

CHAPITRE 4

CONCEPTION ET TIRAGE DE L'ÉCHANTILLON

Ce chapitre technique¹ s'adresse principalement aux spécialistes de sondage, mais aussi au coordinateur et aux autres responsables techniques de l'enquête. Il leur permettra de :

- ▶ Connaître les caractéristiques d'une bonne conception d'un tirage probabiliste ;
- ▶ Estimer si l'échantillon disponible peut être utilisé, ou décider si un nouvel échantillon doit être tiré ;
- ▶ Décider de la taille du nouvel échantillon ;
- ▶ Accorder une certaine importance aux questions relatives aux indicateurs de niveau sous-national, et aux estimations sur l'eau et l'assainissement ;
- ▶ Être mieux informé sur la pondération, l'estimation et les erreurs de sondage ;
- ▶ Apprendre l'échantillonnage avec *ppt* et la stratification implicite ;
- ▶ Prendre connaissance des plans de sondage utilisés dans trois pays ayant participé à la série des enquêtes MICS 2000.

Effectuer une enquête à indicateurs multiples dans votre pays sera fait à partir d'une base de sondage, à l'opposé de la collecte de données pour une population entière. Il y a différentes populations cibles dans ce type d'enquête – les ménages, les femmes âgées de 15 à 49 ans ainsi que les enfants de moins de cinq ans et d'autres groupes d'âges. Malgré cette diversité de cibles, les *personnes enquêtées* seront habituellement les mères ou les personnes qui s'occupent des enfants dans chaque ménage visité². Il est important de reconnaître que la MICS3 est une enquête d'envergure nationale et qu'un échantillon est tiré de l'ensemble des ménages du pays, et pas seulement des ménages comptant des jeunes enfants ou des femmes en âge de procréer.

¹ Les utilisateurs de la version précédente du Manuel d'Enquête par Grappes à Indicateurs Multiples remarqueront que ce chapitre a été l'objet d'une révision importante. Au nombre des modifications intervenues figurent la présentation de la formule de calcul de la taille de l'échantillon, une plus grande attention accordée au développement et à la mise à jour de la base de sondage et au calcul de la marge d'erreur, ainsi que la présentation d'exemples de pays ayant participé à la série d'enquêtes MICS de 2000.

² Dans une Enquête par Grappes à Indicateurs Multiples, un *ménage* est défini comme un groupe de personnes qui vivent et qui prennent leur repas ensemble. Tout adulte informé (*défini, pour les besoins de la MICS3, comme une personne âgée de 15 ans ou plus*) est éligible pour être le répondant principal au Questionnaire Ménage. Toutefois, dans la plupart des cas, le répondant principal sera la mère ou la principale personne en charge des enfants, pour la simple et bonne raison que ce sont ces personnes qui sont le plus, susceptibles d'être trouvées à la maison au moment du passage des agents enquêteurs.

DETERMINER LA TAILLE DE L'ÉCHANTILLON

La taille de l'échantillon est peut-être le paramètre le plus important du plan de sondage car elle affecte la précision, le coût et la durée de l'enquête plus que tout autre facteur. La taille de l'échantillon doit être fixée en fonction, à la fois, du budget de l'enquête disponible et des besoins en précision. Ce dernier facteur doit être considéré du point de vue des besoins en estimations au niveau national et au niveau sous-national. De plus, la taille globale de l'échantillon ne peut pas être fixée indépendamment du nombre de zones d'échantillonnage - ou unités primaires de sondage (UPS) - et de la taille finale des grappes. Ainsi, bien qu'il existe des formules mathématiques pour calculer la taille de l'échantillon, il est nécessaire de prendre en compte tous ces facteurs avant de prendre une décision finale.

Obtenir de l'aide

Bien qu'assez détaillé, ce chapitre du manuel ne prétend pas faire de ses lecteurs des experts statisticiens en sondage. De nombreux aspects de la conception du plan de sondage nécessitent l'assistance d'un spécialiste, soit d'un Bureau National de Statistiques, soit de l'extérieur. Ces aspects peuvent comprendre le calcul de la taille de l'échantillon, la mise en ordre de la base de sondage, l'évaluation des options du plan de sondage, l'application de l'approche des *ppt* à la base de sondage, le calcul des poids et des erreurs de sondage des indicateurs. Dans tous les cas, il est fortement recommandé que le Bureau National de Statistiques de votre pays soit consulté sur le plan de sondage.

