

ANNEXE 3

Organisation et gestion du traitement des données

TABLE DES MATIERES

Gestion de la base de données	A3.4
Suivi des enregistrements	A3.4
Comment installer et utiliser les programmes de saisie et d'analyse des données	A3.6
Le logiciel EPI INFO pour les enquêtes sur les objectifs de la mi-décennie	A3.6
Installation des programmes	A3.6
Exécution des programmes	A3.6
Saisie des données	A3.7
Notes sur les modules spécifiques	A3.8
Contrôles de cohérence	A3.8
Fréquences des variables	A3.9
Analyse des données	A3.9
Sauvegarde des données	A3.9
Entrer dans EPI INFO	A3.9
Quitter pour le sous-répertoire du DOS	A3.9
Exécution de CSAMPLE pour obtenir les erreurs type pour les enquêtes en grappes	A3.9
Notes au programmeur	A3.11
Suppression de Modules	A3.11
Ajout de Modules	A3.11
Changement de la formulation des questions ou des catégories de réponse	A3.12
Changement dans l'analyse/contrôle de cohérence	A3.12
Fusion des fichiers de données de plusieurs ordinateurs	A3.13

GESTION DE LA BASE DE DONNEES

On vous suggère de mener la saisie et l'édition des données en petits lots, en terminant généralement la saisie pour une grappe entière en une seule fois. Les données seront enregistrées sur le disque dur. Après avoir saisi un lot, les données sont vérifiées au moyen d'un programme de contrôle de cohérence et les résultats sont imprimés.

Toute erreur due à une saisie incorrecte doit être immédiatement corrigée. La liste des erreurs doit être donnée au superviseur pour décider de l'action à engager (par exemple, changer les données, consulter les enquêteurs ou laisser les erreurs non corrigées). Une fois que les corrections des erreurs ont été faites, le programme de cohérence doit être exécuté à nouveau pour vérifier les corrections. Quand il n'y a plus d'erreurs qui peuvent être corrigées au bureau, les données doivent être sauvegardées sur disquette.

☛ Etiquetez vos disquettes avec le nom des fichiers, la région et la date des derniers changements.

Quand la base de données complète est prête pour l'analyse, le programme de contrôle doit encore être utilisé pour une vérification finale avant que les fréquences ne soient produites pour chaque question.

SUIVI DES ENREGISTREMENTS

Les opérations de traitement des données deviennent souvent embrouillées à cause du volume de données et des multiples procédures qui doivent être exécutées. Un bon suivi des enregistrements peut limiter ces problèmes.

La Figure A3.1 donne un exemple de formulaire de suivi des enregistrements. Ce formulaire comporte une ligne pour chaque lot de données.

1. Enregistrez le nombre de questionnaires qui reviennent du terrain au bureau pour chaque grappe. Vérifiez ces nombres par rapport aux totaux du travail sur le terrain.
2. Quand un lot de questionnaires est déposé pour la saisie, enregistrez la date.
3. Quand un lot a été saisi, enregistrez la date, le nombre de questionnaires et le nombre de questionnaires avec des erreurs.
4. Notez la date à laquelle les corrections ont été faites.
5. Chaque fois que des rapports d'erreurs sont produits pour un lot, notez le nombre de documents avec des erreurs et la date.
6. Quand la base de données principale est mise à jour avec le lot, notez la date et le nombre d'enregistrements ajoutés.

Les données du questionnaire sont saisies pour chaque ménage dans l'ordre dans lequel elles ont été collectées: les données du cadre d'information sur le ménage, les données sur tous les enfants de la Feuille d'enregistrement des enfants, le module sur l'eau, les sanitaires et l'iodation du sel, les données sur l'éducation pour chaque enfant d'âge supérieur à l'âge d'entrée à l'école, le module sur l'anatoxine tétanique pour chaque mère, le module sur le traitement des IRA pour chaque mère et, en dernier, les modules sur la santé de l'enfant pour chaque enfant. Une

fois que les données ont été saisies pour un ménage, l'agent de saisie passe au ménage suivant. La saisie peut être faite par plusieurs opérateurs de saisie. Les rapports d'erreurs sont produits à la fin de la saisie de chaque grappe avec une description de toute rubrique incohérente.

