

ANNEXE 2 : CONCEPTION DE L'ECHANTILLON

Cet annexe décrit les principales caractéristiques de la conception de l'échantillon. Celles-ci incluent la taille de l'échantillon cible, sa répartition, son cadre d'échantillonnage et son inventaire des participants, le choix des domaines, les étapes de l'échantillonnage, la stratification et le calcul des coefficients de pondération de l'échantillon.

L'objectif premier de la conception d'un échantillon pour l'enquête par grappes à indicateurs multiples du Cameroun était de fournir des estimations statistiquement fiables de la plupart des indicateurs, au niveau national, pour les milieux urbain et rural, mais aussi pour les 12 domaines d'enquête que sont : Douala, Yaoundé, Adamaoua, Centre (hors Yaoundé), Est, Extrême-Nord, Littoral (hors Douala), Nord, Nord-Ouest, Ouest, Sud, Sud-Ouest. Les milieux urbain et rural de chacune des douze régions ont été définis comme des strates pour l'échantillonnage.

Une méthode d'échantillonnage par grappes, stratifiée et à deux degrés a été utilisée pour le choix de l'échantillon de l'enquête.

Taille et répartition de l'échantillon

La taille de l'échantillon cible de l'enquête MICS du Cameroun a été fixée à 9 865 ménages. Pour son calcul, l'indicateur clé utilisé a été le pourcentage d'enfants de 0-59 mois ayant reçu un traitement contre une infection respiratoire aiguë (IRA) au cours des deux dernières semaines. La formule ci-dessous a permis d'estimer la taille moyenne de l'échantillon requise pour ces indicateurs :

$$n = \frac{4(r)(1-r)(f)(1,1)}{(0,12r)^2(p)(n_h)}$$

Où

- 1 *n* est la taille de l'échantillon requise, exprimée en nombre de ménages ;
- 2 4 est le facteur permettant d'obtenir un niveau de confiance de 95% ;
- 3 *r* est la prévalence prévisionnelle ou attendue (taux de couverture) du traitement des IRA chez les enfants de 0-59 mois ;
- 4 1,1 est le facteur nécessaire pour élever la taille de l'échantillon de 10% pour raison de non-participation ;
- 5 *f* est le symbole abrégé de *deff* (effet du plan expérimental) ;
- 6 0,12*r* est la marge d'erreur tolérable à un niveau de confiance de 95% définie comme 12% de *r* (erreur d'échantillonnage relative de *r*) ;
- 7 *p* est la proportion de la population totale sur laquelle l'indicateur *r* est basé ;
- 8 *n_h* est la taille moyenne d'un ménage.

Pour le calcul, *r* (pourcentage d'enfants de 0-59 mois ayant reçu un traitement des IRA) a été fixé à 40%. La valeur de *deff* (effet du plan expérimental) a été fixée à 1,5 sur la base des estimations issues des enquêtes antérieures, celle de *p* (pourcentage d'enfants âgés de 0-4 ans au sein de la population totale) à 19% et celle de *n_h* (taille moyenne d'un ménage) à 4,5.

Le résultat de cette opération est d'environ 800 ménages. Ce chiffre correspond à la taille moyenne de l'échantillon requise dans chaque domaine, soit environ 9 800 ménages. La taille moyenne d'une grappe dans l'enquête MICS du Cameroun a été estimée à 19 ménages par grappe urbaine et 24 par grappe rurale, en raison de plusieurs considérations, dont le budget disponible et le temps nécessaire pour une équipe pour accomplir son travail auprès d'une grappe. La division du nombre total de

ménages par le nombre de ménages par grappe a donné un total de 34 à 45 grappes par région ou domaine, soit 466 grappes pour l'ensemble du pays.

Ce calcul visait une répartition de la taille de l'échantillon dans les 12 domaines d'étude. Pour cette raison, 34 à 45 grappes ont été allouées à chaque domaine, avec une taille finale de l'échantillon équivalent à un plus de 9 800 ménages (environ 39 grappes x 12 domaines x 21). Dans chaque domaine, les grappes (unités d'échantillonnage principales) ont été réparties entre le milieu urbain et rural, proportionnellement à la taille des

populations urbaine et rurale Le tableau ci-dessous présente la répartition des grappes entre les domaines de l'échantillonnage.

Tableau SD.1 : Répartition des grappes des échantillons (unités d'échantillonnage principales) entre les domaines d'échantillonnage

Domaine ou région	Ménages			ZD		
	Urbain	Rural	Total	Urbain	Rural	Total
Douala	822	0	822	45	0	45
Yaoundé	803	0	803	44	0	44
Adamaoua	330	470	800	17	19	36
Centre	219	627	847	12	27	39
Est	275	525	800	14	21	35
Extrême-Nord	201	697	898	11	30	41
Littoral	520	280	800	26	13	39
Nord	201	627	828	11	25	36
Nord-Ouest	274	558	832	15	24	39
Ouest	347	488	835	19	21	40
Sud	262	538	800	13	21	34
Sud-Ouest	311	489	800	17	21	38
Total	4565	5300	9865	244	222	466

Cadre d'échantillonnage et choix des grappes

Les travaux cartographiques de l'EDSC 3 ont servi de cadre pour le choix des grappes. Les zones de dénombrement issues de la cartographie du 3^{ème} Recensement Général de la Population et de l'Habitat (RGPH) ont été définies comme unités d'échantillonnage principales (UEP) et elles ont été choisies dans chacun des domaines d'échantillonnage sur la base de procédures d'échantillonnage ppt (probabilité proportionnelle à la taille) systématiques, en fonction des tailles estimées des zones de dénombrement d'après la cartographie de 2003 du 3^{ème} RGPH. Ainsi, pour exécuter la première étape de l'échantillonnage, le nombre de zones de dénombrement requis dans chacun des 12 domaines a été choisi en faisant la distinction entre le milieu urbain et le milieu rural.

