, et saute les formulaires appartenant aux femmes lors de la saisie du questionnaire d'un enfant. Ainsi, bien que les questionnaires femmes et les questionnaires enfants soient tous deux stockés comme des cas de niveau deux, ils n'ont pas de variables communes, hormis les variables d'identification.

## MODULES

Le dictionnaire des données a été conçu pour refléter la structure modulaire des questionnaires de la MICS3. Chaque module est stocké dans son propre enregistrement (exceptions : le module d'Enregistrement du Ménage et le module des Orphelins et Autres Enfants rendus Vulnérables

du fait du VIH/SIDA comptent chacun deux enregistrements, à cause de leur structure inhabituelle) dans *mics3.dcf* et chaque enregistrement a un formulaire (ou deux, dans le cas du module sur les Caractéristiques du Ménage) qui lui est associé dans *entry.fmf*. Ainsi, si votre pays n'utilise pas un module, vous pouvez l'enlever en supprimant son enregistrement et son formulaire (et en retirant toute autre logique qui y fait référence dans l'application de saisie des données).

Les modules disponibles pour le Questionnaire Ménage (avec leurs codes modules listés entre parenthèses) sont : les Informations du Ménage (HH), l'Enregistrement du Ménage (HL et TO), l'Education (ED), l'Eau et l'Assainissement (WS), les Caractéristiques du Ménage (HC), les Moustiquaires Imprégnées aux Insecticides (TN), les Orphelins et autres Enfants rendus Vulnérables par le VIH/SIDA (OV et OR), le Travail des Enfants (CL), la Discipline de l'Enfant (CD), l'Handicap (DA), la Mortalité Maternelle (MM) et l'Iodation du Sel (SI).

Les modules disponibles pour le Questionnaire Individuel des Femmes sont : le Cadre d'Information des Femmes (WM), la Mortalité des Enfants (CM), l'Anatoxine Tétanique (TT), la Santé de la Mère et du Nouveau-né (MN), les Unions/Mariage (MA), la Sécurité de la Propriété et la Durabilité du Logement (ST), la Contraception (CP), les Mutilations Génitales Féminines/Excision (FG), les Attitudes vis-à-vis de la Violence Conjugale (DV), le Comportement Sexuel (SB) et le VIH/SIDA (HA).

Les modules disponibles pour le Questionnaire des Enfants de Moins de Cinq Ans sont : le Cadre d'Information des Enfants de Moins de Cinq Ans (UF), l'Enregistrement des Naissances et l'Education de la Petite Enfance (BR), le Développement de l'Enfant (CE), la Vitamine A (VA), l'Allaitement (BF), les Soins des Enfants Malades (CA), le Paludisme (ML), la Vaccination (IM) et l'Anthropométrie (AN).

### CONVENTIONS DE LIBELLES DES VARIABLES

Les variables sont nommées selon le module du questionnaire dans lequel elles sont situées et le numéro de la question dont elles contiennent la réponse. Par exemple, la question 9 de l'Enregistrement du Ménage est stockée dans la variable dénommée HL9. Certaines questions sont constituées de deux ou plusieurs parties, dont chacune est identifiée par une lettre distincte. Pour de telles questions, chaque partie est stockée dans une variable distincte. Les noms de ces variables distinctes incluent les lettres qui distinguent les différentes parties de la question. Par exemple, la question 11 du module sur la Santé de la Mère et du Nouveau-né comporte deux parties. La première partie de la question est stockée dans la variable *MN11A*, et la seconde partie est stockée dans *MN11*.

Certaines questions comportent deux ou plusieurs catégories de réponses. Ces questions sont stockées dans une variable individuelle et les catégories de réponses sont définies comme des sous-items. Lorsque ces questions concernent des dates, les lettres "d" (pour jour), "m" (pour mois) et "y" (pour année) sont ajoutées au nom de la variable de base pour créer le nom des

sous-items. Dans la question 6 du Panneau d'Information des Femmes, par exemple, le jour, le mois et l'année de naissance de la femme sont requis. Les réponses correspondantes sont stockées dans *WM8*, qui a trois sous-items : *wm8d*, *wm8m* et *wm8y*.

Certaines questions ont une structure dans laquelle la première partie de la réponse en est la forme alors que la seconde partie est la réponse elle-même. Ces questions sont stockées dans une variable unique ; quant à la forme et à la réponse elle-même, elles sont définies comme des sous-items. Le nom du sous-item stockant la forme de la réponse est le nom de la variable à laquelle la lettre "u" (pour unités) est adjointe, alors que le nom du sous-item stockant la réponse elle-même est le nom de la variable à laquelle la lettre "n" (pour nombre) est adjointe. Par exemple, la question 13 du module sur la Santé de la Mère et du Nouveau-né enregistre la durée écoulée entre la naissance de l'enfant et sa première mise au sein. La personne enquêtée peut donner une réponse en heures ou en jours. La réponse est stockée dans la variable *mn13*, aux sous-items *mn13u* et *mn13n*.

### QUESTIONS A REPONSES MULTIPLES ET VARIABLES ALPHANUMERIQUES

Un certain nombre de questions acceptent des réponses multiples. Dans les questionnaires, ces questions sont caractérisées par des réponses à codes alphanumériques (autrement dit, les lettres A à Z). Dans le dictionnaire des données, la réponse à une question à réponses multiples est stockée dans une variable alphanumérique, dont la longueur est égale au nombre maximum de réponses possibles. Ce sont là les seules variables alphanumériques du dictionnaire. Chaque variable alphanumérique comporte un sous-item pour chaque code de réponse du questionnaire. Le nom de chacun de ces sous-items est le nom de la variable plus le code de la réponse que ce sous-item représente.

Par exemple, la seconde question du module sur la Santé de la Mère et du Nouveau-né enregistre toutes les personnes ayant fourni à la femme des soins prénatals avant sa dernière naissance. Les codes de réponses possibles sont A, B, C, F, G, H, X et Y. Par conséquent, la variable *mn2* a une longueur de huit caractères, et elle comporte huit sous-items : *mn2a*, *mn2b*, *mn2c*, *mn2f*, *mn2g*, *mn2h*, *mn2x* et *mn2y*.

### CONVENTIONS DE CODIFICATION

Les modèles de dictionnaires utilisent des codes standards pour certaines réponses. Nous commencerons par présenter les conventions de codification pour les variables numériques. La réponse "Autre" est toujours codée avec un 6 – précédé d'un ou plusieurs 9. Les réponses incohérentes sont toujours codées avec un 7 – précédé d'un ou plusieurs 9. La réponse "Ne Sait Pas" est toujours codée avec un 8 – précédé d'un ou plusieurs 9. Les questions avec réponse manquante (autrement dit, là où l'agent enquêteur n'a pas enregistré de réponse à une question pourtant applicable) sont toujours codées avec 9 – précédé d'un ou plusieurs 9. Les questions qui ne sont pas applicables à la personne enquêtée sont toujours codées comme vides. Le Tableau 7.3 ci-dessous présente de manière résumée les conventions de codification standards.

**Tableau 7.3**  
**Résumé des Conventions de Codification Standards**

Réponse	Longueur de la Variable				
	Alphabétique	Un caractère	Deux caractères	Trois caractères	Quatre caractères
Autre	X	6	96	996	9996
Non/Aucun	Y	NA	NA	NA	NA
Incohérent	NA	7	97	997	9997
Ne Sait Pas	Z	8	98	998	9998
Manquante	?	9	99	999	9999
Non applicable	Vide	Vide	Vide	Vide	Vide

Etant donné que les codes allant de 6 à 9 sont réservés pour un usage particulier, toute question qui comporte plus de six catégories de réponses devrait avoir des catégories de réponses à deux chiffres, précédés d'un zéro (par exemple, 01, 02, 03, 04, 05, 06, 07, 96, 97, 98 et 99).

Pour les variables alphanumériques, la réponse "Autre" est toujours codée X, la réponse "Ne Sait Pas" est toujours codée Z, une valeur manquante est toujours codée en utilisant le point d'interrogation ( ? ), et non applicable est toujours codé comme un vide.

### ETENDUES

La plupart des questions figurant dans les questionnaires de la MICS3 ont des étendues de réponses déjà prédéfinies. Pour les variables, les étendues sont définies dans le dictionnaire *mics3.dcf*. Lors de la saisie des données, CSPro vérifie que toute valeur saisie dans une variable fait partie de l'étendue prédéfinie pour cette variable. CSPro accepte plusieurs étendues pour chaque variable, ainsi, des questions qui ont des étendues de réponses qui ne se suivent pas (par exemple, 1-8, 96, 98 et 99) devraient être définies en tenant compte de plusieurs étendues (par exemple, 1-8, 96, 98 et 99, plutôt que 1-99). Certes, les étendues du dictionnaire sont utiles pour le contrôle des étendues simples, mais des étendues plus compliquées ou conditionnelles (par exemple, la cohérence entre le jour et le mois au sein d'une variable de date) devraient être également vérifiées dans les applications de saisie des données ou d'apurement de ces dernières.

### LE PROGRAMME DE SAISIE DES DONNEES

L'application de saisie des données est un programme long et complexe. Des contraintes d'espace nous empêchent de le présenter de manière détaillée dans ce chapitre. En lieu et place, cette section va mettre l'accent sur certains des points importants concernant l'application de saisie des données.

## SAUTS

Les questionnaires de la MICS3 font un large usage des sauts. Les sauts sont des instructions figurant dans le questionnaire et qui disent à l'agent enquêteur de sauter toutes les questions situées entre la question actuelle et une question située plus bas dans le questionnaire. Les sauts figurant dans le questionnaire devraient être en conformité avec les sauts correspondants dans le programme de saisie des données. Les sauts du programme de saisie des données définissent le chemin de saisie des données. CSPro applique de manière stricte le chemin de saisie des données lorsque les commandes “*skip to*” ou “*skip to next*” sont utilisées.