Les corrections, notées sur les rapports d'erreurs, sont utilisées pour corriger les enregistrements appropriés sur disquette. Quand on ne rencontre plus d'erreurs qui peuvent être corrigées, les données doivent être sauvegardées sur disquettes.

COMMENT INSTALLER ET UTILISER LES PROGRAMMES DE SAISIE ET D'ANALYSE

LE LOGICIEL EPI INFO POUR LES ENQUETES SUR LES OBJECTIFS DE LA MI-DECENNIE :

INSTALLATION DES PROGRAMMES

Tous les fichiers pour la saisie des données et l'analyse doivent être placés dans un répertoire à part sur le disque dur. Pour installer le logiciel, créez d'abord un nouveau répertoire sur le disque dur, en tapant à l'invite du DOS:

```
MD C:\MIDDECAD
```

Copiez ensuite tous les fichiers de la disquette dans ce répertoire, en tapant :

```
COPY A:*. * C:\MIDDECAD
```

On suppose que EPI INFO v.6 est déjà installé sur le disque dur. Si ce n'est pas le cas, suivez les procédures d'installation décrites dans le manuel de EPI INFO. Le répertoire \EPI6 doit être inclus dans le CHEMIN D'ACCES de l'ordinateur. Vous pouvez vérifier cela en tapant PATH à l'invite du DOS. Si \EPI6 n'est pas inclus dans le CHEMIN D'ACCES, modifiez votre fichier AUTOEXEC.BAT pour l'inclure.

EXECUTION DES PROGRAMMES

On doit toujours utiliser le logiciel à partir du répertoire \MIDDECAD. Commencez par passer à ce répertoire en tapant :

```
C:  
CD \MIDDECAD
```

Puis, pour exécuter le programme, taper :

```
MENU
```

Cette commande fera apparaître un menu comportant comme choix :

1. Saisie des données
2. Contrôle de cohérence
3. Fréquence des variables
4. Analyse des données
- B. Sauvegarde des données
- Q. Quitter pour le sous-répertoire DOS

Saisie des données

Pour saisir de nouvelles données ou pour éditer des données déjà existantes, choisissez l'option 1 du menu. Le choix de cette option fera apparaître le premier écran de saisie des données pour saisir les données du cadre d'information sur le ménage. Un nouveau questionnaire peut immédiatement être entré. Pour éditer un questionnaire existant, pressez <CTRL F> et entrez les numéros de GRAPPE et de MENAGE du questionnaire recherché. Puis, pressez <F3> et <ENTER> pour faire apparaître ce questionnaire pour édition. En pressant <CTRL N> cela permettra d'introduire de nouveaux questionnaires. Pour plus de détails sur la façon de se déplacer dans les écrans de saisie de EPI INFO, voir le manuel de EPI INFO sur la saisie des données.

La saisie des données est organisée en plusieurs modules, y compris :

1. Informations sur le ménage (occurrence simple)
2. Feuille de la mère et de l'enfant (occurrences multiples)
3. Eau et sanitaires (occurrence simple)
4. Iodation du sel (occurrence simple)
5. Education (occurrences multiples)
6. Anatoxine tétanique (occurrences multiples)
7. Traitement des infections respiratoires aiguës (occurrences multiples)
8. Santé de l'enfant (occurrences multiples)

Cette organisation permet au personnel de saisie de compléter l'entrée des données de toute une page du questionnaire avant de passer à la page suivante et, par conséquent, d'éviter les retours en arrière dans le questionnaire.

L'entrée dans chaque module est contrôlée au bas de l'écran ménage. Pour commencer à entrer les données de la feuille d'enregistrement de la mère et de l'enfant, pressez simplement <ENTER> à côté de Passer-A-l'Enfant. Ceci fera apparaître un écran pour saisir la première ligne de la feuille d'enregistrement. Quand les données sont saisies et enregistrées sur disque, le même écran est répété puisque plusieurs lignes sont permises. Après avoir saisi toutes les lignes de la feuille d'enregistrement, pressez <F10> (RETURN) et l'écran ménage réapparaîtra.. Pressez <ENTER> à côté de Passer-A-l'Eau pour passer au Module sur l'Eau et les Sanitaires. Puisqu'une seule occurrence du Module sur l'Eau et les Sanitaires est permise pour chaque ménage, l'écran ménage réapparaîtra immédiatement dès que les enregistrements sur l'eau seront sauvegardés. Les mêmes procédures s'appliquent pour les autres modules. S'il n'y a pas de données pour un module, le module peut être sauté en déplaçant le curseur au-delà du module avant de presser <ENTER> ou en pressant simplement <F10> aussitôt après être entré dans le module. Quand les données ont été saisies pour tous les modules, la saisie sera terminée en pressant <ENTER> à côté de Fin.

