

II - DESCRIPTION DETAILLÉE DU SYSTÈME DE TRAITEMENT DES DONNÉES POUR
L'ENQUÊTE SUR LA FÉCONDITÉ DU RWANDA

L'Enquête Nationale sur la Fécondité a été conduite par l'Office National de la Population (ONAPO) entre la fin d'août et le début de décembre 1983. L'enquête a été faite sur un échantillon nationalement représentatif comprenant 100 unités d'échantillonnage rurales et 30 urbaines (districts). Les données ont été collectées à l'aide de trois questionnaires : un questionnaire sur les ménages donnant la liste de toutes les personnes résidant habituellement ou en visite dans le ménage ; un questionnaire individuel détaillé pour toutes les femmes âgées de 15 à 50 ans ; et un questionnaire sur les maris posé à un sous-échantillon des maris des femmes interviewées. Le questionnaire sur les ménages a été utilisé principalement pour sélectionner les femmes éligibles pour l'interview individuelle, mais aussi pour collecter l'information démographique de base (âge, sexe, statut marital, etc.). Le questionnaire individuel couvre les renseignements socio-économiques des femmes, les historiques complets des grossesses et des mariages, les préférences en matière de fécondité, la connaissance et l'utilisation de la contraception et l'information détaillée sur les deux dernières grossesses. Le questionnaire sur les maris était essentiellement une version abrégée du questionnaire individuel.

Les questionnaires avaient été conçus sur le modèle des questionnaires utilisés par l'Enquête Mondiale sur la Fécondité (EMF) dans les 42 pays en voie de développement entre 1974 et 1982. Le système de traitement des données était basé sur les procédures contenues dans le document de l'EMF "Manuel de Traitement des Données". Ces procédures ont été adaptées aux conditions particulières au pays et modifiées pour pouvoir produire les résultats de base de l'enquête aussi rapidement que possible. La proposition initiale pour le traitement des données a été préparée en mai, après une visite de deux semaines au Rwanda. Le système a été désigné pour marcher sur un mini-ordinateur NCR 8250 situé au bureau du recensement rwandais et a été installé au cours de deux missions de 10 semaines au Rwanda. La première mission, qui s'est déroulée de la mi-août à la fin d'octobre 1983, a été consacrée à l'organisation du bureau de l'enquête et à la conception, la rédaction et l'installation des programmes nécessaires à l'entrée des données. La seconde mission, de la fin de janvier à la première semaine d'avril 1984, a été consacrée à la validation, l'imputation, la construction des variables nécessaires à l'analyse et la production des tableaux exigés pour le rapport de l'enquête.

Le système de traitement des données comprend trois phases principales : la préparation des données ; l'imputation et la recodification ; et la tabulation. La phase de préparation des données inclut deux sous-systèmes séparés mais en relation : un pour la saisie, la mise à jour et la liste des informations de l'échantillon de l'enquête et le second pour la saisie et la validation des données à partir des questionnaires de l'enquête. Le système final comprenait 15 programmes, dont quatre programmes généraux développés par l'EMF et deux programmes utilisés pour l'Enquête sur la Fécondité au Nigeria qui ont été adaptés à l'enquête rwandaise. Les résultats de l'enquête ont été stockés dans cinq fichiers de données informatiques, dont un fichier d'archive avec tous les tableaux et les distributions de fréquence.

La phase de préparation des données comprenait le traitement manuel et informatique à partir du moment où les questionnaires arrivaient du terrain jusqu'à leur saisie sur ordinateur. Tout le travail de préparation des données, la saisie des données incluse, a été fait par le personnel du bureau de l'enquête comprenant 10 employés pour la préparation des données et deux superviseurs. Ces employés avaient suivis la formation d'enquêteur et étaient familiers avec les questionnaires de l'enquête. Les questionnaires étaient reçus du terrain par unité d'échantillonnage (district) et étaient immédiatement comptés pour s'assurer qu'aucun questionnaire ne manquait. Le nombre de questionnaires complets et incomplets individuels, pour les ménages et pour les maris étaient enregistrés par district dans un registre et ces totaux étaient utilisés pour la mise à jour des fichiers de chaque district stockés dans l'ordinateur. Après la procédure de réception, les questionnaires étaient conservés au bureau de l'enquête pour un traitement ultérieur, consistant dans le contrôle, la codification et la saisie des données.

Le questionnaire était largement pré-codé et le programme qui informatisait la saisie des données incluait pratiquement toutes les vérifications normalement faites à la main. En conséquence, le traitement manuel était limité à une révision superficielle des questionnaires pour localiser les erreurs importantes et à la codification d'un nombre limité de questions ouvertes, telles que l'occupation, le nombre d'enfants désirés, etc. Quand le traitement manuel était fini, les questionnaires étaient saisis directement sur l'ordinateur, généralement par la personne qui avait fait le contrôle et la codification. Le fichier du district, mis à jour avec le résumé des résultats du travail sur le terrain, était utilisé pour vérifier l'identification des questionnaires et empêcher des questionnaires d'être saisis deux fois.