Deux règles générales régissent le choix du nombre d'UPS et la taille des grappes : plus le nombre d'UPS sélectionnées est important, plus la représentation géographique et la fiabilité d'ensemble de l'échantillon seront importantes ; plus la taille des grappes est petite, plus les estimations seront fiables.

EXEMPLE :

Une enquête nationale de 600 UPS avec une taille de grappe de 10 ménages dans chacune d'elles fournira des résultats plus fiables que 400 UPS avec des tailles de grappes de 15 ménages, même si les deux enquêtes ont une taille globale d'échantillon de 6.000 ménages. De plus, une taille de grappe de 10 ménages est meilleure que 15 parce que la fiabilité de l'enquête s'améliore également du fait des plus petites tailles des grappes. Ainsi, en résumé, à condition que les autres facteurs soient les mêmes, il est préférable d'utiliser un plus grand nombre d'UPS et des UPS de plus petite taille.

En général, plus le nombre d'UPS sélectionnées est important, meilleure est la qualité de votre enquête. Toutefois, le nombre d'UPS de l'enquête sera affecté, dans une grande mesure, par des considérations de coût et par le besoin éventuel d'estimations à un niveau sous-national (*les estimations à un niveau sous-national sont présentées plus loin dans le chapitre*). Le coût du transport est un facteur-clé. Si la distance entre les UPS de l'échantillon est grande et que les mêmes équipes d'enquêteurs doivent voyager de grappe en grappe (par opposition à l'utilisation

d'enquêteurs résidant dans chaque UPS), alors la réduction du nombre d'UPS sélectionnées diminuera considérablement le coût global de l'enquête. Par contre, si l'enquête a pour objectif de fournir des estimations à un niveau sous-national, il sera nécessaire de sélectionner davantage d'UPS.

Le choix de la taille des grappes de l'enquête est un autre paramètre qui doit être pris en compte lors de la détermination de la taille de l'échantillon. Son effet peut être évalué par ce que l'on appelle l'effet du plan de sondage, ou *deff*. Le *deff* est une mesure qui compare les ratios de la variance d'échantillonnage de l'échantillon de l'enquête stratifié par grappe (l'enquête MICS3 dans notre cas) à un échantillon aléatoire simple³ de même taille globale. Si, par exemple, la valeur calculée de l'effet du plan de sondage, *deff*, à partir de l'indicateur de l'enquête était 2,0, cela voudrait dire que l'estimation a une variance d'échantillonnage deux fois plus grande que la variance d'échantillonnage d'un échantillon aléatoire simple de même taille globale.

Plusieurs exemples spécifiques sur la manière de choisir le nombre d'UPS et de décider de la taille des grappes sont fournis à la fin de cette section sur la taille de l'échantillon.

Du fait de son coût, l'utilisation d'un échantillon aléatoire simple n'est pas envisageable dans le cas de la MICS3, ce qui explique pourquoi on utilise plutôt des sondages par *grappe*. Les facteurs qui déterminent l'effet du plan de sondage sont la stratification, la taille des grappes et l'homogénéité des grappes – autrement dit dans quelle mesure deux personnes (ou ménages) de la grappe ont les mêmes caractéristiques. L'homogénéité des grappes peut être illustrée, par exemple, par le fait que deux enfants vivant à proximité l'un de l'autre ont une plus forte probabilité d'avoir reçu les mêmes vaccinations que deux enfants vivant à deux endroits différents choisis au hasard dans la population.

La stratification diminue généralement la variance d'échantillonnage, alors que la mesure d'homogénéité et la taille de la grappe l'augmentent. Par conséquent, un objectif de votre plan de sondage sera de choisir une taille de grappe de façon à équilibrer l'homogénéité, pour laquelle une petite taille est préférable, et le coût, pour lequel une grande taille est généralement préférable.

L'effet du plan de sondage doit être pris en compte dans la formule de calcul de la taille de l'échantillon. Cependant, deux problèmes se posent. Tout d'abord, alors que la valeur de *deff* peut être facilement calculée *après l'enquête*, elle est souvent inconnue avant l'enquête, à moins que des enquêtes portant sur les mêmes variables aient été réalisées précédemment. En outre, la valeur de *deff* diffère pour chaque indicateur et, en fait, pour chaque groupe cible parce que l'homogénéité des grappes varie selon les caractéristiques. Bien sûr, même si l'on connaissait les valeurs de *deff*, il ne serait pas pratique de mener une enquête avec *différentes* tailles d'échantillon pour chacune des caractéristiques basées sur leurs différents *deffs*.