Les numéros de grappe et de ménage sont mis ensemble pour former un HOUSEID qui sert de clé pour relier tous les fichiers ensembles. Le HOUSEID doit uniquement identifier le ménage et ne peut pas être répété. De plus, le numéro de ligne de la mère ou de l'enfant sert de clé pour relier l'information sur l'individu à la ligne

correspondante de la feuille d'enregistrement de la mère et de l'enfant. Il est important que ce numéro de ligne soit saisi correctement dans la mesure où la plupart des indicateurs dépendent de la mise en relation des données de la feuille d'enregistrement et des modules spécifiques.

• *Notes sur les modules spécifiques:*

Cadre d'information sur le ménage : La date de l'interview doit être saisie parce que tous les âges et dates de naissance doivent être vérifiés par rapport à la date de l'interview.

Feuille d'enregistrement de la mère et de l'enfant : Le numéro de ligne des mères et des personnes en charge des enfants doivent se terminer par des 0 (c'est-à-dire, 1-0, 2-0, etc.). Leur date de naissance n'est pas saisie. Le numéro de ligne des enfants doit se terminer par un nombre allant de 1 à 9 (c'est-à-dire, 1-1, 1-2, etc.)

Anatoxine tétanique : Le numéro de ligne de la mère (se terminant par 0) doit être saisi ici pour permettre de relier la mère à son plus jeune enfant. (La clé pour relier les enfants est basée sur le numéro de ligne plus 1, puisque les dates de vaccination doivent être comparées à la date de naissance du plus jeune enfant.) Il n'est pas nécessaire de saisir les données de la question 7 puisque la réponse à la question est simplement le total des questions 3 et 5.

Traitement des IRA: Le numéro de ligne de la mère (se terminant par un 0) doit être saisi, bien que la clé pour la liaison est basée sur le numéro de ligne plus 1.

Enfants de moins de cinq ans : Comme tous les modules pour les enfants de moins de cinq ans (diarrhée, Vitamine A, allaitement, vaccination et anthropométrie) doivent être remplis pour un enfant avant de passer à l'enfant suivant, ils sont traités comme un module unique pour la saisie des données. Les données sur la date de l'interview et sur la date de naissance de l'enfant, doivent être saisies de nouveau dans ce module, parce que EPI INFO calcule les indices anthropométriques au moment de la saisie des données.

Contrôles de cohérence

La cohérence interne des données de chaque grappe doit être vérifiée dès que toutes les données de la grappe ont été saisies. Choisissez l'option 2 pour effectuer le contrôle de cohérence. On vous demandera si vous voulez envoyer directement les résultats à l'imprimante ou dans un fichier pour examen ultérieur. On vous demandera aussi quel intervalle de grappes doivent être vérifiées. Si vous êtes seulement intéressés par la grappe 101, tapez 101 au début et à la fin des numéros de grappes. Si vous voulez effectuer des contrôles pour le fichier entier, tapez 000 à la grappe de départ et 999 pour la dernière grappe.

Les erreurs de cohérence seront inscrites dans un fichier appelé CHECK.TXT ou directement imprimées. Chaque déclaration d'erreur fournit : (1) le numéro d'identification du ménage, (2) le module du questionnaire qui contient l'erreur, (3) une description de l'erreur et (4) le numéro de l'individu, s'il s'agit d'un individu.

Fréquences des variables

Quand toutes les données ont été saisies, il est utile d'examiner le listing des fréquences de toutes les variables dans le fichier de données. Le choix de l'option 3 produira un listing des fréquences de toutes les variables à l'exception de HOUSEID et CHILDDID.