Le programme de saisie des données suivait l'ordre du questionnaire et seules les questions ayant été posées à la personne interviewée étaient entrées. Les données étaient entrées question par question et chacune était vérifiée pour s'assurer que la valeur entrée était une valeur comprise dans le domaine des codes. Le programme suivait les sauts dans le questionnaire bien que certaines devaient être confirmées par l'opérateur. Les questions filtres qui avaient servi à contrôler la séquence des questions n'étaient pas entrées mais elles étaient déterminées par l'ordinateur et devaient être confirmées par l'opérateur. Les contrôles de cohérence entre les questions étaient incorporées au programme et les erreurs détectées devaient être corrigées ou le questionnaire était refusé et donné à un des superviseurs pour être vérifié de plus près. Les contrôles de cohérence comprenaient la plupart des vérifications faites normalement pendant le contrôle informatique recommandé par l'EMF. Il s'agit de contrôler toutes les dates dans les historiques des grossesses et des mariages et de vérifier que les grossesses indiquées dans l'historique des grossesses se trouvaient dans la section 5 du questionnaire, car cela avait été un gros problème dans d'autres enquêtes. Certains contrôles de cohérence n'étaient pas inclus, car ils étaient considérés trop complexes ou certaines dispositions avaient été prises dans les derniers programmes pour les détecter et les corriger automatiquement.

Quand le programme détectait des erreurs, les opérateurs vérifiaient d'abord qu'ils avaient bien entré les bonnes informations du questionnaire. Si les données étaient visiblement incohérentes, l'opérateur essayait de résoudre cette incohérence, si nécessaire avec l'aide des superviseurs. Si l'erreur ne pouvait être corrigée facilement, le questionnaire était abandonné pour plus ample vérification par le superviseur. Certaines erreurs étaient le résultat de conditions qui n'avaient pas été prévues par le programme de saisie des données. De tels questionnaires furent mis de côté jusqu'à la deuxième mission au cours de laquelle le programme fut corrigé. Le nombre final de questionnaires abandonnés a été inférieur à 100, ces problèmes étant principalement dus au traitement incorrect des naissances multiples qui se trouvaient à la fin de l'historique des grossesses. Ce mélange de contrôle au bureau, de saisie des données et de recodification n'avait pas été prévu au départ dans le système proposé, mais a été adopté à cause du manque d'expérience du personnel de saisie des données. Au début, nous avions peur que cette procédure ne soit extrêmement lente, mais nos craintes ne furent pas justifiées car après quelques semaines d'expérience, les opérateurs de saisie étaient capables d'entrer les données bien plus vite que prévu. L'entière opération de saisie ne prit finalement qu'un peu plus de trois mois, ce qui est à peu près le temps nécessaire lorsque l'on utilise des méthodes conventionnelles de saisie des données. Le tableau ci-dessous montre un résumé des opérations de saisie des données.

| Mois | Inter-views | Nombre de Ménages | Quests entrés Indivd | % | Temps moyen par entrée (mn) | Ménag Indivd Total | Temps médian par entrée (mn) | Ménag Indivd Total |
|-----------|-------------|-------------------|----------------------|------|-----------------------------|--------------------|------------------------------|--------------------|
| Mois | : | : | : | : | : | : | : | : |
| Mois | : | Nombre de Ménages | Quests entrés Indivd | % | Temps moyen par entrée (mn) | Ménag Indivd Total | Temps médian par entrée (mn) | Ménag Indivd Total |
| Août | : | 715 | | | | | | |
| Septembre | : | 1466 | | | | | | |
| Octobre | : | 1843 | 494 | 463 | 9 | 4.4 | 13.6 | 18.0 |
| Novembre | : | 1830 | 1801 | 1799 | 32 | 3.4 | 7.9 | 11.3 |
| Décembre | : | 33 | 1227 | 1248 | 21 | 4.2 | 6.8 | 11.0 |
| Janvier | : | | 1982 | 1973 | 34 | 3.6 | 6.0 | 9.6 |
| Février | : | | 214 | 256 | 4 | 3.3 | 5.5 | 8.8 |
| TOTAL | : | 5887 | 5718 | 5739 | | 3.7 | 6.8 | 10.5 |

Note : Les temps moyens ne prennent en compte que les questionnaires complets.

La moyenne totale de 7-10 minutes par questionnaire sur les ménages donne un taux de 6-8 ménages par personne et par heure ou 30-40 ménages par personne et par jour.