## MESSAGES D'ERREURS

Lorsqu'un agent de saisie entre, pour une variable, une valeur qui est incohérente avec une information déjà entrée, il est utile d'afficher un message d'erreur. Ce message d'erreur devrait expliquer la nature du problème et fournir toute information qui pourrait aider à la résolution de l'incohérence. Dans CSPro, la fonction *errmsg* affiche un message d'erreur, avec un texte défini par l'utilisateur, à chaque fois qu'elle est sollicitée. Les messages d'erreur du programme de saisie des données sont nombreux, et ils sont stockés dans le fichier *entry.mgf*. Le texte, les nombres et les incohérences qui entraînent l'affichage de chacun de ces messages d'erreur sont listés dans l'Annexe Sept, de même que les directives à suivre pour leur résolution.

Il vous appartiendra de passer en revue le questionnaire pour déterminer si les questions qui y ont été ajoutées nécessitent des contrôles de cohérence. Le cas échéant, vous devriez ajouter des logiques pour vérifier la cohérence soit dans le programme de saisie des données, soit dans le programme d'apurement des données, ou alors dans les deux programmes. Lorsque vous y ajoutez un contrôle de cohérence, assurez-vous que le message d'erreur correspondant a été ajouté au fichier de messages de saisie des données ou d'apurement des données. De même, si vous ajoutez des messages d'erreur, veillez à ne pas utiliser un numéro de message d'erreur déjà existant.

Certains messages d'erreur sont suivis de la commande *reenter*, qui renvoie au champ qui est entrain d'être saisi actuellement. Cela oblige l'agent de saisie à régler l'erreur avant de continuer la saisie. Etant donné qu'il sera souvent demandé à l'agent de saisie de procéder aux corrections, une supervision rapprochée de cette activité sera nécessaire. Lorsque vous insérez vos propres messages d'erreur, il est important que vous réfléchissiez soigneusement au fait de savoir si vous voulez obliger l'agent de saisie à résoudre le problème avant de poursuivre la saisie. Si c'est le cas, faites suivre votre message d'erreur par la commande *reenter*.

## VARIABLES ALPHANUMERIQUES

L'application de saisie des données vérifie que les variables alphanumériques sont correctement saisies. Elle effectue quatre contrôles sur chaque variable alphanumérique. D'abord, elle vérifie que la valeur saisie fait partie des codes listés dans le questionnaire (autrement dit, elle effectue un contrôle de l'étendue). Ensuite, elle vérifie que les réponses sont saisies selon l'ordre alphabé-

tique (autrement dit, *ACG* et non *GAC*). Puis elle vérifie, pour les cas où les codes “Ne Sait Pas” ou “Aucun” (généralement la lettre “Y”) font partie de la réponse, qu’aucune autre réponse n’est présente (autrement dit, elle n’acceptera pas la réponse “*ACY*”). Enfin, elle vérifie, pour les cas où un code manquant (“ ? ”) fait partie de la réponse, qu’aucune autre réponse n’est présente (autrement dit, elle n’acceptera pas la réponse “*AC?*”)

L’application de saisie des données réarrange les valeurs saisie dans les variables alphanumériques, de sorte que chaque réponse est stockée à l’endroit exact qui définit son sous-item. Par exemple, pour la variable “*mn2*”, la réponse *ACG* sera réarrangée en *A C G* , où il y a un vide entre *A* et *C*, et entre *C* et *G*, et trois vides après *G*.

### FONCTIONS DEFINIES PAR L’UTILISATEUR

Une caractéristique très intéressante de CSPro réside dans le fait qu’il accepte que les programmes définissent leurs propres fonctions. De telles fonctions sont dénommées “*fonctions définies par l’utilisateur*”, et elles peuvent être très utiles. En particulier, elles évitent de réécrire un code fréquemment utilisé. Les fonctions définies par l’utilisateur sont toujours indiquées tout en haut d’une application CSPro. L’application de saisie des données *entry.app* comprend 14 fonctions définies par l’utilisateur. Il n’est pas nécessaire de modifier ces fonctions, mais vous devriez comprendre ce à quoi elles servent, si vous voulez comprendre l’application de saisie des données.

La fonction *valid* vérifie si la valeur de la variable est l’une des valeurs spéciales : incohérente, ne sait pas, manquante ou non applicable. Si la valeur de la variable est non applicable, la fonction *natozero* la change en “0”, lui permettant d’être ajoutée à une autre variable (Voir la procédure *cm9* pour un exemple de son utilisation).

Les sept fonctions définies par l’utilisateur suivantes (*zscoef*, *dabs*, *zspct*, *zseval*, *zscr*, *zsanth* et *agemth*) sont utilisées dans le Questionnaire des Enfants de Moins de Cinq Ans, pour calculer les scores anthropométriques qui se trouvent à la fin du module sur l’Anthropométrie. On recourt à la fonction *agemth* pour calculer l’âge de l’enfant en mois. Il est alors fait recours à la fonction *zsanth*. Cette fonction fait appel à *zseval*, *zscr* et à *zspct*. La fonction *zseval* fait appel à *zscoef*, et *zspct* fait appel à *dabs*. Vous ne rencontrerez ces fonctions que pour les variables anthropométriques, et si vous les rencontrez, vous saurez qu’elles sont entrain de calculer, puis de vérifier les scores anthropométriques.

Le code de la fonction *agemth* calcule l’âge de l’enfant en mois. Etant donné que l’anthropométrie est particulièrement sensible à l’âge, l’âge de l’enfant devra être basé sur l’âge de l’enfant en jours. Tout d’abord, le code calcule le nombre de jours écoulés entre le début de l’année et la naissance de l’enfant. Ensuite, il calcule le nombre de jours écoulés entre le début de l’année et la date de l’interview. La différence entre ces deux nombres de jours donne l’âge de l’enfant en jours. Ce chiffre est alors converti en âge de l’enfant en mois en le divisant par 30,4375 (qui est le nombre moyen de jours dans un mois au cours d’une période de quatre ans).

Pour des raisons de précision, l'âge de l'enfant en mois est calculé avec deux chiffres après la virgule.

Les fonctions *vdvalid*, *vdoi* et *vdob* vérifient que les dates de vaccination saisies dans le module sur la Vaccination sont cohérentes, ne sont situées, ni après la date de l'interview, ni avant la date de naissance de l'enfant, respectivement. La fonction *endmess* (abréviation pour "end message") affiche un message à la fin du questionnaire, qui demande à l'agent de saisie s'il désire passer en revue le questionnaire en cours ou passer au suivant. Enfin, la fonction *alphachk* effectue les contrôles sur les variables alphanumériques présentées en détail dans la sous-section précédente.

### **DATES ET CENTURY MONTH CODE**

Les programmes modèles (y compris l'application de saisie des données) utilisent les century month codes (CMC) pour la plupart des dates. Le CMC pour une date est le nombre de mois écoulés depuis décembre 1899. Par exemple, le CMC pour janvier 1900 est 1 ; le CMC pour mars 2000 est 1203. Le CMC pour une date est calculé de la manière suivante : on soustrait 1900 de l'année de la date, nombre d'années que l'on multiplie par 12, ensuite on y ajoute le mois de la date. Par exemple, le CMC pour mars 2000 est calculé comme suit :  $(2000-1900) \times 12 + 3$ .

L'application de saisie des données utilisent quatre fonctions pour simplifier le travail avec les century month codes. Deux de ces fonctions, *setlb* et *setub*, calculent les limites inférieures et supérieures, respectivement, pour le CMC de la date d'un événement. Les deux autres fonctions, *adjlba* et *adjuba*, ajustent les limites inférieures et supérieures, respectivement, de la date d'un événement (par exemple, l'anniversaire d'un enfant) lorsqu'un âge est également spécifié. Le Tableau 7.4 ci-dessous résume ces fonctions.

**Tableau 7.4**  
**Fonctions CPro Simplifiant le travail avec les Century Month Codes**

<pre>lcmc = setlb (month, year, minimum);</pre>	<p>Les arguments de la fonction sont : un mois (<i>month</i>), une année (<i>year</i>) et un CMC <i>minimum</i>. Si <i>year</i> et <i>month</i> sont tous deux valides, le CMC est calculé et rendu. Si <i>year</i> n'est pas valide, <i>minimum</i> est rendu. Si <i>month</i> n'est pas valide, le CMC pour le mois de janvier de <i>year</i> est rendu.</p>
<pre>ucmc = setub (month, year, maximum);</pre>	<p>Les arguments de la fonction sont : un mois (<i>month</i>), une année (<i>year</i>) et un CMC <i>maximum</i>. Si <i>year</i> et <i>month</i> sont tous deux valides, le CMC est calculé et rendu. Si <i>year</i> n'est pas valide, <i>maximum</i> est rendu. Si <i>month</i> n'est pas valide, le CMC pour le mois de décembre de <i>year</i> est rendu.</p>
<pre>t = adjlba (lcmc, ucmc, di, di, age); if t &lt; 0 then   e = errmsg (2171); else   lcmc = t; endif;</pre>	<p>Les arguments de la fonction sont : un CMC minimum pour la date (<i>lcmc</i>), un CMC maximum pour la date (<i>ucmc</i>), le CMC pour la date de l'interview (<i>di</i>), et un <i>age</i>. La fonction augmente <i>lcmc</i> en utilisant <i>age</i> et <i>di</i>. La fonction rend la date produite si elle est supérieure ou égale à <i>lcmc</i> et inférieure ou égale à <i>ucmc</i>. Si la date qui en est issue est inférieure à <i>lcmc</i>, la fonction rend <i>lcmc</i>. Si la date qui en résulte dépasse <i>ucmc</i>, la fonction rend -1.</p>
<pre>t = adjuba (lcmc, ucmc, di, di, age); if t &lt; 0 then   e = errmsg (2171); else   ucmc = t; endif;</pre>	<p>Les arguments de la fonction sont : un CMC minimum pour la date (<i>lcmc</i>), un CMC maximum pour la date (<i>ucmc</i>), le CMC pour la date de l'interview (<i>di</i>), et un <i>age</i>. La fonction réduit <i>ucmc</i> en utilisant <i>age</i> et <i>di</i>. La fonction rend la date produite si elle est supérieure ou égale à <i>lcmc</i> et inférieure ou égale à <i>ucmc</i>. Si la date qui en est issue est supérieure à <i>ucmc</i>, la fonction rend <i>ucmc</i>. Si la date qui en résulte est moindre que <i>lcmc</i>, la fonction rend -1.</p>

## RECEPTION DES QUESTIONNAIRES VENANT DU TERRAIN

Lorsque les questionnaires d'une grappe arrivent du terrain, l'archiviste doit contrôler le nombre de questionnaires sur la base de la fiche de contrôle de la grappe. Si des questionnaires sont manquants, l'archiviste doit contacter l'équipe de terrain pour essayer de retrouver ces questionnaires manquants. Si les questionnaires manquants ne peuvent pas être retrouvés, l'équipe de terrain devra les reprendre si c'est possible (autrement dit, si l'équipe de terrain se trouve toujours à proximité de la grappe). Si cela n'est pas possible, et que le questionnaire manquant est un Questionnaire Ménage, l'archiviste devra ajouter un questionnaire vierge (hormis pour les variables d'identification) avec "6" comme code de résultat, aux questionnaires de la grappe. Si le questionnaire manquant est un questionnaire femme ou enfant, l'archiviste devra remplacer le nombre total de questionnaires remplis sur la page de garde du questionnaire ménage auquel la femme ou l'enfant appartient.