³ Un type de sondage probabiliste dans lequel n unités de sondage ont été sélectionnées à probabilité égale dans une population de N unités, généralement sans remplacement et en utilisant une table de nombre au hasard.

Les différents *deffs* ne seront généralement pas connus avant l'enquête pour tous les indicateurs, mais on peut s'attendre à ce qu'ils soient relativement faibles pour de nombreux indicateurs, c'est-à-dire ceux basés sur des *sous-groupes très spécifiques* (par exemple, les enfants de 12-23 mois)⁴. S'il existe une enquête ménage antérieure qui a collecté des données similaires à celles de la MICS, et qui a utilisé un plan de sondage similaire, vous pouvez utiliser les *deffs* de cette enquête précédente pour estimer les effets probables des erreurs du plan de sondage pour la MICS3. Peu d'enquêtes ménages calculent les *deffs*, mais les enquêtes démographiques et de santé (EDS) constituent une bonne source d'information dans ce domaine.

Dans la formule de calcul ainsi que dans le tableau de la taille de l'échantillon présentés dans les sections qui suivent, on a supposé que l'effet du plan de sondage est de 1,5 (ce qui est une approche prudente, car cet effet du plan de sondage est un peu plus important qu'en réalité). En optant pour une valeur prudente de *deff*, on veut s'assurer que la taille de l'échantillon serait assez grande pour permettre le calcul de tous les principaux indicateurs. Néanmoins, dans le choix de la taille des grappes et, implicitement, le nombre de grappes, une règle générale et l'approche recommandée consistent à *s'assurer que la taille des grappes soit aussi petite que possible pour être utilisée efficacement sur le terrain*, tout en tenant compte des autres facteurs associés comme le nombre d'UPS et les coûts du terrain (*présentés ci-dessus*), et tout en garantissant une charge de travail raisonnable pour les enquêteurs.

CALCUL DE LA TAILLE DE L'ECHANTILLON

Le calcul de la taille de l'échantillon à l'aide de formules mathématiques appropriées nécessite que certains facteurs soient spécifiés et, que pour d'autres, vous posiez des hypothèses ou que vous utilisiez des valeurs tirées d'enquêtes précédentes ou similaires. Ces facteurs sont les suivants :

- La précision ou la marge d'erreur *relative* recherchée ;
- Le niveau de confiance souhaité ;
- La proportion de la population estimée (ou connue) dans un groupe cible donné ;
- Le taux de couverture - ou la prévalence - prévu ou anticipé d'un indicateur donné ;
- L'effet du plan de sondage (*deff*) ;
- La taille moyenne du ménage ;
- Un coefficient d'ajustement pour les cas éventuels de non-réponses.

Le calcul de la taille de l'échantillon est compliqué par le fait que certains de ces facteurs varient selon l'indicateur. Nous avons déjà mentionné le fait que les *deffs* diffèrent. Il en est de même de la marge d'erreur qui ne sera probablement pas la même pour chaque indicateur (en pratique, elle ne peut pas l'être). Cela signifie que différentes tailles d'échantillon seront nécessaires pour

⁴ L'expression mathématique du *deff* est une fonction du *produit* de l'homogénéité de la grappe et de la taille de la grappe. Même si la taille de la grappe est importante en terme de ménages, elle peut être faible en terme de populations cibles particulières (enfants âgés d'un an) et, par conséquent, il est probable que le *deff* soit également peu important.

différents indicateurs pour obtenir la précision nécessaire. Manifestement, nous devons décider d'une taille d'échantillon pour l'enquête.

Le calcul de la taille de l'échantillon s'applique seulement aux personnes, même si elle est exprimée en termes de nombre de ménages que vous devriez visiter pour interviewer des individus. Cela s'explique par le fait que les indicateurs les plus importants de l'évaluation de la MICS3 sont basés sur des individus. Les variables ménages ne doivent pas être utilisés dans les calculs des tailles d'échantillon de ménages car elles exigeraient non seulement une formule différente, mais également des valeurs différentes d'effets de plan de sondage (*deff*) qui s'élèveraient à 10 ou plus.