L'aspect le plus important de cette approche du contrôle fait au bureau et de la saisie des données est la cohérence des données produites. Les contrôles ultérieurs n'ont fait apparaître aucune erreur de valeurs non comprises dans le domaine des codes, de sauts ou de filtres et très peu d'erreurs de cohérence. Les données étaient assez cohérentes pour que l'analyse commence aussitôt que la saisie des données a été terminée et cela a été fait sans que cela affecte sérieusement le temps requis pour la saisie des données.

La phase de validation, d'imputation et de recodification des données a suivi de très près les recommandations de l'EMF, principalement parce que nous avons décidé de produire un fichier final des données dans le même format que les fichiers recodés standard de l'EMF. La principale raison derrière cette décision était de profiter du logiciel disponible à l'EMF qui nécessitait un fichier dans le format standard. Une autre raison, aussi importante, pour recoder les données c'est que les variables nécessaires pour l'analyse démographique n'apparaissent pas directement dans le questionnaire mais doivent provenir (ou être recodés) des données du questionnaire. Quand il y a un grand nombre de questions (le questionnaire rwandais a 249 questions plus huit questions pour chaque grossesse et six pour chaque mariage précédent), il est nécessaire d'avoir un programme spécial qui crée un fichier recodé pour simplifier la tabulation et une analyse plus approfondie.

Comme une vérification extensive des erreurs avait été faite au moment de la saisie des données, la validation des données n'a pas été faite avec des programmes spécifiques pour identifier et imprimer les questionnaires erronés, mais a été incorporée aux programmes utilisés pour la construction du dernier fichier des données. Le premier but de la validation des données était de confirmer que les données saisies étaient cohérentes et aussi de faire les vérifications qui avaient délibérément été exclues du programme de saisie des données. La validation a confirmé que les données étaient fondamentalement cohérentes et bien que des erreurs aient été détectées, elles n'ont pas été corrigées soit parce qu'elles n'étaient pas graves, soit parce que certaines dispositions avaient été prises pour leur correction (dates non déclarées et utilisation de la contraception contre connaissance).

Les données finales ou fichier recodé ont été créées selon les standards de l'EMF mais ont été modifiées à cause des restrictions physiques du nombre d'enregistrements de l'ordinateur NCR qui est de 512 caractères. Le fichier recodé standard de l'EMF contient plus de 200 variables, plus les historiques détaillés des grossesses et des mariages. Il est aussi stocké dans un format conçu pour faciliter la distribution et être utilisé sur une grande variété d'ordinateurs. La longueur minimum d'un enregistrement est alors de 690 caractères et après avoir ajouté les variables du module FOTCAF (facteurs autres que la contraception affectant la fécondité) et les variables spécifiques au pays, celle-ci approche des 1000 caractères. Le seul moyen qui nous permettait d'avoir toutes ces variables et de respecter la restriction de 512 caractères était d'utiliser un format spécifique du NCR (unsigned packed decimal) permettant d'avoir deux positions par caractère. Cette modification nécessitait des changements dans les programmes de l'EMF, mais cela n'a pas été un problème très important. Trois programmes furent utilisés pour la recodification : un programme d'extraction des dates pour créer le fichier dans la forme standard requise par le programme de contrôle, d'imputation et de recodification des dates (FDEIR) ; le programme de contrôle des dates ; et le programme de recodification pour créer le fichier final recodé comprenant les dates recodées et les variables spécifiques au Rwanda. Les programmes d'extraction et de recodification des données ont été créés en adaptant des programmes similaires utilisés pour l'Enquête sur la Fécondité au Nigeria et cela a grandement simplifié la tâche de programmation. Une description complète du fichier recodé se trouve en Appendix A de ce rapport.

Les tabulations produites pour l'enquête sont un sous-ensemble substantiel des tabulations recommandées pour le rapport national No 1 dans la publication "Recommandations pour le Rapport National No 1" et tous les tableaux standard associés au module FOTCAF. L'idée originale était de produire un nombre minimum de tableaux pour avoir un rapport montrant les résultats de base de l'enquête aussi vite que possible. La cohérence des données produites par le programme de saisie et les tableaux additionnels générés par le logiciel COCENTS ou qui étaient déjà parus dans la série des rapports nationaux No 1 standard nous ont permis d'aller bien plus loin que prévu à l'origine. Nous pensons toujours produire un rapport le plus tôt possible (la publication en est prévue en juin), mais ce rapport sera plus détaillé. Les tableaux additionnels nous ont aussi permis une analyse plus approfondie que prévue. Tous les tableaux, sauf les tableaux des ménages, ont été produits en utilisant le logiciel générateur de tableaux COCENTS de l'EMF et le progiciel de tabulation COCENTS. Une liste complète de tous les tableaux se trouve en Appendix B de ce rapport.