L'arrivée des questionnaires du terrain devrait être enregistrée par l'archiviste sur la fiche de suivi de la grappe. Cette dernière est disponible en formats papier et électronique. Un exemple de la version papier figure à la fin de ce chapitre ; la version électronique du formulaire est disponible en passant par le menu du superviseur.

Avant de recevoir le moindre questionnaire venant du terrain, les numéros d'ordre des grappes doivent être listés par ordre numérique croissant en utilisant autant de fiches de suivi des grappes que nécessaire. Pour les questionnaires des grappes reçus du terrain, il faut enregistrer les informations relatives à chaque grappe dans l'espace correspondant. De plus, pour chaque grappe, l'archiviste devra vérifier les points suivants :

- Tous les questionnaires enregistrés sur la fiche de contrôle de la grappe sont là ;
- Les Questionnaires ménage, femme et enfant de chaque ménage enquêté sont ensemble ;
- Les questionnaires sont classés par ordre numérique selon les numéros du ménage ;
- Tous les questionnaires femme sont mis ensemble par ordre croissant - selon le numéro de ligne de la femme, suivi de tous les questionnaires des enfants, eux aussi classés par ordre croissant ;
- Tous les codes relatifs aux informations d'identification géographique et de l'interview sont remplis sur le Panneau d'Information sur le Ménage, au début du Questionnaire Ménage.

Une fois que le contrôle est achevé et que tout problème rencontré est résolu, l'archiviste doit enregistrer le nombre de questionnaires ménages, femmes et enfants sur la fiche de suivi de la grappe. Il/elle devra aussi créer une fiche de synthèse de la grappe. Cette dernière devrait être un morceau de papier cartonné à poser au-dessus du paquet de questionnaires pour afficher le numéro de la grappe en gros caractères. La fiche de synthèse de la grappe devra aussi résumer le nombre de questionnaires de la grappe, tel qu'indiqué au tableau 7.5 ci-dessous.

Une fois que la fiche de synthèse de la grappe aura été créée, le superviseur devra assigner la grappe à un agent de saisie (il faut noter que l'agent de saisie auquel la grappe aura été assignée devra saisir tous les questionnaires de cette grappe). L'archiviste devra ensuite enregistrer le nom de cet agent de saisie sur la fiche de suivi de la grappe.

**Table 7.5**  
**Exemple de Fiche de Suivi de la Grappe**

<b>101</b>	
<b>Nombre Total de Questionnaires Ménages</b>	_____
<b>Nombre Total de Ménages Remplis</b>	_____
<b>Nombre Total de Questionnaires Femmes Remplis</b>	_____
<b>Nombre Total de Femmes Remplis</b>	_____
<b>Nombre Total de Questionnaires Enfants Remplis</b>	_____
<b>Nombre Total d'Enfants Remplis</b>	_____

Chaque fois que les questionnaires sont déplacés de leur lieu de stockage pour la saisie ou l'apurement, ils doivent de nouveau être mis en liasse dans le même ordre et retournés au même endroit. Le fait de manipuler les questionnaires d'une manière systématique et stricte rendra les différentes phases du traitement des données plus rapides et plus efficaces.

Quand chaque grappe a été enregistrée sur le formulaire de suivi de la grappe, les questionnaires doivent être remis au superviseur du traitement des données, qui devra s'assurer qu'ils sont prêts à être envoyés à la saisie. Les éléments de contrôle spécifiques qui doivent être exécutés figurent dans le Tableau 7.6 ci-dessous. Une fois que le contrôle est achevé et que tout problème rencontré est résolu, le superviseur affecte la grappe à un agent de saisie.

### SAISIE DES DONNEES PRIMAIRES

Il est préférable de commencer la saisie des données pendant que les enquêteurs sont sur le terrain. Cela permet de repérer et de corriger des erreurs que certains enquêteurs ou équipes peuvent avoir commis. Les problèmes sérieux qui peuvent échapper à l'attention du superviseur de terrain peuvent ainsi être détectés rapidement, suffisamment tôt pour reprendre la formation du personnel de terrain et corriger les erreurs importantes. Les superviseurs de terrain ont pour responsabilité de contrôler la complétude et la cohérence des questionnaires, et de classer toutes les réponses que l'enquêteur n'a pas pu coder. Comme nous l'avons indiqué ci-dessus, le personnel de bureau doit faire le minimum de contrôle nécessaire des questionnaires, une fois que ces derniers arrivent au bureau. Aussitôt que les données d'une grappe arrivent au bureau central, la saisie des données doit commencer.

**Tableau 7.6**  
**Contrôle de Bureau**

1. Assurez-vous que tous les Questionnaires de ménage sont classés par ordre croissant par rapport au numéro du ménage dans la grappe.
2. Assurez-vous que tous les codes relatifs aux informations d'identification géographique et de l'interview sont complétés sur la page du Cadre d'information sur le ménage.
3. L'éligibilité pour l'interview de chaque membre du ménage doit être vérifiée dans la Feuille d'Enregistrement du Ménage. Pour être éligible aux modules de la femme, une personne doit être de sexe féminin et être âgée de 15-49 ans. Pour être éligible au module sur le travail des enfants, une personne doit être âgée de 5-14 ans (ou une tranche d'âges spécifique au pays). Pour être éligible au module sur la santé de l'enfant, une personne doit être âgée de moins de 5 ans. Pour chaque personne éligible, le code d'éligibilité encerclé devrait être celui de son numéro de ligne ou du numéro de ligne de la mère ou de la personne en charge de l'enfant (HL6-HL8, Feuille d'Enregistrement du Ménage). Pour tous les autres membres du ménage, les colonnes d'éligibilité devraient être laissées vides sur la Feuille d'Enregistrement du Ménage, et devront être saisies comme 00. Suivez les instructions figurant dans les Directives d'Apurement des Données (Annexe Six) pour les messages 0101-0132 afin de résoudre les problèmes éventuels.
4. Le nombre total de femmes et d'enfants éligibles enregistrés dans le Cadre d'Information sur le Ménage doit correspondre au nombre éligible enregistré sur la Feuille d'Enregistrement du Ménage.
5. Le nombre de questionnaires femme et enfants complètement remplis enregistrés dans le Cadre d'Information sur le Ménage doit correspondre au nombre de questionnaires ayant "1" (rempli) pour code résultat pour ce ménage.
6. Assurez-vous que tous les Questionnaires femme sont mis ensemble par ordre croissant selon le numéro de ligne de la femme (WM4), suivi de tous les Questionnaires des enfants mis ensemble par ordre croissant selon le numéro de ligne de l'enfant (UF4).

Le processus de saisie des données est guidé par une application de CSPro appelée *entry.ent*. Cette application effectue deux tâches principales. Premièrement, pour un ménage donné, il veille à ce que tous les questionnaires (et seulement ces questionnaires) sont saisis et que toutes les informations contenues dans ces questionnaires (et seulement ces informations) sont saisies. Plus particulièrement, l'application de saisie des données commence par saisir toutes les données du Questionnaire Ménage. Une fois que ces données ont été saisies, elle passe en revue la Feuille d'Enregistrement du Ménage et détermine quels sont les membres du ménage qui sont éligibles pour le Questionnaire Individuel Femme. Une fois qu'elle a identifié ces membres, l'application de saisie des données se positionne pour recevoir les données des questionnaires femmes l'un après l'autre, en suivant un ordre ascendant selon le numéro de ligne de la femme. Ensuite, l'application de saisie des données exécute la même procédure pour les membres du ménage qui

sont éligibles pour le questionnaire des enfants. Une fois que tous les questionnaires d'un ménage ont été saisis, l'ensemble des données du ménage sont sauvegardées et l'application passe au ménage suivant dans la grappe, s'il y en a.

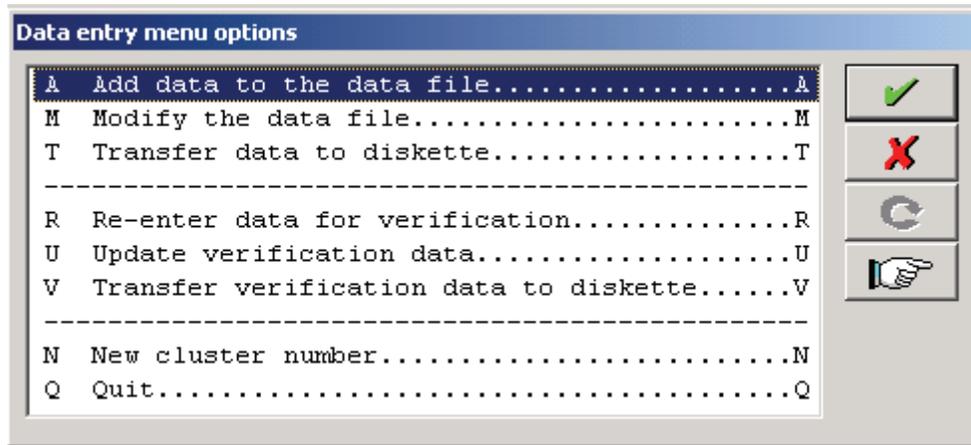
Etant donné que les données ne sont sauvegardées que lorsque les questionnaires ménage, ceux de toutes les femmes et de tous les enfants éligibles ont été saisis, les agents de saisie ne devraient pas quitter leurs machines alors qu'ils sont au milieu de la saisie d'un ménage. Avant de prendre un temps de repos ou d'arrêter le travail pour la journée, tous les questionnaires d'un ménage doivent être complètement saisis. En outre, il est recommandé, par précaution, que les données saisies sur le disque dur de l'ordinateur soient copiées sur l'ordinateur du superviseur ou sur une disquette. De plus, le superviseur doit copier – chaque soir – le contenu du répertoire *c:\mics* et de tous ses sous-répertoires dans le périphérique de stockage secondaire. Cette précaution permettra au superviseur de récupérer les données si son ordinateur a une panne.