Activités d'inventaire

Le cadre d'échantillonnage (travaux cartographiques de l'EDSC 3) n'étaient pas à jour, l'inventaire des ménages a été actualisé dans l'ensemble des zones de dénombrement choisies avant la sélection des ménages. A cet effet, des équipes d'inventaire ont été constituées, et elles ont séjourné dans chaque zone de dénombrement pour répertorier les logements occupés (ménages). Les équipes d'inventaire étaient en fait les agents enquêteurs et les contrôleurs de la MICS 3. Durant l'enquête, ils étaient chargés d'actualiser la cartographie et la liste des ménages dans les zones de dénombrement choisies avant de procéder au tirage des ménages et à l'interview de ceux-ci.

Choix des ménages

Pour chaque zone de dénombrement, les équipes d'enquêteurs sur le terrain ont répertoriés les ménages. Ensuite, ceux-ci ont été numérotés par ordre de 1 à n (n étant le nombre total de ménages dans chaque zone de dénombrement) et le contrôleur a procédé à la sélection d'un nombre variable de ménages fixé au départ par la Coordination pour chaque zone de dénombrement sur la base des procédures de sélection systématiques.

Calcul des pondérations de l'échantillon

Au bout du compte, l'échantillon de l'enquête par grappes à indicateurs multiples du Cameroun n'est pas auto-pondéré. En fait, faute de disposer des résultats du 3^{ème} RGPH la non uniformité du nombre de ménages alloué à chaque domaine (région) et à chaque zone de dénombrement a permis d'utiliser des fractions d'échantillonnage

différentes sur la base de la taille de chaque domaine et de chaque zone de dénombrement. Pour cette raison, les pondérations de l'échantillon ont été calculées et utilisées dans les analyses ultérieures des données de l'enquête.

La principale composante de cette pondération est l'inverse de la fraction d'échantillonnage qui a permis de déterminer le nombre de ménages par échantillon dans ce domaine d'échantillonnage spécifique :

$$W_h = 1 / f_h$$

Le terme f_h , la fraction d'échantillonnage au niveau de la strate h , est le produit des probabilités de choix à chaque étape de chaque domaine d'échantillonnage :

$$f_h = P_{1hi} * P_{2hi}$$

où

P_{1hi} est la probabilité de sélectionner la grappe i dans la strate h au premier degré ;

P_{2hi} est la probabilité de sélectionner un ménage dans la grappe i de la strate h au deuxième degré.

Les fractions d'échantillonnage des ménages de chaque zone (grappes) ont été calculées en raison de la différence entre le nombre estimatif de ménages par zone de dénombrement avant la sélection de la première étape (choix des unités d'échantillonnage principales) et le nombre révisé de ménages par zone. Ces fractions incluent donc la probabilité de choix de la zone de dénombrement appartenant à ce domaine d'échantillonnage spécifique ainsi que celle du choix d'un ménage appartenant à la zone de dénombrement de l'échantillon (grappe).

Une seconde composante à prendre en compte dans le calcul des pondérations de l'échantillon est le niveau de non-réponse aux enquêtes auprès des ménages et des personnes dans chaque grappe. L'ajustement pour la non-réponse des ménages est égal à l'inverse de la valeur suivante :

$$RR = \text{Nombre de ménages interrogés} / \text{Nombre de ménages occupés répertoriés}$$

À l'issue du travail de terrain, les taux de réponse ont été calculés pour chaque domaine d'échantillonnage. Ces taux ont servi à ajuster les pondérations d'échantillon calculées pour chaque grappe. Les taux de réponse de l'enquête MICS du Cameroun sont présentés dans le Tableau HH.1 du présent rapport.

De même, l'ajustement des cas de non-réponse au niveau individuel (femmes et enfants de moins de 5 ans) est égal à l'inverse de la valeur ci-dessous :

RR = Questionnaires pour femme (ou enfant de moins de 5 ans) remplis / Femmes (ou enfants de moins de 5 ans) concernés

Le nombre de femmes de 15-49 ans et d'enfants de moins de 5 ans concernés est issu de l'inventaire des ménages fait dans le Questionnaire dédié au ménage ayant répondu à l'enquête.

Pour chaque zone de dénombrement, les pondérations non ajustées des ménages ont été calculées en multipliant les facteurs ci-dessus. Ensuite, elles ont été standardisées (ou normalisées). Cette opération a pour finalité d'harmoniser la somme des unités d'échantillonnage interrogées avec la taille totale de l'échantillon au niveau national. La normalisation consiste à multiplier les pondérations non ajustées susmentionnées par le ratio du nombre de ménages interrogés sur le nombre total de ménages à pondération non ajustée. Une procédure de standardisation similaire a été utilisée pour calculer les pondérations standardisées des questionnaires dédiés aux femmes et aux enfants de moins de 5 ans. Les pondérations ajustées (normalisées) ont varié entre 0,217 et 4,669 pour les 466 zones de dénombrement (grappes).

Les pondérations de l'échantillon ont été annexées à l'ensemble des groupes de données et appliquées à chaque ménage, femme ou enfant de moins de 5 ans pour les besoins de calcul des indicateurs et de l'analyse.