

2.
CEPED
CENTRE FRANÇAIS SUR LA POPULATION
ET LE DÉVELOPPEMENT
15, rue de l'École-de-Médecine
75270 PARIS CEDEX 06
Tél. : (1) 46 33 89 41

COMITE INTERETATS DE LUTTE CONTRE LA
SECHERESSE DANS LE SAHEL

C. I. L. S. S.

INSTITUT DU SAHEL

(U.S.E.D.)

ETUDE SUR LA MORTALITE DANS LE SAHEL

(E.M.I.S)

NOTE TECHNIQUE No 2.-

Vérification des données et
apurement des fichiers.-

BAMAKO, septembre 1982.-

INSTITUT DU SAHEL - EMIS

NOTE TECHNIQUE No 2

Vérification des données et apurement des fichiers.

La présente note technique décrit la saisie des données de l'enquête sur la mortalité infantile et juvénile à Bobo-Dioulasso en Haute-Volta. L'exploitation mécanographique de l'enquête se fait à l'Institut du Sahel à Bamako sur micro-ordinateur IBM Personal Computer. Les méthodes exposées ici sont conditionnées par une série de facteurs spécifiques à l'enquête. En premier lieu, l'usage d'un micro-ordinateur a rendu possible la saisie directe de l'enquête, sans codage préalable des questionnaires. En second lieu, la saisie a débuté avant la fin des opérations sur le terrain. Comme l'enquête de Bobo est une enquête à passages répétés, cette procédure a permis de vérifier les réponses reçues lors des passages précédents, de reposer les questions omises et de résoudre les incohérences constatées.

Saisie directe et tests de cohérence.

Le questionnaire de l'enquête de Bobo-Dioulasso se conforme très largement au prototype mis au point par l'IFORD. C'est un instrument relativement complexe, tant par sa structure (questions-filtres et renvois) que par le nombre des variables (47 pour le questionnaire de base, et 51 pour chacun des questionnaires de passage). Le codage de ces informations n'avait pas été prévu sur le questionnaire même. Dans la conception initiale, on comptait procéder à un codage manuel préalable à la saisie mécanographique. Le recours au micro-ordinateur permet de sauter ce stade. L'opérateur transcrit les réponses directement en clair sur l'écran, dans une grille qui reproduit le questionnaire. C'est le programme de saisie qui effectue le codage. Les programmes, écrits en BASIC, reconnaissent une série de variables (par exemple le nom de l'ethnie, l'occupation, le quartier de résidence, l'établissement sanitaire...), acceptent les résultats numériques (dates, nombre d'enfants précédents, poids du bébé) et codent les choix binaires (oui/non) ou multiples (catégorie de revenu, état matrimonial, type d'accouchement...). Les programmes CREER et NEONE sont utilisés pour saisir le questionnaire de base, et le programme PASS pour les questionnaires de passage. La discussion qui suit se limite à la saisie du questionnaire de base.

Les programmes de saisie contiennent une série de tests de

cohérence. Un message sur l'écran attire l'attention de l'opérateur et demande confirmation de ce que l'incohérence constatée se situe bien au niveau du questionnaire et n'est pas le résultat d'une erreur de saisie. Dans l'affirmative, l'ordinateur code une non-réponse, ce qui déclenchera un message à vérifier lors du prochain passage sur le terrain. Quelques exemples concrets feront mieux comprendre la procédure.

a) Revenus du ménage.

Dans l'esprit des organisateurs, la question sur les revenus ne devait être posée que lors du premier passage (c'est à dire un mois après que le questionnaire de base ait été rempli en maternité). En effet, la plupart des femmes ne connaissent pas le revenu de leur mari; en outre, la question paraissait de nature à inspirer la méfiance, et on s'était dit qu'il valait mieux attendre pour la poser qu'une atmosphère de confiance se soit établie entre enquêtrice et enquêtée. Mais à Bobo, la réponse n'avait dans bien des cas pas encore été portée sur le questionnaire de base lors du troisième passage. -Lors de la saisie, quand l'opérateur n'a coché aucune des quatre cases correspondant à une catégorie de revenu, l'ordinateur affiche la question: "Revenus inconnus ? Oui/Non ". Si l'opérateur inscrit n (pour Non) le curseur réapparaît à la première case des catégories de revenu, pour permettre à l'opérateur de réparer son omission. Si toutefois il avait inscrit o (pour Oui), l'ordinateur aurait donné à la variable Revenus le code 0 (revenu inconnu), ce qui déclencherait plus tard un "message d'erreur" (voir plus loin.).

b) Nombre d'enfants déjà eus.