En plus du contrôle sur le nombre de questionnaires qui sont saisis, l'application de saisie des données contrôle de manière rigoureuse la structure des sauts à l'intérieur d'un questionnaire. Autrement dit, elle n'attendra des réponses que pour les questions qui devaient être posées, en fonction des réponses aux questions précédentes. Par exemple, si la valeur 2 est saisie pour la variable *cm1* (c'est-à-dire que la femme n'a jamais accouché), l'application de saisie des données s'attendra ensuite à ce que la valeur qui sera immédiatement saisie à la suite sera celle de la variable *ma1*, sautant ainsi toutes les variables destinées seulement aux femmes ayant accouchées.

La seconde tâche de l'application de saisie des données consiste à minimiser les erreurs de saisie. Pour ce faire, l'application de saisie des données effectue des contrôles au fur et à mesure que les données sont saisies. Si une valeur saisie est située hors de l'intervalle des valeurs du questionnaire, ou si toute autre incohérence simple est détectée, l'application de saisie des données affiche un message d'erreur et demande à l'agent de saisie de corriger l'incohérence avant de poursuivre. Les corrections d'incohérences plus complexes, qui retarderaient grandement la saisie des données, ne sont pas faites lors de la saisie des données, mais plutôt lors de l'apurement secondaire des données.

### **MENU DE SAISIE DES DONNEES**

Le menu de saisie des données simplifie l'opération de saisie des données. Ce menu est créé par l'application de CSPro appelée *entry\_menu.bch* (il n'est pas nécessaire de modifier cette application). L'illustration ci-dessous montre le menu de saisie des données.



Les options A, M et T ne sont à utiliser que si l'agent de saisie est entrain de saisir les données dans le fichier primaire (fichier des données primaires). Les options R, U et V ne sont à utiliser que si l'agent de saisie est entrain de procéder à la vérification des données (fichier des données de vérification). Quant aux options N et Q, elles peuvent être utilisées en toute circonstance.

L'option A permet à l'agent de saisie d'ajouter des cas au fichier primaire de données, l'option M lui permet de modifier les cas déjà enregistrés dans le fichier primaire de données, et l'option T lui permet de copier le fichier de données primaires complet dans une disquette, de sorte que celui-ci puisse être transféré sur l'ordinateur du superviseur. L'option R permet à l'agent de saisie d'ajouter des cas au fichier des données de vérification, l'option U lui permet de modifier les cas déjà enregistrés dans le fichier de travail, et l'option V lui permet de copier le fichier des données de vérification complet dans une disquette, de sorte que celui-ci puisse être transféré sur l'ordinateur du superviseur.

L'option N permet à l'agent de saisie de modifier le numéro de grappe, de façon à pouvoir saisir les données d'une autre grappe. L'option Q permet de sortir du menu de saisie des données.

### STRUCTURE DES REPERTOIRES SUR UN ORDINATEUR DE SAISIE DES DONNEES

Sur un ordinateur de saisie des données, tous les fichiers et programmes sont situés dans le répertoire *c:\mics\CSPro* ou dans l'un de ses sous-répertoires. Les sous-répertoires sont nommés *data*, *dicts*, *entry* et *veri*. Le répertoire *data* contient tous les fichiers des données primaires qui ont été saisis sur cet ordinateur. Le répertoire *dicts* contient tous les dictionnaires des données. Le répertoire *entry* contient l'application de saisie des données et l'application qui crée le menu de saisie des données. Le répertoire *veri* contient tous les fichiers des données de vérification qui ont été saisis sur cet ordinateur.

La structure des répertoires sur l'ordinateur de l'agent de saisie est la suivante :

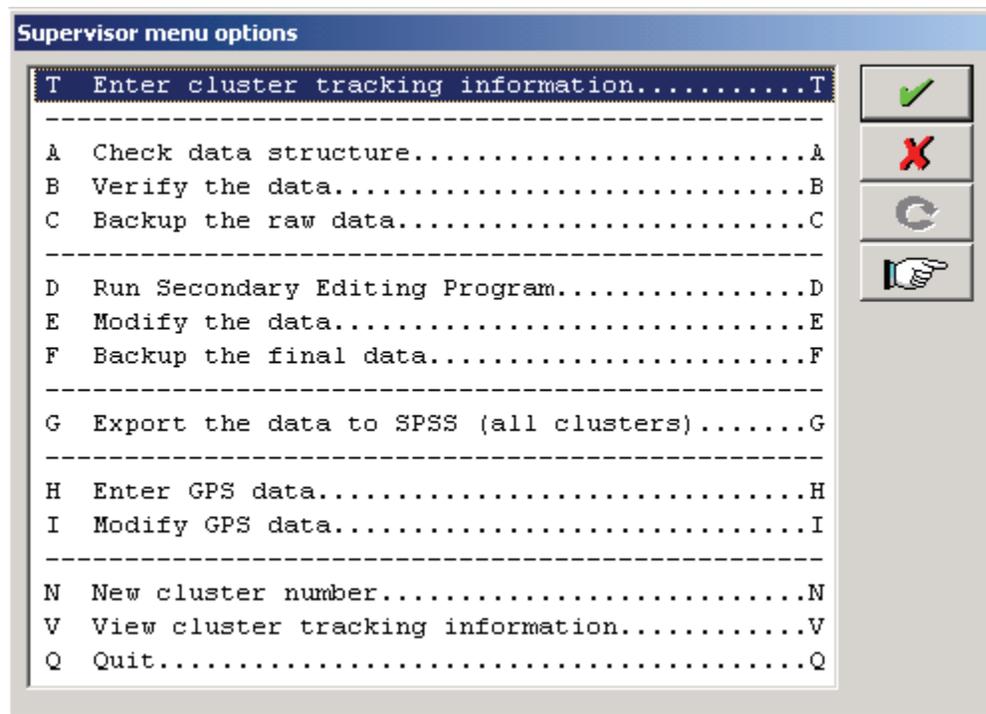
MICS

CSPRO

DATA	Répertoire des données primaires
DICTS	Répertoire des dictionnaires des données
ENTRY	Répertoire de l'application de saisie des données
VERI	Répertoire des données de vérification

## MENU DU SUPERVISEUR

Le menu du superviseur aide le superviseur du traitement des données à gérer le système de traitement des données de la MICS3. Ce menu est lancé par l'exécution de l'application CPro *super\_menu.pff*. Pour s'exécuter, le menu commence par demander le numéro de grappe. Une fois que le numéro de grappe est saisi, l'illustration ci-dessus affiche le menu du superviseur, suivi par une présentation résumée de chaque option.



### OPTION T : SAISIE DES INFORMATIONS DE SUIVI DE LA GRAPPE

Cette option affiche la version électronique du formulaire de suivi du traitement de la grappe, de sorte que le superviseur du traitement des données puisse saisir les informations de suivi du traitement de la grappe. Des informations devraient être saisies lorsque les questionnaires d'une

grappe arrivent du terrain, lorsque la grappe est affectée à un agent de saisie pour la saisie primaire des données, et lorsque la grappe est affectée à un agent de saisie pour la saisie de vérification des données. Les autres informations de la version électronique du formulaire sont mises à jour automatiquement au fur et à mesure que le superviseur avance dans le menu du superviseur.

#### **OPTION A : CONTROLE DE LA STRUCTURE DES DONNEES**

Cette option vérifie la structure du fichier de données de la grappe, en exécutant l'application *check.bch*. Après vérification de la structure du fichier de données, *check.bch* sort un fichier de résultat (qui s'affiche automatiquement sur l'écran) qui présente de manière résumée le nombre de chaque type de questionnaires et montre le nombre de questionnaires qui ont été remplis. Le superviseur du traitement des données doit vérifier ces informations en les comparant avec celles du formulaire de suivi du traitement des données, et s'assurer que les deux sources s'accordent. Si ce n'est pas le cas, le superviseur doit rechercher le problème (par exemple, l'agent de saisie aurait pu omettre de saisir un ménage) et le résoudre en contrôlant soigneusement les questionnaires de la grappe. Une fois le problème résolu (en mettant à jour soit le fichier de données, soit le formulaire de suivi du traitement des données de la grappe), le superviseur du traitement doit relancer le programme *check.bch*. Ce n'est que lorsque *check.bch* produit le même nombre de questionnaires que celui du formulaire de suivi que le superviseur devra affecter les questionnaires à un second agent de saisie pour des fins de vérification de la saisie des données.

#### **OPTION B : VERIFICATION DES DONNEES**

Cette option permet de comparer le fichier des données primaires d'une grappe à son fichier de données de vérification en utilisant l'outil de comparaison de CPro. Si la moindre différence entre les deux fichiers des données est détectée, elle sera affichée sur l'écran. Ce résultat devrait être imprimé et remis aux agents de saisie responsables de la saisie des données de cette grappe. En travaillant en équipe de deux, les agents de saisie devront revoir les questionnaires et déterminer la valeur correcte pour chacun des cas où les deux fichiers de données diffèrent. Une fois qu'ils auront déterminé les valeurs correctes, chaque agent de saisie devra mettre à jour son fichier de données. Cela fait, les deux fichiers devront de nouveau être comparés. Si n'y a plus de différence entre les fichiers, le traitement des données de la grappe peut se poursuivre.

#### **OPTION C : ARCHIVAGE DES DONNEES BRUTES**

Cette option archive les données brutes en copiant le fichier des données primaires dans le sous-répertoire *back-up* de l'ordinateur du superviseur. Cette option doit être lancée après que les contrôles et les vérifications de structures auront été terminés, et avant qu'aucun apurement secondaire ne soit effectué. Les données brutes sont archivées pour documenter l'état des données avant leur apurement.