Analyse des données

Choisissez l'option 4 pour analyser les données. Cette option exécute l'ensemble des commandes contenues dans INDICATR.PGM, qui produit les résultats pour les indicateurs selon les spécifications exposées à l'Annexe 4. Vous pouvez obtenir un listing complet du fichier .PGM en imprimant simplement le fichier, soit à l'écran, soit sur l'imprimante. Les résultats sont limités aux indicateurs pour l'ensemble de l'échantillon. Le programme peut être modifié pour examiner seulement certains sous-groupes particuliers de la population, puis il peut être exécuté à nouveau pour ce sous-groupe. Les instructions pour modifier le programme sont données à la section "Notes au programmeur" qui figure ci-dessous.

Le programme d'analyse ne fournit pas les erreurs type des indicateurs. Ces erreurs peuvent être calculées en utilisant l'option CSAMPLE de EPI INFO v.6. Voir ci-dessous les instructions sur la façon d'utiliser CSAMPLE.

Sauvegarde des données

Les données doivent être sauvegardées quotidiennement sur disquettes, en utilisant différentes disquettes les différents jours de la semaine. Le choix de l'option B vous permettra de copier les fichiers appropriés sur les disquettes. Cette procédure suppose que la taille des fichiers de données ne dépassera pas la capacité d'une simple disquette. Dans le cas où les fichiers de données sont trop importants pour une disquette, le responsable des données devra trouver une autre méthode pour recopier les données et modifier le MENU.BAT en conséquence. On peut utiliser, par exemple, PKZIP, BACKUP (DOS), ou PCTOOLS.

Entrer dans EPI INFO

S'il est nécessaire de travailler directement avec le logiciel EPI INFO, choisissez l'option E qui vous mènera à l'écran de démarrage de EPI INFO v.6. Tous les programmes et options de EPI INFO peuvent être exécutés à partir de cet écran. Lorsque vous sortirez de EPI INFO, vous retournerez au menu. (L'option E n'apparaît pas vraiment sur la liste du menu dans la mesure où le personnel de saisie n'a pas besoin d'entrer directement dans EPI INFO de façon régulière. Cependant, cette option fonctionne.)

Quitter pour le sous-répertoire DOS

Choisissez l'option Q pour sortir de l'application.

EXECUTION DE CSAMPLE POUR OBTENIR LES ERREURS TYPE POUR LES ENQUETES EN GRAPPES

Pour estimer correctement les erreurs type pour les indicateurs avec EPI INFO, on doit créer sur le disque un jeu de données qui contient les variables créées pour l'indicateur auquel on s'intéresse. Par exemple, supposons que nous voulions obtenir les erreurs type pour l'indicateur 11.1, le pourcentage d'enfants de moins de 5 ans ayant un faible poids-pour-âge. Pour créer un tel jeu de données, ajoutez deux lignes au fichier INDICATR.PGM après la ligne "report waz.rpt":

```
ROUTE TEMP.REC
WRITE "RECFILE" CLUSTER LOW2WAZ
```

et exécutez l'analyse. Cela créera un fichier temporaire appelé TEMP.REC, qui contient seulement la variable grappe et la variable déterminant si l'enfant a, ou non, un faible poids-pour-âge.

Choisissez ensuite l'option CSAMPLE à partir du menu PROGRAMS en EPI INFO v.6. Tapez TEMP comme nom du fichier. Un écran d'options apparaîtra. En-dessous de Main, tapez LOW2WAZ, la variable pour laquelle vous avez besoin d'erreurs type. Précisez la variable strate seulement si l'échantillon a été stratifié. Précisez la variable UPS (au Chapitre 4, le terme "petite zone" se rapporte à ces Unités Primaires de Sondage), à laquelle on donne généralement le nom CLUSTER (grappe). On n'a pas besoin de pondérer à moins que le plan de sondage n'ait introduit des pondérations. Spécifiez une variable de tabulation croisée seulement si vous avez besoin d'examiner les erreurs type dans des sous-groupes de population. Les résultats peuvent être lus directement sur l'écran, imprimés directement, ou mis en mémoire dans un fichier en vue d'une utilisation ultérieure. Si les données ne sont pas déjà triées par GRAPPE, choisissez alors SORT. Choisissez TABLES, et l'analyse sera effectuée.