La formule de calcul est donnée ci-dessous :

$$n = \frac{[4 (r) (1-r) (f) (1.1)]}{[(0,12r)^2 (p) (n_h)]}$$

Où :

- **n** est la taille requise de l'échantillon - exprimée en nombre de ménages - pour l'indicateur-CLÉ (voir la section qui suit pour la détermination de cet indicateur clé)
- **4** est un facteur pour atteindre 95% d'intervalle de confiance
- **r** est la prévalence (taux de couverture) prévue ou anticipée pour l'indicateur-clé à estimer
- **1,1** est le facteur nécessaire pour augmenter la taille de l'échantillon de 10% afin de tenir compte du taux de non réponse
- **f** est le symbole représentant l'effet du plan de sondage - *deff*
- **0,12r** est la marge d'erreur raisonnable pour un intervalle de confiance de 95%, définie comme 12 pour cent de r (12% est donc la marge d'erreur relative de r)
- **p** est la proportion de la population totale sur laquelle l'indicateur, r, est basé, et
- **n_h** est la taille moyenne du ménage.

Si, pour une enquête, la taille de l'échantillon est calculée en utilisant un indicateur-clé basé sur le plus petit groupe cible, en termes de proportion de la population totale, alors la précision des estimations de l'enquête pour les autres principaux indicateurs n'en seront que meilleures.

Les utilisateurs particulièrement observateurs du manuel de MICS2 remarqueront que la différence dans cette formule réside dans le fait que la *marge d'erreur relative* (valeur de 0,12r) remplace la *marge d'erreur* (**e** dans l'édition précédente, avec une valeur de 0,05 ou de 0,03 selon qu'il s'agit d'indicateurs à couverture élevée ou faible respectivement). Dans le manuel de MICS2, une estimation fiable pour l'enquête a été définie différemment, selon qu'elle représente une couverture élevée ou faible.

Pour estimer les indicateurs, il avait été recommandé de fixer la marge d'erreur ou de précision à 5 points de pourcentage pour les taux de couverture (par exemple, les vaccinations) qui sont comparativement élevés, supérieurs à 25%, et à 3 points de pourcentage pour les taux de couverture qui sont faibles, 25% ou moins. Quoiqu'une justification possible ait été donnée à la présentation de ces deux marges d'erreur définies de la sorte, les utilisateurs étaient néanmoins souvent confrontés à une difficulté, à savoir quelle marge utiliser pour leur enquête, en

particulier lorsque les tailles d'échantillon calculées étaient très différentes. L'utilisation de la marge d'erreur relative permet d'éviter ce problème, car réduisant la marge d'erreur et résultant en une précision comparable, indépendamment du fait qu'un indicateur à couverture élevée ou faible soit retenu comme indicateur-clé pour la détermination de la taille de l'échantillon. Vous remarquerez, cependant, que la taille de l'échantillon est tout de même plus grande pour des indicateurs à faible couverture, raison pour laquelle il est crucial de choisir soigneusement quel est l'indicateur qui est *le plus clé* pour l'enquête (voir la section suivante).

DEFINIR ET CHOISIR L'INDICATEUR-CLE POUR CALCULER LA TAILLE DE L'ECHANTILLON

Choisir un indicateur important qui permettra la plus grande taille d'échantillon constitue la stratégie recommandée pour calculer la taille de l'échantillon. Cela signifie qu'en premier lieu, il faut choisir la population cible qui représente une petite proportion de la population totale (p dans la formule ci-dessus). C'est généralement *une population cible ne comprenant qu'un seul groupe d'âges*⁵. Dans la MICS3, il s'agira des enfants de 12-23 mois, qui représentent environ 2,5% de la population totale dans de nombreuses enquêtes pays MICS3. *Nous recommandons d'utiliser 2,5% sauf si vous avez de meilleures estimations disponibles pour votre pays*. Si, par exemple, votre chiffre est plus élevé (3,5, 4 ou 5%), les tailles de vos échantillons seront largement moindres que celles du Tableau 4.3, aussi est-il crucial d'utiliser votre meilleure estimation de p pour cette population cible. Deuxièmement, l'indicateur particulier doit être choisi pour cette même population cible. Nous le qualifierons d'*indicateur-clé* (mais seulement à des fins de calcul de la taille de l'échantillon).