**OPTION D : EXECUTER UN SECOND PROGRAMME DE CONTROLE**

Cette option vérifie les incohérences plus complexes en exécutant l'application *editing.bch*. Le résultat de ce programme s'affiche sur l'écran, et il devrait être imprimé s'il contient des messages d'erreurs. Si c'est le cas, la liste des incohérences et les questionnaires de la grappe devraient être remis à un agent de vérification. L'agent de vérification, sur la base de ses connaissances du questionnaire et du manuel d'apurement des données en Annexe Sept, devra résoudre chacune des incohérences listées. Lorsque les agents de vérification auront achevé leur travail, ils rendront la liste des incohérences et des actions requises au superviseur du traitement des données, qui aura à mettre en œuvre personnellement les modifications requises (voir Option E ci-dessous). Cela fait, le superviseur relance *editing.bch*. S'il n'y a plus de message d'erreur, le traitement des données peut se poursuivre ; par contre, si des messages d'erreurs demeurent, une liste de ces messages et les questionnaires devraient être remis à l'agent de vérification pour un contrôle plus approfondi. Dans certains cas, les messages seront considérés comme acceptables par l'agent de vérification, et il ne sera pas nécessaire d'aller plus loin dans la correction des données. Si les seuls messages qui continuent à s'afficher sont ceux d'erreurs que l'agent de vérification trouve acceptables, alors on peut considérer le processus d'apurement secondaire des données comme terminé.

**OPTION E. MODIFICATION DES DONNEES**

Cette option exécute l'application *entry.ent* qui permet au superviseur du traitement des données de mettre en œuvre les modifications requises par les agents de vérification. Après avoir utilisé cette option, le superviseur devra retourner à l'Option D pour s'assurer que les modifications dans les données ont effectivement permis de corriger les incohérences comme souhaité, et qu'aucune nouvelle incohérence n'a été introduite.

**OPTION F. ARCHIVAGE DES DONNEES APUREES**

Cette option copie le fichier des données apurées dans le sous-répertoire *final* de l'ordinateur du superviseur du traitement des données. Les données stockées dans ce répertoire seront par la suite concaténées, puis exportées vers SPSS.

**OPTION G. EXPORTATION DES DONNEES VERS SPSS**

Cette option concatène tous les fichiers de données du répertoire *final* en un seul fichier, qui est ensuite exporté par l'exécution de l'application *export.bch*. Cette application crée 4 fichiers textes ASCII et le fichier de syntaxe SPSS correspondant pour chacun de ces fichiers textes. L'un des fichiers textes contient les ménages, l'autre contient les membres des ménages, le troisième contient les femmes et le dernier contient les enfants.

**OPTION H. SAISIE DES DONNEES DU SYSTEME DE POSITIONNEMENT GEOGRAPHIQUE (GPS)**

Cette option permet au superviseur du traitement des données de saisir les données de localisation GPS en exécutant l'application *gpsentry.ent*. Contrairement au programme de saisie des données primaires, cette application permet au superviseur du traitement des données de saisir autant de grappes en même temps qu'il le désire. Comme mesure de prévention des erreurs de saisie, l'application exige du superviseur que les données GPS soient saisies deux fois.

**OPTION I. MODIFICATION DES DONNEES GPS**

Cette option permet au superviseur du traitement des données de modifier les données de localisation GPS en lançant l'application *gpsentry.ent*.

**OPTION V. AFFICHAGE DES INFORMATIONS DE SUIVI DE LA GRAPPE**

Cette option affiche, pour toutes les grappes, les informations contenues sur le formulaire de suivi du traitement de la grappe.

**OPTION N. SELECTIONNER UNE NOUVELLE GRAPPE**

Cette option change le numéro de la grappe, de sorte que le superviseur du traitement peut passer au traitement d'une nouvelle grappe.

**STRUCTURE DU REPERTOIRE SUR L'ORDINATEUR DE TRAITEMENT DES DONNEES DU SUPERVISEUR**

Le système de traitement des données de la MICS3 utilise une structure de répertoire spécifique sur les ordinateurs des agents de saisie et sur l'ordinateur du superviseur de traitement des données. La structure des ordinateurs de saisie des données est présentée ci-dessous. Sur l'ordinateur du superviseur, tous les fichiers et programmes relatifs à SPSS sont stockés dans le répertoire *c:\mics\spss* ou dans l'un de ses sous-répertoires. Sur l'ordinateur du superviseur, tous les fichiers et programmes relatifs à CSPro sont stockés dans le répertoire *c:\mics\CSPro* ou dans l'un de ses sous-répertoires. Les sous-répertoires sont dénommés *backup*, *dicts*, *entry*, *export*, *gps*, *final*, *raw*, *super* et *weights*.

La structure des répertoires de l'ordinateur du superviseur est la suivante :

MICS

CSPRO	
BACKUP	Répertoire Back-up
DICTS	Répertoire des dictionnaires des données
ENTRY	Répertoire de l'application de saisie des données
EXPORT	Répertoire de l'application Export
FINAL	Répertoire des données apurées
GPS	Répertoire de l'application de saisie des données GPS
RAW	Répertoire des données brutes
SUPER	Répertoire de l'application de contrôle du superviseur
WEIGHTS	Répertoire de l'application des poids de l'échantillon

Le répertoire *back-up* contient des copies de sauvegarde des fichiers de données dont la structure a été contrôlée et qui ont été vérifiés, mais non apurés. Le répertoire *dicts* contient tous les dictionnaires de données. Le répertoire *entry* contient l'application de saisie des données, ainsi que l'application qui crée le menu de saisie des données. Le répertoire *export* contient les programmes qui permettent d'exporter les données de CSpRo vers SPSS. Le répertoire *final* contient des copies de sauvegarde des fichiers de données dont la structure a été contrôlée, qui ont été vérifiés et apurés. Le répertoire *raw* contient les fichiers de données qui ont été transférés des ordinateurs des agents de saisie. Le répertoire *super* contient les applications qui effectuent les contrôles de structure, la vérification et l'apurement secondaire, ainsi que l'application qui crée le menu du superviseur. Le répertoire *weights* contient la feuille de calcul qui permet de calculer les poids de l'échantillon.

## CONTROLE DE STRUCTURE

Il est crucial que les données soient structurellement solides. Le programme de saisie des données met en application la plupart des cohérences structurelles, mais il ne peut tout vérifier sans être extrêmement contraignant. Par conséquent, il est nécessaire d'exécuter un programme de contrôle de structure une fois que la saisie des données est terminée. Le programme de contrôle de structure s'assure que le nombre de questionnaires figurant dans le fichier de données correspond au nombre de questionnaires réceptionnés du terrain, et il effectue quelques contrôles supplémentaires sur la structure d'un questionnaire individuel.

Le programme de contrôle de structure est dénommé *check.bch*. Cette section met davantage l'accent sur ce que le programme peut faire, plutôt que sur la manière dont il le fait. Le programme est complexe et il ne se prête pas à des explications faciles. La meilleure manière de comprendre la logique du programme est de l'étudier attentivement, une fois que vous avez compris ce que le programme est susceptible de faire. Un exemple des résultats majeurs issus de *check.bch* est montré ci-dessous.

MICS Data Structure Check  
Cluster: 3

Households			Women				Children			
Total	Comp	Incomp	Eligible	Interviewed	Eligible	Interviewed	Eligible	Interviewed	Eligible	Interviewed
			HH12 FOUND	HH13 FOUND	HH14 FOUND	HH15 FOUND				
2	1	1	5	5	4	4	4	4	3	3

Le premier bloc de résultats est un résumé du nombre total de ménages et de leurs codes de réponses. Le second bloc de résultats présente les résultats de quatre décomptes du nombre de femmes éligibles. Les décomptes des colonnes *hh12* et *hh13* représentent les nombres de femmes éligibles et interviewées selon le Panneau d'Information du Ménage. Les décomptes dans les colonnes “*found*” représentent les nombres de questionnaires femmes et de questionnaires femmes remplis, respectivement, dans le fichier de données.

Le troisième bloc de résultats présente les résultats de quatre décomptes du nombre d'enfants de moins de cinq ans éligibles. Les décomptes des colonnes *hh14* et *hh15* représentent les nombres d'enfants éligibles et interviewés selon le Panneau d'Information du Ménage. Les décomptes dans les colonnes “*found*” (c'est-à-dire “*trouvé*”) représentent les nombres de questionnaires d'enfants et de questionnaires d'enfants remplis, respectivement, dans le fichier de données.

Le résultat de *check.bch* devrait être imprimé par le superviseur, et les informations contenues devraient être comparées au formulaire de suivi du traitement de la grappe. Si une différence est constatée entre les eux décomptes de questionnaires, le superviseur et l'agent de saisie devront se servir du listing des erreurs et des questionnaires de la grappe pour identifier la cause du problème de structure. Lorsque ces causes auront été identifiées, elles devront être corrigées par l'agent de saisie. L'application de vérification de la structure devra alors être exécutée de nouveau, afin de s'assurer que le problème a bien été réglé sans en créer un nouveau. Ce ne sera que lorsque les décomptes générés par *check.bch* correspondent à ceux du formulaire de suivi du traitement de la grappe que la saisie des données de vérification pourra commencer.

L'application *check.bch* produit une liste de tous les ménages de la grappe. Chacun des numéros de ménage et des codes de résultat est affiché, en même temps que le décompte des questionnaires de femmes et d'enfants éligibles, si l'interview du ménage a été remplie (c'est-à-dire, le code de résultat du ménage est égal à 1). Un exemple de résultat d'un ménage est donné ci-dessous.

MICS Data Structure Check  
Household: 1  
Result: 1

Women				Children			
Eligible	Interviewed	Eligible	Interviewed	Eligible	Interviewed	Eligible	Interviewed
HH12 FOUND	HH13 FOUND	HH14 FOUND	HH15 FOUND				
4	4	3	3	2	2	1	1

Le listing des ménages peut être d'une grande utilité pour l'identification de la source du problème au niveau de la grappe. Supposons, par exemple, que la feuille de suivi du traitement de la grappe liste 20 ménages dans la grappe, mais que seuls 19 sont retrouvés dans le fichier de don-

nées. En comparant le listing des ménages aux questionnaires de la grappe, on peut aisément retrouver le ménage qui n'a pas été saisi.