Figure A3.2 Exemple de résultat de CSAMPLE pour l'analyse du poids-pour-âge

CTABLES COMPLEX SAMPLE DESIGN ANALYSIS

Analysis of LOW2WAZ
LOW2WAZ

	Total
0	
Obs	2168
Percent V	79.009
SE%	0.887
LCL%	77.271
UCL%	80.746
1	
Obs	576
Percent V	20.991
SE%	0.887
LCL%	19.254
UCL%	22.729
Total Obs	2744
Design eff.	1.300

LCL = lower confidence limit
UCL = upper confidence limit

Sample Design Included:

Sampling Weights--None
Primary Sampling Units from CLUSTER
Stratification--None

0 records with missing values

Les résultats donnent l'erreur type pour l'indicateur et les valeurs supérieure et inférieure de l'intervalle de confiance à 95 pour cent. Le programme calcule aussi l'effet du plan de sondage obtenu dans l'enquête pour la variable à laquelle on s'intéresse. Par exemple, dans l'analyse du poids-pour-âge, le résultat ressemblera à l'exemple donné à la Figure A3.2, dans laquelle le pourcentage d'enfants avec un faible poids-pour-âge (arrondi à l'unité) est de 21,0 avec un intervalle de confiance de (19,3; 22,7).

NOTES AU PROGRAMMEUR

Certaines modifications des programmes seront nécessaires dans chaque pays où on utilise le logiciel. Par exemple, la définition de "l'eau salubre à distance acceptable" est spécifique à chaque pays et, par conséquent, le programme d'analyse INDICATR.PGM doit être révisé pour prendre en compte la définition utilisée dans le pays. La définition de l'âge d'entrée à l'école varie d'un pays à l'autre. Il existe aussi des modules différents pour la Vitamine A et pour la couverture vaccinale. Un programmeur qui connaît EPI INFO doit adapter le programme aux spécificités du pays. Il est préférable que le programmeur comprenne parfaitement la structure de la base de données avant d'entreprendre quelque modification que ce soit. Cela est particulièrement vrai s'il faut effectuer de nombreux changements dans le questionnaire.

Tout changement aux fichiers .QES ou .CHK exige que le fichier .REC correspondant soit recréé. Cela peut être fait en choisissant ENTER en EPI INFO v.6 et en sélectionnant l'option 2, "Create new data file from .QES file" (créer un nouveau fichier de données à partir du fichier .QES). Comme le fichier .REC contient toutes les données, il est important que tous les changements soient faits aux fichiers .QES et .CHK avant de commencer la saisie des données. Il est nécessaire de faire quelques modifications de ces fichiers après que les données aient été saisies, consultez le manuel de EPI INFO à propos de "Révision de la structure du fichier de données (Choix 3 du menu)", du chapitre 8 "Saisie des données".

Suppression de modules

N'importe quel module du questionnaire modèle qu'un pays décide de ne pas utiliser peut facilement être sauté. Si le module à supprimer est l'un de ceux listés ci-dessus à la page A3.7, une simple suppression dans le fichier HOUSEHLD.CHK évitera d'entrer dans le module. La variable correspondant à ce module dans le HOUSEHLD .QES sera également supprimée .

Si seulement une partie du module de saisie des données doit être supprimée (par exemple, les questions sur l'allaitement dans le module sur la santé des enfants), les lignes se rapportant à ces questions devront alors être supprimées des fichiers .QES et .CHK. Les modules alternatifs sur la Vitamine A et sur la vaccination qui ne sont pas utilisés doivent être supprimés de cette façon. Un nombre correspondant de lignes vierges doit être ajouté ou supprimé du fichier .QES pour permettre qu'à l'écran, chaque page se termine à un endroit logique.

Ajout de modules

Si des modules supplémentaires doivent être ajoutés au questionnaire, il sera nécessaire de développer de nouveaux fichiers .QES et .CHK. Les fichiers existants peuvent servir de modèle sur la façon de procéder. Dans la plupart des cas, des questions supplémentaires pourront être ajoutées aux modules existants : dans ces cas, il suffira de modifier les fichiers .QES et .CHK.