Tableau 4.1

Taux de couverture des indicateurs (Prévalence ou Proportion)

Faible couverture *pas souhaitable* :

- Accès à l'eau salubre potable ou aux toilettes
- Fréquentation scolaire
- Soins prénatals et accouchements assistés
- Taux d'allaitement
- Taux de couverture de la vaccination

Faible couverture *souhaitable* :

- Taux de mortalité
- Prévalence de la malnutrition (émaciation, retard de croissance, insuffisance pondérale)
- Travail des enfants

⁵ En choisissant les groupes qui représentent les plus faibles proportions dans la population, nous vous recommandons fortement d'exclure les enfants de quatre mois qui forment la base des indicateurs de l'allaitement parce que les tailles nécessaires de l'échantillon risqueraient probablement d'être beaucoup trop grandes

Vous devez choisir votre indicateur-clé parmi ceux qui ont un faible taux de couverture. Cependant, certains indicateurs ayant un faible taux de couverture doivent être exclus. Le Tableau 4.1 qui présente les indicateurs vous aidera à comprendre pourquoi certains choix sont exclus. En effet, ce tableau donne des exemples d'indicateurs pour lesquels une faible couverture n'est pas *souhaitable*, et dont l'objectif est focalisé sur l'*augmentation* du taux (par exemple le taux de vaccination du DTCOq, c'est-à-dire la diphtérie, le tétanos et le choléra). Le deuxième groupe d'indicateurs présenté dans le Tableau 4.1 donne des exemples de la situation opposée - une faible couverture est *souhaitable* et l'objectif est de réduire davantage ce taux dans l'avenir (par exemple, la prévalence du retard de croissance). *Il ne serait pas pertinent de baser votre taille d'échantillon sur des indicateurs pour lesquels une faible couverture est souhaitée et dont la couverture est déjà très faible ; de tels indicateurs devraient être exclus lors du choix de l'indicateur-clé.*

Le Tableau 4.2 présente certaines suggestions pour choisir un groupe cible et un indicateur-clé en vue de calculer directement la taille de votre échantillon ou pour trouver la taille de l'échantillon dans le Tableau 4.3. Notez que ni le taux de mortalité infantile, ni le taux de mortalité maternelle (MMR)⁶ ne figurent parmi les possibilités de choix de l'indicateur-clé. Cela s'explique par le fait que les tailles d'échantillons qui sont nécessaires pour mesurer ces deux indicateurs sont beaucoup trop grandes – de l'ordre de dizaines de milliers – ce qui est pratiquement irréalisable. Cela ne signifie pas que de tels indicateurs ne devraient pas être mesurés dans l'enquête, mais plutôt que la taille de l'échantillon de l'enquête ne doit pas être basée sur ces indicateurs. Les résultats de l'enquête pour ces indicateurs seront entachés d'erreurs de sondage plus importantes, et par conséquent, d'intervalles de confiance plus grands que pour les autres indicateurs.

⁶ En ce qui concerne la taille de l'échantillon nécessaire pour le taux de mortalité maternelle, un guide préparé par l'OMS et l'UNICEF en 1997, intitulé "*The Sisterhood Method for Estimating Maternal Mortality (Estimation de la Mortalité Maternelle par la Méthode des Sœurs)*" recommande que si le taux est de 300 (pour 100 000 naissances vivantes), il peut être estimé avec un échantillon d'environ 4.000 enquêtés avec une marge d'erreur de 60 environ, en utilisant la méthode indirecte des sœurs.

Tableau 4.2**Liste de contrôle des indicateurs et des groupes-cibles**

Pour décider de quel groupe cible et de quel indicateur vous avez besoin pour déterminer la taille de votre échantillon :

- 1 Choisissez deux ou trois groupes-cibles de la population représentant les plus petits pourcentages de la population totale. Normalement, ces groupes-cibles ne devraient pas être aussi réducteurs que des groupes d'âges de un an, ou plus grands que des groupes d'âges quinquennaux. Dans la MICS3, le cas typique sera constitué des enfants de 12-23 mois, ou des enfants de moins de 5 ans qui, dans la plupart des pays, seront compris entre 2 et 4%, et 10 et 20% de la population totale respectivement.
- 2 Passez en revue les principaux indicateurs relatifs à ces groupes, en laissant de côté ceux qui ont des prévalences très faibles (inférieures à 5%) ou très élevées (supérieures à 50%). Commencez vos calculs par le plus petit groupe. Si les indicateurs se rapportant à ce groupe ont une couverture importante, effectuez les calculs avec le groupe le plus grand, pour lesquels les indicateurs devraient en principe avoir une couverture moins élevée.
- 3 De manière générale, choisissez un indicateur qui affiche un taux de couverture relativement bas pour les populations-cibles représentant 10 à 15% de la population, de l'ordre de 15 ou 20%. Pour les populations-cibles représentant moins de 5% de la population totale, choisissez un indicateur qui a une couverture relativement plus grande, supérieure à 20%, mais inférieure à 50%.
- 4 Ne choisissez pas, parmi les indicateurs pour lesquels ont souhaite trouver un faible taux de couverture, un indicateur qui est déjà reconnu comme ayant un niveau bas et acceptable.