### VERIFICATION

La vérification des données ayant fait l'objet d'une double saisie est effectuée par une application de comparaison de CSPro. Cette application de comparaison est dénommée *compare.cmp*. Elle contient une liste de toutes les variables (items). Tel que le programme est actuellement configuré, les items cochés seront comparés lors de la vérification et les items non cochés ne seront pas comparés au cours de la vérification. Une seule variable n'est pas cochée (le code de l'agent de saisie, *hh16*), et il est conseillé qu'**aucune autre variable** ne soit non cochée puisque des différences sur toutes ces autres variables sont susceptibles de compromettre la qualité des données.

L'application de comparaison compare le fichier des données primaires (qui a été copié sur la machine du superviseur) et le fichier des données de vérification (qui a été copié sur une disquette, ou sur un serveur, dans le cas où un réseau est utilisé), et il ressort une liste des différences, s'il y en a. si aucune différence n'existe, le superviseur devrait archiver les données brutes, puis passer à l'apurement secondaire des données.

Si des différences existent, la liste de celles-ci devrait être imprimée et remise aux deux agents de saisie. Les agents de saisie devront alors se servir de la liste des différences et des questionnaires de la grappe pour contrôler chacune des différences relevées, et inscrire sur la liste les fichiers qui devront être corrigés. Lorsque toutes les différences auront été examinées, les agents de saisie devront corriger toutes les erreurs de leurs fichiers. Ensuite, ils devront recopier les fichiers de données sur leurs disquettes (ou sur le serveur), et les fichiers seront comparés de nouveau. Ce processus se poursuivra jusqu'à ce que les deux fichiers soient identiques.

### APUREMENT SECONDAIRE

L'expérience a montré que les incohérences simples peuvent être facilement identifiées et corrigées pendant la saisie des données. Mais les incohérences plus complexes devraient être résolues à travers un examen méticuleux du questionnaire. La meilleure manière de procéder à ce type de vérification des incohérences est d'en faire une étape distincte, avec les erreurs transcrites sur un document imprimé qui peut être utilisé pour marquer les corrections. Cette étape est connue sous le nom d'apurement secondaire.

Le programme d'apurement secondaire est dénommé *editing.bch*. Il s'acquiesce d'une longue liste de contrôles de cohérences (par exemple, est ce que les âges et les dates de naissance sont cohérents ?) et produit une liste des incohérences détectées dans le fichier de données primaires. Le superviseur du traitement des données devra imprimer cette liste et la remettre, accompagnée des

questionnaires de la grappe, aux agents de vérification. L'agent de vérification procède alors à une revue de la liste des erreurs et des réponses du questionnaire. En utilisant les directives d'apurement des données (qui se trouvent dans l'Annexe Sept) et sur la base de sa maîtrise du questionnaire, l'agent de vérification aura alors, soit à rédiger sur le listing des erreurs la correction à effectuer, soit à écrire qu'il n'y a aucune action à prendre. Lorsque l'agent de vérification aura passé en revue tous les messages d'erreurs, l'un après l'autre, il devra remettre le listing d'erreurs annoté au superviseur du traitement des données. Ce dernier aura alors à intégrer les modifications proposées dans les données brutes. Une fois cela fait, le superviseur du traitement des données relance le programme d'apurement. Ce n'est que lorsque le programme d'apurement ne produit plus de message d'erreur que le traitement de la grappe peut continuer.

Trois aspects du processus d'apurement des données revêtent une importance cruciale. Premièrement, pour chaque erreur affichée, l'agent de vérification devra retourner au questionnaire concerné et l'examiner attentivement. En second lieu, l'agent de vérification doit se référer aux directives d'apurement avant de proposer une solution au problème. Troisièmement, le processus d'apurement des données devra être répété autant de fois que nécessaire, jusqu'à l'élimination de toutes les erreurs.

Une fois que cette troisième tâche est achevée, le superviseur du traitement des données peut archiver les données apurées. Ces données peuvent maintenant être considérées comme propres, et elles peuvent être utilisées pour la création des fichiers d'analyse.

## EXPORTATION DES DONNEES VERS SPSS

Lorsque le traitement des données primaires est terminé, vous avez en votre possession un fichier de données propre pour chacune des grappes de votre échantillon. Alors que le traitement des données primaires est effectué avec CPro, le traitement secondaire des données est essentiellement effectué avec SPSS. La première étape du traitement secondaire des données consiste donc à convertir les données du format CPro au format SPSS. Cela est fait en utilisant l'option 'Export the data to SPSS<sup>3</sup>' du menu du superviseur.

Lorsque vous sélectionnez cette option, tous les fichiers de données du sous-répertoire *final* (autrement dit, tous les fichiers de données qui ont été vérifiés, contrôlés et apurés) sont concaténés en un seul fichier dénommé *all.dat*. Ce fichier de données est ensuite exporté vers SPSS par l'application *export.bch*. Cette application crée quatre fichiers de données ASCII (*mych.dat*, *myhh.dat*, *myhl.dat* et *mywm.dat*) et quatre fichiers de syntaxe SPSS (*mych.sps*, *myhh.sps*, *myhl.sps* et *mywm.sps*) dans le répertoire *c:\mics\spss*. Lorsque les fichiers de syntaxe SPSS liront les fichiers de données ASCII dans SPSS, ils ne les sauvegarderont pas. Si l'on veut que les fichiers de syntaxe sauvent les données sous le format SPSS, la commande SPSS

---

<sup>3</sup> Exporter les données vers SPSS.

```
save outfile = `filename.sav`.
```

doit être ajoutée à la fin de chaque fichier de syntaxe. Le mot *filename* devrait être remplacé par *hh*, *hl*, *wm* ou *ch*, selon le type du fichier de données. Une fois que cette commande aura été correctement modifiée et ajoutée à chacun des fichiers de syntaxe, l'exécution des fichiers de syntaxe SPSS permettra la création des fichiers de données SPSS *hh.sav*, *hl.sav*, *wm.sav* et *ch.sav*.

## CREATION D'UN FICHIER D'ANALYSE

La structure du fichier de données lors du traitement des données primaires simplifie le processus de saisie des données. Toutefois, cette structure n'est pas optimale pour l'analyse des données collectées et c'est la raison pour laquelle la première tâche à exécuter après que les données ont été transférées vers SPSS consiste à recoder les variables, pour faciliter l'analyse et la rendre plus efficace. Cette tâche est connue sous le nom de création d'un fichier d'analyse. Cette section présente en détail les étapes de la création des fichiers d'analyse de la MICS3. Les fichiers d'analyse qui résulteront de l'application de ces étapes peuvent être utilisés par les plans de tabulation modèles et peuvent tout à fait être disséminés auprès de chercheurs.

### RECODIFICATION DES VARIABLES

Les programmes *makehl.sps*, *makewm.sps* et *makech.sps* recodent les variables existantes pour en créer de nouvelles. Les variables qui sont utilisées dans plusieurs tabulations sont recodées dans ces programmes avant d'être sauvées ; toutes les autres recodifications sont effectuées dans les programmes de tabulation et sont temporaires.

La recodification de la plupart des variables utilise les commandes standards de SPSS, et ne sera donc pas présentée ici. Toutefois, il y a une approche couramment utilisée qui doit être expliquée : la recodification des variables en 0 ou 100. Cette recodification inhabituelle n'est faite qu'à des fins d'illustration. Lorsque SPSS affiche des pourcentages dans un tableau, il affiche toutes les catégories possibles de la variable. Pour plusieurs tableaux du plan de tabulation, ce n'est qu'une seule catégorie qui nous intéresse. Si nous affectons la valeur 100 à cette catégorie et la valeur 0 à toutes les autres catégories, la moyenne de cette variable est la proportion de personnes enquêtées dans cette catégorie. Ainsi, le fait de demander à SPSS d'afficher la moyenne de cette nouvelle variable aura comme résultat de n'afficher que le pourcentage qui nous intéresse.

Par exemple, la variable *hal* enregistre le fait pour une femme d'avoir entendu parler ou non du SIDA. Elle prend la valeur 1 dans le cas où la femme a déjà entendu parler du SIDA, et la valeur 0 dans le cas contraire. Si nous voulons afficher le pourcentage de femmes ayant déjà entendu parler du SIDA dans le programme *makewm.sps*, la variable *hal* est recodée dans la variable *aids*. La variable *aids* prend la valeur 100 dans le cas où la femme a déjà entendu parler du

SIDA, et la valeur 0 sinon. Pour comprendre pourquoi il en est ainsi, regardez l'exemple ci-dessous :

Femmes ayant entendu parler du SIDA	10
Nombre total de femmes	20
Pourcentage de femmes ayant entendu parler du SIDA	$10 / 20 \times 100 = 50$
Moyenne de la variable <i>aids</i>	$(10 \times 100 + 10 \times 0) / 20 = 10 \times 100 / 20 = \mathbf{10/20 \times 100 = 50}$

### CALCUL ET AJOUT DES POIDS D'ECHANTILLON

Si des bases de sondage distinctes ont été utilisées pour différentes régions (ou domaines) pour le tirage au premier degré, l'échantillon national n'a pas été choisi avec probabilité proportionnelle à la taille. Cela peut également se produire si vous avez procédé à une stratification selon certains autres facteurs (par exemple, urbain/rural ou quartiers pauvres/autres quartiers) et utilisé des taux de sondage (proportions) différents dans chaque strate. Ces échantillons ne sont pas auto-pondérés, et vous devez pondérer votre échantillon quand vous calculez les estimations nationales. Cela signifie que vous devez vous assurer que la part de chaque sous-échantillon distinct – par exemple, chaque région (domaine) – contribue à l'échantillon national seulement pour la part qu'il aurait eu si l'échantillon de l'enquête, au niveau national, avait été tiré avec une probabilité proportionnelle à la taille.