Changements de la formulation des questions ou des catégories de réponse

Les changements de la formulation des questions (y compris la traduction en langues différentes de l'anglais) nécessitent simplement que les écrans définis dans le fichier .QES soient modifiés pour faire apparaître la formulation appropriée. Les noms d'origine des variables qui sont utilisés dans les fichiers .QES, .CHK et .PGM n'ont pas besoin d'être changés. Cependant, si le nom des variables change, les changements doivent être faits dans tous ces fichiers.

Si les catégories de réponses admissibles pour les questions sont modifiées, les changements correspondants doivent être faits au fichier .CHK (pour permettre toutes les réponses valides, mais pour ne pas autoriser les réponses invalides) et au fichier .PGM (pour garantir que les indicateurs seront calculés en utilisant les catégories correctes). Le logiciel n'attribue pas explicitement d'étiquettes aux réponses, de sorte qu'il n'est généralement pas nécessaire de composer des étiquettes pour les réponses.

Changement dans l'analyse/contrôle de cohérence

Le contrôle de cohérence ainsi que le calcul des indicateurs sont faits à travers les fichiers .PGM files. Ces fichiers peuvent être édités pour supprimer les références à n'importe quel module non utilisé et pour préciser la définition de certains indicateurs (par exemple, l'eau salubre et à distance acceptable, l'âge d'entrée à l'école et la fréquence de l'utilisation des capsules de Vitamine A). On souhaite généralement obtenir les indicateurs pour des sous-groupes particuliers comme les régions ou le sexe.

- ☛ Comment obtenir des estimations pour des sous-groupes : pour calculer chaque indicateur pour des sous-ensembles de données, la modification la plus simple consiste à choisir seulement les observations du sous-ensemble et à relancer le programme d'analyse. Par exemple, pour calculer les indicateurs seulement pour les garçons, sélectionnez les observations concernant les garçons :

```
SELECT SEX = 1.
```

Comme l'énoncé sélectionné doit être exécuté chaque fois qu'un nouvel ensemble de données est lu, INDICTR.PGM comprend, à la fin, un sous-programme appelé "SELECTSEX" pour exécuter l'analyse pour n'importe quel sous-ensemble de données qui est choisi. Editez ce sous-programme et exécutez le programme révisé pour calculer les indicateurs pour les garçons seulement. Pour les filles seulement, modifiez le sous-programme avec la ligne :

```
SELECT SEX = 2
```

et lancez à nouveau le fichier INDICATR.PGM. L'énoncé sélectionné peut aussi être modifié pour prendre en compte seulement les observations d'une REGION particulière, d'un groupe d'AGES ou d'une catégorie URBAIN/RURAL. Bien entendu, le programmeur devra s'assurer que la variable à laquelle on s'intéresse est toujours disponible en émettant une commande RELATE appropriée avant la sélection.

- ☛ Des fichiers .PGM supplémentaires peuvent être créés pour d'autres raisons—par exemple, pour calculer la durée moyenne de l'allaitement ou pour examiner la distribution par âge des enfants. L'étude des fichiers .PGM existants peut être très utile pour démarrer de nouvelles analyses.

☛ Tous les programmes d'analyse de EPI INFO sont rédigés sous l'hypothèse que l'échantillon du pays est conçu pour être auto-pondéré (c'est-à-dire que tous les poids sont égaux à 1,0). Cependant, un sous-programme est inclus pour procéder à des analyses pondérées aussi simples que possibles. Le programmeur aura simplement besoin d'éditer le sous-programme ":WEIGHT" à la fin du fichier INDICATR.PGM pour inclure l'attribution de poids comme cela est décrit au chapitre 7.

Fusion des fichiers de données de plusieurs ordinateurs

Dans certains cas, il sera nécessaire de saisir les données des questionnaires sur plusieurs ordinateurs, puis de fusionner les données pour l'analyse. Ceci peut se faire au moyen du programme MERGE en EPI INFO. Les fichiers .REC d'un ordinateur doivent être copiés sur un autre, mais placés dans un sous-répertoire à part. Le programme MERGE, en utilisant l'option CONCATENATE, créera un nouveau fichier de données comportant les données des deux fichiers. La procédure peut être répétée si l'on utilise plus de deux ordinateurs. La fusion peut être réalisée séparément pour chaque module.