Le questionnaire de l'enquête permet plusieurs recoupements au sujet des accouchements précédents. Les questions portent sur la grossesse précédente, et sur le nombre total d'accouchements précédents. L'ordinateur effectue les tests suivants:

- Si la grossesse précédente a donné un mort-né, un avortement, ou un enfant décédé depuis, y a-t-il un mort-né, un avortement ou un décès mentionné parmi les réponses portant sur le nombre total d'accouchements?

- Le nombre total d'accouchements est-il égal au nombre d'enfants nés vivants plus le nombre de morts-nés? (A noter toutefois la possibilité de naissances gemellaires.)

On constate de nombreuses incohérences dues à la confusion fréquente entre les morts-nés et les enfants décédés peu de temps après la naissance.

[Voir par exemple le cas #2318:

4 enfants nés-vivants

0 encore en vie

4 morts-nés

Nombre total d'accouchements: 4

ou le cas #2896:

Grossesse précédente: mort-né

Date de naissance: janvier 79; date de décès: septembre 79

Avez-vous eu des morts-nés? Non

Nombre total d'enfants nés-vivants: 1; encore en vie: 0.]

Dans le cas d'une incohérence de ce type, l'ordinateur avertit l'opérateur pour le mettre en garde contre une erreur de frappe, et après confirmation donne à la variable incriminée le code correspondant à une non-réponse, pour vérification sur le terrain lors du passage suivant.

Vérification.

Directement après la saisie complète des données d'un mois, et avant le retour sur le terrain pour le passage suivant, les questionnaires sont vérifiés par échantillonnage systématique. La taille de l'échantillon et la fréquence des vérifications dépendent du nombre d'erreurs identifiées. Le programme VERBO est utilisé pour la vérification. Ce programme opère une seconde saisie d'une série de questionnaires choisis au hasard, et en compare le produit aux renseignements existant déjà dans le fichier approprié dans la banque de données. Le numéro du questionnaire sélectionné apparaît sur la feuille de l'imprimeuse et, après vérification les valeurs des variables pour lesquelles les deux saisies divergent sont aussi imprimées. La réconciliation sera faite après coup par le contrôleur sur la liste obtenue. Quand la saisie et la vérification concordent, l'ordinateur choisit et affiche le numéro de la naissance suivante à vérifier correspondant à l'intervalle d'échantillonnage. S'il y a désaccord, l'ordinateur sélectionne le numéro de la naissance suivante immédiatement celle qui vient d'être vérifiée. Le but de cette procédure est de faire des vérifications plus fréquentes quand on a identifié des erreurs. La fraction d'échantillonnage, en l'absence

d'erreur, est à présent d'un questionnaire sur huit, ce qui paraît suffisant pour contrôler la qualité de la saisie. Les résultats obtenus à Bobo-Dioulasso sont réconfortants quant à cette qualité, les erreurs identifiées affectant moins d'un questionnaire vérifié sur dix. Le type d'erreur rencontré le plus souvent est d'ailleurs mineur, et porte sur l'interprétation du code des professions, où l'opérateur a une certaine marge de choix. (Comment faut-il classer un captureur d'insectes du Centre Muraz ?)

Le but de la vérification n'est évidemment pas d'identifier toutes les erreurs commises lors de la saisie; il faudrait pour cela procéder à la vérification intégrale. L'objectif poursuivi est celui d'un contrôle de qualité. Il est possible de mesurer la part d'imprécision introduite par la saisie. Par exemple pour septembre, sur les 586 naissances du mois on en a vérifié 110, ou 18%. La vérification a porté sur 44 variables (c'est à dire toutes les variables à l'exclusion des noms de personnes). Pour les 110 questionnaires, on a donc vérifié 110 x 44, soit 4840 valeurs numériques, 38 erreurs ont été identifiées, soit moins d'un pour cent. C'est un niveau qui paraît acceptable.

Messages d'erreur.

Après la saisie et la vérification, on soumet le fichier au programme MESSAGE qui produit des messages d'erreur correspondant à chaque code "inconnu" rencontré. Les messages attirent donc l'attention

sur les non-réponses et les incohérences constatées à la saisie, plutôt que sur des erreurs à proprement parler; mais le terme "message d'erreur" est commode.

Les messages sont imprimés en deux exemplaires, et vérifiés individuellement pour s'assurer qu'ils ne résultaient pas d'une omission lors de la saisie. Le premier exemplaire est découpé en papillons qui sont insérés dans les questionnaires individuels où l'enquêtrice les trouvera lors du prochain passage sur le terrain. L'enquêtrice reposera les questions litigieuses, recueillera tous les compléments d'information sur le papillon et corrigera le questionnaire même. Au bureau de l'enquête, le contrôleur transcrira l'information collectée du papillon au second exemplaire imprimé des messages. C'est ce document qui sera renvoyé au centre de saisie, où les disquettes de saisie seront complétées et corrigées.