Si votre échantillon n'est pas auto-pondéré, vous devez calculer les poids d'échantillon et les ajouter à vos fichiers d'analyse. Cette tâche est réalisée en utilisant la feuille de calcul *weights.xls* et les programmes SPSS *weights\_table.sps*, et *weights\_merge.sps*. La feuille de calcul est utilisée pour calculer les poids de l'échantillon. Elle comporte deux feuilles de travail, *calculations* et *output*. La feuille de travail *calculations* exécute les calculs. Quant à la feuille de travail *output*, elle ne contient que les poids de l'échantillon et une liste des numéros de grappes ; ce format est d'une grande utilité pour la lecture des données dans SPSS. Le programme *weights\_table.sps* produit les données requises pour le calcul des poids de l'échantillon. Le programme *weights\_merge.sps* ajoute les poids de l'échantillon appropriés aux fichiers d'analyse. Le programme *weights.sps*, que vous n'aurez jamais à exécuter directement, décrit la structure des données dans la feuille de travail de sortie des résultats.

La procédure de calcul des poids de l'échantillon et de leur ajout à vos fichiers d'analyse peut être subdivisée en sept étapes :

**Etape 1 :** Ajustez le nombre de lignes des feuilles de travail *calculations* et *output*, de sorte qu'il n'y ait qu'une seule ligne par grappe de votre enquête. Une fois que vous aurez fini de supprimer ou d'ajouter des lignes, veuillez à vérifier que cette opération n'a pas modifiée le nombre total de lignes de la feuille de travail *calculations*.

**Etape 2 :** Saisissez les poids qui ont été construits dans le plan de sondage dans la feuille *weights.xls*. Si vos poids sont variables entre les grappes d'une strate ou d'un domaine particulier, il vous faudra remplir à la fois la colonne des indices de pondération au niveau grappe et la colonne des indices de pondération au niveau strate (ou domaine), cela à partir des informations qui vous seront fournies par votre spécialiste en sondage. Si vos poids sont variables entre les strates (ou domaines), mais ne le sont pas entre les grappes au sein d'une strate (autrement dit, l'échantillon est auto-pondéré à l'intérieur des strates ou domaines), saisissez la valeur 1 dans la colonne des indices de pondération au niveau grappe et remplissez la colonne des indices de pondération au niveau strate (ou domaine) en vous servant des informations qui vous seront fournies par votre spécialiste en sondage.

**Etape 3 :** Procédez à une mise à jour de la définition des strates (ou domaines) sur les lignes 3 à 10 du programme *weights\_table.sps*. Les programmes standards supposent que les strates sont formées par les combinaisons de toutes les zones (c'est-à-dire urbain et rural) et régions, et qu'il y a quatre régions (le programme devrait être modifié pour refléter les strates ou domaines qui sont utilisés dans votre échantillon).

**Etape 4 :** Exécutez le programme *weights\_table.sps*.

**Etape 5 :** Copiez les informations du tableau et collez les dans la feuille de travail *calculations* de *weights.xls*. Lorsque cette étape sera terminée, *weights.xls* calculera automatiquement les poids de sondage de l'échantillon.

**Etape 6 :** Sauvegardez votre feuille de travail *output* sous la forme d'un fichier séparé par un point (\*.csv), sous le nom *weights.csv* dans le répertoire *c:\mics\weights*.

**Etape 7 :** Exécutez le programme *weights\_merge.sps*. Une fois que vous aurez achevé la septième étape, pensez à vérifier la feuille de sortie des résultats pour voir si elle comporte des messages d'erreurs, et ouvrez les fichiers d'analyse pour vous assurer que les poids ont été correctement ajoutés aux fichiers de données SPSS.

## CALCUL ET AJOUT DE L'INDICE DE RICHESSE

Parmi les variables socio-économiques utilisées, le plan de tabulation de la MICS3 inclut l'indice de richesse (encore appelé indice de bien-être économique). Cet indice de richesse est calculé par le programme *wealth.sps* qui crée le fichier de données *wealth.sav* contenant les variables d'identification, une variable contenant le score de richesse de chaque ménage et une variable contenant l'indice de richesse de chaque ménage. En premier lieu, le programme *wealth.sps* sort des fréquences de toutes les variables du ménage qui concernent la richesse ou les biens possédés. Ensuite, il recode les variables décrivant les biens du ménage ou de l'individu, transformant ces variables en variables dichotomiques. Puis le programme utilise l'analyse factorielle (plus précisément l'analyse en composantes principales) pour calculer un score de richesse pour chaque ménage. Pour terminer, il utilise le score de richesse pour créer les quintiles de richesse

(c'est-à-dire l'indice de bien être économique) qu'il sauvegarde dans un fichier de données SPSS. Le choix des variables à prendre en compte dans l'analyse factorielle est crucial, et ne devrait être fait qu'après un examen minutieux des fréquences produites. Les informations sur les variables incluses dans l'analyse sont disponibles sur le site Web [www.childinfo.org](http://www.childinfo.org). Une fois que l'indice de richesse aura été calculé, l'exécution du programme *wealth\_merge.sps* l'adjoindra à vos fichiers d'analyse. Pensez à vérifier la feuille de sortie des résultats pour voir si elle comporte des messages d'erreurs, et ouvrez les fichiers d'analyse pour vous assurer que les valeurs de l'indice de richesse ont été correctement ajoutées aux fichiers de données SPSS.

### AJOUT DES COORDONNEES GPS

Certaines enquêtes MICS3 recueilleront les coordonnées GPS de leurs grappes lors du travail de terrain. Ces coordonnées GPS, qui donnent la localisation précise des grappes, peuvent être utilisées pour adjoindre une série de données géographiques (par exemple des données pluviométriques) à une série de données MICS3. Si votre enquête recueille des coordonnées GPS, vous voudriez probablement les intégrer à vos fichiers d'analyse. Cette tâche est réalisée par deux programmes SPSS : *gps.sps* et *gps\_merge.sps*. Le programme *gps.sps*, que vous n'aurez jamais à exécuter directement, décrit la structure du fichier de données *gps.dat* (qui est créé par l'application de saisie des données de CSPro *gpsentry.bch*). Si vous avez modifié le dictionnaire CSPro *gps.dic*, vous devriez mettre à jour *gps.sps* pour prendre en compte vos modifications. Le programme *gps\_merge.sps* adjoind les coordonnées GPS aux fichiers d'analyse. Vous n'avez pas besoin de modifier ce programme. Pour adjoindre les coordonnées GPS à vos fichiers d'analyse, exécutez *gps\_merge.sps*. Pensez à vérifier la feuille de sortie des résultats pour voir si elle comporte des messages d'erreurs, et ouvrez les fichiers d'analyse pour vous assurez que les coordonnées GPS ont été correctement ajoutées aux fichiers de données SPSS.

## PREPARATION DE LA TABULATION

Le Chapitre 8 décrit en détail le processus d'analyse des données et de préparation des rapports. Faire tourner les tableaux est une composante importante de cette activité. Un plan de tabulation modèle (Chapitre 8 et Annexe Six) et un programme de tabulation pour SPSS sont joints à ce manuel. Il est prévu un programme de tabulation SPSS pour chaque table dans le plan de tabulation de la MICS3. Chaque nom de programme est composé de la lettre "t" suivie du numéro du tableau dans le plan de tabulation. Par exemple, le programme *t1.sps* crée le Tableau 1 du plan de tabulation.

Avant de faire tourner les tableaux pour le rapport, il est essentiel de produire une série de distributions de fréquences (non pondérées) pour chaque variable dans le fichier des données. Les fréquences doivent être vérifiées pour y repérer les valeurs anormales, celles qui sont en dehors de l'étendue de la plupart des réponses et celles qui semblent être des réponses invraisemblables à des questions qui se rapportent à l'enquête. Par exemple, une réponse de '53' à la question sur le nombre d'heures de travail effectué par un enfant au cours de la dernière semaine semble à la

fois invraisemblablement élevée et trop précise. Les informations d'identification de tels cas doivent être écrites, et les questionnaires concernés devraient être vérifiés par rapport aux valeurs contenues dans le fichier des données.

### REVISION DES PROGRAMMES MODELES

Chaque programme de tabulation devrait être examiné avec soin. Il est important de vérifier que les variables utilisées dans le programme de tabulation figurent bien dans votre fichier de données. Si ce n'est pas le cas, vérifiez si cette variable est de première importance, ou plutôt a une importance secondaire. S'il s'agit d'une variable de première importance, qui ne figure pas dans votre fichier de données, vous devrez, soit éliminer le tableau complètement, soit demander à un analyste de revoir la conception du tableau. Par contre, si c'est une variable d'importance secondaire qui est manquante, éliminez toutes les références à cette variable dans votre programme de tabulation et faites tous les ajustements qu'implique son retrait du programme.

Toute opération de recodification devrait être menée avec beaucoup de soin. Si votre questionnaire comporte des variables ayant des catégories non-standards, toute opération de recodification impliquant ces variables devrait être examinée. Si votre questionnaire comporte des variables non-standards, elles devraient faire l'objet d'une recodification s'il est prévu de les utiliser dans une tabulation.

Il vous faudra aussi contrôler minutieusement toute opération d'adjonction, si votre questionnaire utilise des identifiants qui ne figurent pas dans le questionnaire standard. Plusieurs adjonctions figurant dans les programmes de tabulation ne marcheront que si des identifiants uniques sont utilisés.

### ADJONCTION DES POIDS DE L'ECHANTILLON

La pondération dans les programmes de tabulation est simple, sauf dans les cas où la commande SPSS *aggregate* est concernée. Si le but de la commande *aggregate* est de faire un cumul des cas pour calculer un numérateur et un dénominateur, les poids devraient être ajoutés avant l'exécution de la commande *aggregate*. Ils ne devraient pas être utilisés lorsque l'on travaillera avec le fichier qui en résultera ; ce fichier étant déjà pondéré.

Par exemple, le Tableau HH.1, montré dans l'Annexe Sept, contient le taux de réponses des ménages. Le taux de réponse des ménages est difficile à calculer, car il nécessite de diviser une variable par une autre à l'intérieur du tableau. Une des solutions à ce problème consiste à créer un fichier agrégé qui contient le décompte des ménages échantillonnés, des ménages occupés et des ménages interviewés. Le fichier agrégé contient un cas pour chaque catégorie de la variable discriminante concernée (par exemple, urbain/rural).

Les poids devraient être adjoints lorsque le fichier agrégé est créé pour générer le numérateur pondéré (le décompte des ménages interviewés) et le dénominateur pondéré (le décompte des

ménages occupés). Une fois que le fichier agrégé aura été créé, le taux de réponses des ménages pour chaque catégorie de la variable discriminante est le numérateur divisé par le dénominateur.