En effectuant vos choix, vous devrez également tenir compte de l'importance *relative* des différents indicateurs de votre pays. Par exemple, vous n'auriez pas à utiliser un indicateur qui exigera une taille d'échantillon très grande si, comparativement, cet indicateur ne présente que peu d'intérêt pour votre pays.

UTILISER LE TABLEAU SUR LA TAILLE DE L'ÉCHANTILLON

Le Tableau 4.3 présente des tailles d'échantillons déjà calculés sur la base des besoins de la MICS3 et de certaines hypothèses de départ. Vous devriez utiliser les valeurs du tableau, s'ils correspondent à la situation de votre pays, pour obtenir la taille de votre échantillon. Dans le cas contraire, votre spécialiste en échantillonnage devra calculer la taille de l'échantillon directement, en se servant de la formule donnée ci-dessous.

Si les paramètres du Tableau 4.3 correspondent à la situation de votre pays, vous pouvez trouver la taille de votre échantillon sans avoir besoin de la calculer en vous servant de la formule ci-dessus. Dans le Tableau 4.3, le niveau de confiance pour la précision des estimations est

prédéfini à 95%. Des valeurs variables pour la taille moyenne des ménages et pour le taux de couverture, r , sont utilisées – de 4,0 à 6,0 et de 0,25 à 0,40. L'effet du plan de sondage – $deff$ – est fixé à 1,5 et le niveau de précision (marge d'erreur) est fixé à 12% de r , autrement dit à 12% de la marge d'erreur relative. Dans le tableau, il a été effectué un ajustement vers le haut de 10% dans la taille de l'échantillon pour tenir compte des possibilités de non-réponses dans l'enquête.

Il est très important de noter que le tableau postule également que la population-cible pour votre indicateur-clé représente 2,5% de la population totale. Dans le cas où cette valeur est différente, vous ne pouvez pas utiliser le tableau pour trouver la taille de l'échantillon dont vous avez besoin. De manière générale, le tableau ne peut pas être utilisé si une seule des valeurs postulées comme paramètres de la formule ne correspond pas à votre situation. Plus loin dans cette section, davantage de détails sont donnés sur ce qu'il convient de faire dans cette situation.

Tableau 4.3

Taille d'Échantillon (Ménages) pour Différentes Estimations des Taux de Couverture pour la Plus Petite Population Cible

(avec une Marge d'Erreur Relative de 12% du Taux de Couverture au Niveau de Confiance de 95%)

Taille moyenne des ménages (nombre de personnes)	Taux de Couverture (r)			
	$r = 0,25$	$r = 0,30$	$r = 0,35$	$r = 0,40$
4,0	13.750	10.694	8.512	6.875
4,5	12.222	9.506	7.566	6.111
5,0	11.000	8.556	6.810	5.500
5,5	10.000	7.778	6.191	5.000
6,0	9.167	7.130	5.675	4.583

Utilisez ce tableau quand votre

- Population-cible représente 2,5% de la population totale ; ce sont généralement les enfants de 12-23 mois
- Effet du plan de sondage – $deff$ – est fixé à 1,5 et le taux de non-réponse attendu est estimé à 10%
- Marge d'erreur relative est fixée à 12% de l'estimation du taux de couverture, r

Si toutes les hypothèses pour les valeurs des paramètres de la formule sont applicables à votre pays, alors l'une des tailles d'échantillon du Tableau 4.3 devrait correspondre à votre situation. Dans certains cas, les paramètres peuvent s'appliquer, mais le taux de couverture que vous avez choisi nécessite d'être interpolé. Par exemple, si votre taux de couverture se situe entre 30 et 35%, vous trouverez la valeur de la taille de votre échantillon en interpolant une valeur située entre la 3^{ème} colonne et la 4^{ème} colonne du tableau. A titre d'exemple, sur la dernière ligne pour un taux de couverture de 32,5%, votre taille d'échantillon devrait être comprise entre 7.130 et 5.675, soit environ 6.403 ménages.