A Bobo-Dipulasso, les messages d'erreur sont peu nombreux, ce qui témoigne de l'excellente qualité du travail sur le terrain. Seule la question sur les revenus du ménage, qui n'avait pas été posée lors de la phase initiale de l'enquête, a également été souvent oubliée des passages à domicile. Il y a en moyenne bien moins d'un message d'erreur par questionnaire. Les non-réponses des femmes concernent avant tout les caractéristiques du mari, particulièrement son niveau d'éducation. Les incohérences constatées portent principalement sur le nombre d'accouchements et résultent souvent d'une confusion entre les morts-nés et les décès précoces.

Consultation de sources extérieures à l'enquête.

Dans le milieu urbain qui fait l'objet de ce type d'enquêtes, on dispose souvent de listes de naissances assemblées à des fins d'état civil, ou pour les besoins du système hospitalier. Comme l'univers des enquêtes est identifié lors de l'accouchement en maternité, c'est en tout premier lieu les registres des maternités qui devraient être systématiquement dépouillés. A ce titre, l'expérience de Bobo-Dioulasso est de nouveau instructive. Les registres des deux maternités de la ville, celles de l'Hopital et de Guimbi, et ceux des sages-femmes retraitées qui assistent les naissances à domicile, permettent non seulement de vérifier la couverture de l'enquête, mais aussi certains renseignements essentiels sur l'enfant au moment de la naissance. Tel enfant qui était déclaré mort-né au moment de l'enquête, à la lecture du registre de la maternité s'avère avoir manifesté des signes de vie à la naissance. Tel autre avait en fait été évacué d'une ville voisine pour une intervention chirurgicale, et ne relevait donc pas du champ de l'enquête.

Comme nous l'avons discuté dans une Note Technique précédente, les enquêtes de ce type ont tendance à sous-estimer la mortalité périnatale, parce qu'elles perdent des naissances difficiles où l'accouchée est transférée vers un autre service de l'hopital avant l'interview. Souvent, la mère est trop souffrante pour se prêter à l'enquête. Il se peut même qu'elle évite l'enquêteuse pour cacher sa honte d'avoir perdu l'enfant. C'est seulement en compulsant systématiquement les registres de

la maternité qu'on peut espérer corriger partiellement ces omissions. Partiellement seulement, car le registre ne saurait remplacer l'enquête. Il permet tout au plus de réintroduire les décès manquant dans le calcul des quotients. Les renseignements socio-démographiques du questionnaire sont perdus pour de bon.

Apurement des fichiers.

Lors de l'enquête, à Bobo-Dioulasso, les questionnaires sont classés par catégories pour faciliter le travail sur le terrain. C'est ainsi que les questionnaires des enfants qui sont toujours sous observation constituent le gros des dossiers, que les enquêtrices complètent et tiennent à jour au fur et à mesure des passages. Par contre, les décédés, les morts-nés, les émigrés et les non-retrouvés constituent des piles séparées. Cependant, à une petite exception près (les naissances dites "traditionnelles", c'est-à-dire non assistées par une sage-femme), tous les accouchements survenus à Bobo lors de l'année d'observation, ont fait l'objet d'une numérotation continue unique. A notre sens, la numérotation unique constitue un instrument indispensable à la bonne gestion du fichier. Il faut éviter les numérotations parallèles (naissances traditionnelles ou morts-nés), et les duplications de numéros (par exemple dans le cas des jumeaux) sont sources de confusion et d'erreur. Un stade essentiel de la vérification consiste à identifier la cause des trous dans la numérotation et des doubles numéros. Cela permet de s'assurer que la saisie a été complète.

Après la saisie et la vérification, les données sont classées en ordre par fichiers de cinquante (numéros d'ordre de 1 à 49, 50 à 99, etc.) Cette taille facilite la correction séquentielle des fichiers. Le reclassement des fichiers est effectué au moyen du programme NUMERBO. Le programme MODIF permet d'éditer et d'apurer les fichiers en y intégrant, cas par cas, toutes les corrections identifiées lors des stades décrits précédemment. Le programme INVENT permet d'imprimer le numéro d'ordre, le nom de la mère et la date de la naissance. La séquence des numéros et des noms est comparée avec la série des questionnaires individuels pour s'assurer qu'il n'y a eu ni double compte ni omission.

On observera que l'usage du micro-ordinateur permet la correction et l'apurement des fichiers en mode interactif. Lors de la saisie des données d'une enquête, le stade de la correction est toujours complexe. La correction des données individuelles apparaissant en clair sur l'écran, tout en restant laborieuse, représente un apport considérable aux techniques de collecte et de saisie.