Si le but de la commande *aggregate* est de créer un résumé des statistiques des cas individuels, les poids devraient être adjoints après la commande *aggregate*. Par exemple, le Tableau HH.3 de l'Annexe Sept contient des informations sur le pourcentage de ménages abritant au moins un enfant de moins de 15 ans. Cette information ne figure pas dans le fichier de données ménage, mais il peut être créé en agrégeant le fichier d'énumération du ménage. Les variables discriminantes sont le numéro de grappe et le numéro de ménage. Les poids sont adjoints après agrégation, parce que ce qui nous intéresse, c'est le pourcentage pondéré de ménages avec au moins un enfant de moins de 15 ans, et non le nombre pondéré d'enfants de moins de 15 ans dans chaque ménage.

### LA COMMANDE INCLUDE

Le programme SPSS *tables.sps* exécutera tous les programmes de tabulation en une seule fois. Il consiste en une suite de la commande SPSS *include*, qui exécutent les programmes de tabulation de manière individuelle. Si SPSS rencontre une erreur dans l'un des programmes qui sont inclus (autrement dit, exécutés par une commande *include*), il arrêtera immédiatement l'exécution du programme et retournera au programme qui inclut le programme (c'est-à-dire le programme qui contient la commande *include*).

Pour cette raison, *tables.sps* ne devrait être utilisé qu'une fois que tous les fichiers individuels de tabulation auront été contrôlés, modifiés et testés. Assurez-vous d'éliminer la commande *include* qui exécute un programme de tabulation que vous n'utilisez pas.

La commande *include* impose quatre contraintes aux programmes qu'elle exécute. La première contrainte est que chaque commande doit commencer dans la première colonne du programme. Il ressort de cette contrainte qu'elle limite le retrait du programme, alors qu'une ligne devrait compter un retrait si elle commence par le caractère "+". La commande qui suit illustre l'utilisation du caractère "+".

```
do if (cage >= 6 and cage <= 9).  
+ compute solids = 0.  
+ if (BF3G = 1) solids = 100.  
end if.  
variable labels solids "Solid foods".
```

La seconde contrainte imposée par la commande *include* est que si cette commande se poursuit sur plusieurs lignes, la colonne 1 des lignes de continuation devrait rester vide. L'exemple ci-dessous illustre une commande multi-ligne qui respecte cette contrainte.

```
add files  
/file=*  
/file='tmp6.sav'.
```

Notez que les sous-commandes sur les 2<sup>ème</sup> et 3<sup>ème</sup> lignes ont un retrait de deux colonnes (alors qu'elles n'avaient besoin d'avoir un retrait d'une colonne pour respecter la contrainte, elles ont eu un retrait de deux colonnes pour être en conformité avec le style de programmation de la MICS3).

Les troisième et quatrième contraintes imposées par la commande *include* sont que les commandes de fin de programmation sont optionnelles et qu'un astérisque (\*) dans la première colonne d'une ligne indique que celle-ci est une ligne de commentaire. Aucune de ces contraintes n'a d'incidence sur nos programmes de tabulation.

En plus de *tables.sps*, il existe un programme SPSS qui rend automatique la création de fichiers d'analyse. Ce programme s'appelle *CSPro.sps*. Ce programme ne devrait être utilisé qu'après que chacun des programmes qui le composent aura été exécuté et les résultats affichés. Cela garantit que tous les programmes de création de fichier d'analyse seront exécutés dans le bon ordre.

## ARCHIVAGE ET DISTRIBUTION DES DONNEES

Une composante importante, mais souvent négligée du traitement et de l'analyse de l'enquête est l'archivage et la documentation des fichiers de données. En plus du fait que les fichiers des données sont largement disponibles ou bien limités à une seule institution, il est très utile d'établir bien à l'avance quelques directives de distribution. Ces étapes, archivage - documentation et distribution du fichier des données, exigent un investissement en temps et en efforts. Toutefois, cet investissement en vaut la peine, et ce, pour plusieurs raisons :

- **Cela permet d'augmenter le coût-efficacité de la collecte des données.** Recueillir les données d'une enquête est une activité coûteuse et demande un travail intense. Afin de justifier cet investissement, les données recueillies doivent être exploitées de la manière la plus exhaustive possible. Rendre les fichiers des données accessibles à d'autres chercheurs augmente le coût-efficacité de l'activité de l'enquête.
- **Il permet d'augmenter, au niveau national, l'appropriation des données et l'acceptation des résultats de l'enquête.** Quand le fichier des données peut être utilisé par d'autres, le processus de collecte des données gagne en crédibilité. Les responsables de la collecte des données sont perçus comme ayant confiance en leurs résultats, et l'accessibilité au fichier des données par les autres chercheurs signifie que les mêmes résultats peuvent être reproduits et vérifiés par d'autres personnes.
- **Il permet d'examiner les tendances.** Les résultats publiés à partir de différentes enquêtes ne sont pas, souvent, directement comparables. Par exemple, le rapport d'une enquête antérieure peut définir les enquêtés adultes comme ceux âgés de 15 ans ou plus, alors qu'une autre définirait les adultes comme ceux âgés de 18 ans ou plus. Sans les fichiers des données, le mieux que l'on puisse faire c'est d'avoir une comparaison imprécise des deux séries de résul-

tats. Par contre, lorsque les fichiers de données des deux enquêtes sont disponibles, les résultats peuvent souvent être re-tabulés, de telle sorte qu'ils deviennent directement comparables, ce qui permet de tirer des conclusions valables sur les tendances.

- **Il permet de faire des comparaisons intra-pays et inter-pays.** Il est souvent instructif de comparer les résultats entre pays, ou bien au sein d'une sous-région ou entre régions. Ces comparaisons facilitent l'identification des zones où on a besoin d'attirer l'attention sur un programme particulier, ou dans lesquelles des programmes ont été particulièrement réussis. En plus, il peut être utile de comparer les résultats provenant de différentes enquêtes dans un même pays. Quelquefois, cela est fait pour valider des résultats inattendus (par exemple, un niveau de mortalité infantile plus faible que prévu) ou pour évaluer les effets d'une méthodologie particulière de collecte des données (par exemple, le fait de se baser sur les cartes de vaccination par rapport au fait de recueillir les déclarations des mères sur les vaccinations). Afin d'entreprendre ce type d'analyses, les chercheurs doivent pouvoir accéder aux fichiers des données, de telle sorte que des chiffres directement comparables puissent être calculés.
- **Il permet l'analyse approfondie des domaines d'études importants, par des spécialistes en la matière.** En raison de la pression sur les responsables de l'enquête de présenter rapidement les résultats, les informations souvent fournies dans un rapport d'enquête comprennent seulement les résultats de base de cette enquête. Un fichier des données bien documenté et disponible permettra d'entreprendre des analyses approfondies sur des domaines d'études particuliers ; et ces analyses peuvent être exécutées par des spécialistes en la matière qui peuvent ne pas faire partie du personnel de l'institution responsable de la collecte des données.

Les fichiers d'analyses des données créés à partir de l'enquête MICS3 doivent être archivés (sauvegardés dans un format accessible), documentés et distribués. Au minimum, la documentation qui est jointe aux fichiers d'analyse des données doit inclure une copie du rapport complet, une copie des questionnaires et une description du plan de sondage. La documentation doit également comprendre un manuel de codification contenant l'emplacement et la description de chaque variable dans le fichier d'analyse (qui peut facilement être créé dans SPSS). Vous devez archiver une copie de tous les programmes et fichiers utilisés au cours du traitement et de l'analyse des données de l'enquête, et les rendre disponibles à la demande. Une copie des fichiers d'analyse et de leur documentation doit être envoyée au Bureau régional et au siège central de l'UNICEF à New York (Section des Statistiques et du Suivi – *Statistics and Monitoring Section*). Par ailleurs, une politique et une procédure pour la distribution du fichier des données aux autres parties prenantes doivent être établies.

**Tableau 7.7**  
**Liste de contrôle**

Avant la descente sur le terrain :

- Obtenir les ordinateurs et les autres équipements de traitement et d'analyse des données.
- Mettre en place une salle ou un espace pour le traitement des données.
- Recruter un superviseur du traitement et de l'analyse des données et le personnel additionnel.
- Mettre en place un système pour l'organisation des activités de traitement des données.
- Adapter les programmes pour qu'ils soient en cohérence avec le questionnaire du pré-test.
- Saisir et apurer les questionnaires du pré-test.
- Finaliser les programmes en se basant sur l'expérience du pré-test et les adapter selon le questionnaire final.

Pendant le travail de terrain :

- Réceptionner les questionnaires venant du terrain.
- Procéder à l'affectation de la saisie des données primaires.
- Vérifier la structure du fichier des données primaires saisies.
- Procéder à l'affectation de la saisie des données de vérification.
- Vérifier que les données primaires et les données de vérification sont identiques.
- Sauvegarder le fichier de données brutes.
- Procéder à l'apurement secondaire.
- Sauvegarder le fichier de données apurées (ou fichier final).

Après la tournée de collecte sur le terrain :

- Exporter les données vers SPSS.
- Recoder les variables.
- Calculer et ajouter les poids de l'échantillon, l'indice de richesse et les données GPS.
- Lancer les programmes de tabulation.
- Archiver les données et développer un système et des procédures de distribution des données (par exemple, un site Web).
- Envoyer les fichiers d'analyse, la documentation y afférant et tous les programmes à l'UNICEF.

**Tableau 7.8**  
**Exemple de Formulaire de Suivi du Traitement de la Grappe**

Numéro de grappe	Date reçue	Ménages sélectionnés	Nombre de Questionnaires			Saisie des données			Date de simple contrôle final	Date de vérification finale	Date de sauvegarde des données brutes	Date d'apurement	Date de sauvegarde finale
			Ménages	Femmes	Enfants	Nom de l'agent de saisie	Numéro de l'agent de saisie	Date de création					
						P:							
						V:							
						P:							
						V:							
						P:							
						V:							
						P:							
						V:							
						P:							
						V:							
						P:							
						V:							
						P:							
						V:							
						P:							
						V:							
						P:							
						V:							