Un exemple des étapes à suivre pour l'utilisation du Tableau 4.3 est fourni ci-dessous :

- Tout d'abord, assurez-vous que toutes les valeurs des paramètres utilisées dans le Tableau 4.3 s'appliquent à votre situation.
- Ensuite, dans le Tableau 4.2, choisissez l'indicateur qui a la plus faible couverture, en excluant tout indicateur qui a déjà un niveau bas et acceptable. Supposons qu'il s'agit de la couverture vaccinale contre la rougeole, à 35%.
- Puis retrouvez, dans le Tableau 4.3, la taille moyenne des ménages qui se rapproche le plus de celle de votre pays (en supposant qu'elle fasse partie des intervalles spécifiés). Supposons qu'elle est de 5,5 personnes.
- Enfin, dans le Tableau 4.3, trouvez la valeur correspondant à une taille moyenne de ménage de 5,5 personnes et à un taux de couverture de 35%. Cette valeur est de 6.191.

Cependant, les chiffres ne devraient pas être considérés comme exacts, mais uniquement comme des tailles d'échantillon approximatives. Il serait pertinent d'arrondir les tailles d'échantillons trouvés à la hausse ou à la baisse, en fonction des contraintes budgétaires. Dans l'exemple ci-dessus, il vous faudra décider si 6.100 ou 6.200 seraient plus indiqués, une fois que vous auriez apprécié les coûts des déplacements entre les unités primaires de sondage, les tailles des grappes et la charge de travail des agents enquêteurs.

UTILISATION DE LA FORMULE DE CALCUL DE L'ECHANTILLON⁷

Que faire pour calculer la taille de l'échantillon si, dans votre pays, toutes les hypothèses des paramètres s'accordent, hormis la proportion des enfants de 12-23 mois qui ne serait pas de 2,5%, mais serait plus proche de 2% ? Dans ce cas, il vous suffit simplement de multiplier tous les nombres du Tableau 4.3 par 2,5/2, ou 1,25, pour obtenir les tailles de l'échantillon. Cette taille est, bien sûr, trop grande puisque les tailles d'échantillon seraient de 25% plus grandes que celles du tableau.

Il existe, cependant, des situations où il est préférable d'ignorer le Tableau 4.3 et de calculer la taille de l'échantillon directement, en utilisant la formule ci-dessous. Cette formule doit être utilisée à chaque fois que l'une des valeurs des paramètres dans votre pays est différente des hypothèses utilisées dans le Tableau 4.3. Le Tableau 4.4 met en exergue les conditions d'application de la formule.

Les raisons pour lesquelles la taille de l'échantillon devrait être plus grande si p est inférieure à 0,025 ont déjà été présentées. A moins que les exemples illustrés ci-dessus ne correspondent à la situation de votre pays, la formule ci-dessous devrait être utilisée pour le calcul de la taille de l'échantillon. A titre de rappel, la formule ci-dessous devrait être utilisée à chaque fois que l'une des valeurs des paramètres remplit les critères indiqués au Tableau 4.4.

⁷ Un tableau Excel de calcul de la taille de l'échantillon est disponible à www.childinfo.org

Tableau 4.4**Liste de Contrôle pour utiliser la Formule de Calcul de la Taille de l'Echantillon**

La formule pour déterminer la taille de l'échantillon :

$$n = \frac{[4 (r) (1-r) (f) (1.1)]}{[(0.12r)^2 (p) (n_h)]}$$

Utiliser la formule ci-dessus si l'une des conditions suivantes correspond à la situation de votre pays :

- La proportion des enfants âgés d'un an est différente de 0,025 ;
- La taille moyenne du ménage (n_h) est en dehors de 4, - 6,0 personnes ;
- Le niveau de l'indicateur-clé (r) est en dessous de 25% ;
- L'effet de sondage (f) pour l'indicateur-clé est différent de 1,5, selon les estimations validées et tirées d'autres enquêtes dans le pays ;
- Le taux de non-réponse anticipé est plus grand ou plus petit que 10%.

Ne changez pas le niveau de confiance dans la formule, mais donnez-lui comme valeur "4".

L'utilisation de la formule est relativement facile, étant donné que c'est de l'arithmétique simple, une fois que les valeurs des paramètres sont insérées. Par exemple, pour un $r = 0,25$, $f = 1,6$, et un coefficient d'ajustement des non réponses = 1.05, $p = 0.035$ et $n_h = 6$, nous avons :

$$n = \frac{[4 (0,25) (1-0,25) (1,6) (1,05)]}{[(0,12 \times 0,25)^2 (0,035) (6)]} = \frac{1,26}{0,000189} = 6,667$$

Lors des enquêtes MICS précédentes, la taille de l'échantillon se situait habituellement entre 4.000 et 8.000 ménages. Cet intervalle est une cible réaliste à atteindre pour vous lors du calcul de la taille de l'échantillon, en tenant compte à la fois des impératifs de fiabilité et des contraintes budgétaires. Comme susmentionné, la MICS3 devrait fournir des estimations pour de nombreux indicateurs, chacun de ces derniers ayant son propre niveau de précision. Par conséquent, il est utile d'examiner les niveaux de fiabilité approximatifs – erreurs-types et limites de confiance – sur vos indicateurs pour une taille donnée de l'échantillon.

Le Tableau 4.5 illustre ces niveaux de fiabilité pour un échantillon de 6.000 ménages, qui devrait être considéré comme une taille d'échantillon courante pour la production d'estimations comparativement fiables pour la plupart des indicateurs de référence de la MICS3.

Tableau 4.5
Mesures de la fiabilité attendue (Erreurs-types et Intervalles de Confiance) pour des
échantillons de 6.000 ménages sous différentes situations démographiques

Taille moyenne des ménages	Taille de l'indicateur r	Taille de la sous-population P	Nombre de personnes échantillonnées dans la sous-population	Nombre de personnes ayant l'indicateur	Ecart-type	Intervalle de Confiance (niveau de 95%)	
						Marge inférieure	Marge supérieure
4	0,10	0,025	540	54	0,016	0,068	0,132
		0,05	1.080	108	0,011	0,078	0,122
		0,125	2.700	270	0,007	0,086	0,114
		0,20	4.320	432	0,006	0,089	0,111
	0,20	0,025	540	108	0,021	0,158	0,242
		0,05	1.080	216	0,015	0,170	0,230
		0,125	2.700	540	0,009	0,181	0,219
		0,20	4.320	864	0,007	0,185	0,215
	0,30	0,025	540	162	0,024	0,252	0,348
		0,05	1.080	324	0,017	0,266	0,334
		0,125	2.700	810	0,011	0,278	0,322
		0,20	4.320	1.296	0,009	0,283	0,317
	0,50	0,025	540	270	0,026	0,447	0,553
		0,05	1.080	540	0,019	0,463	0,537
		0,125	2.700	1.350	0,012	0,476	0,524
		0,20	4.320	2.160	0,009	0,481	0,519
5	0,10	0,025	675	68	0,014	0,072	0,128
		0,05	1.350	135	0,010	0,080	0,120
		0,125	3.375	338	0,006	0,087	0,113
		0,20	5.400	540	0,005	0,090	0,110
	0,20	0,025	675	135	0,019	0,162	0,238
		0,05	1.350	270	0,013	0,173	0,227
		0,125	3.375	675	0,008	0,183	0,217
		0,20	5.400	1.080	0,007	0,187	0,213
	0,30	0,025	675	203	0,022	0,257	0,343
		0,05	1.350	405	0,015	0,269	0,331
		0,125	3.375	1.013	0,010	0,281	0,319
		0,20	5.400	1.620	0,008	0,285	0,315
	0,50	0,025	675	338	0,024	0,453	0,547
		0,05	1.350	675	0,017	0,467	0,533
		0,125	3.375	1.688	0,011	0,479	0,521
		0,20	5.400	2.700	0,008	0,483	0,517
6	0,10	0,025	810	81	0,013	0,074	0,126
		0,05	1.620	162	0,009	0,082	0,118
		0,125	4.050	405	0,006	0,088	0,112
		0,20	6.480	648	0,005	0,091	0,109
	0,20	0,025	810	162	0,017	0,166	0,234
		0,05	1.620	324	0,012	0,176	0,224
		0,125	4.050	810	0,008	0,185	0,215
		0,20	6.480	1.296	0,006	0,188	0,212
	0,30	0,025	810	243	0,020	0,261	0,339
		0,05	1.620	486	0,014	0,272	0,328
		0,125	4.050	1.215	0,009	0,282	0,318
		0,20	6.480	1.944	0,007	0,286	0,314
	0,50	0,025	810	405	0,022	0,457	0,543
		0,05	1.620	810	0,015	0,470	0,530
		0,125	4.050	2.025	0,010	0,481	0,519
		0,20	6.480	3.240	0,008	0,485	0,515

La colonne 4 du Tableau 4.5 indique le nombre attendu de personnes à interviewer dans un échantillon de 6.000 ménages, en supposant un taux de non réponse de 10%. Par exemple, dans

un pays où la taille moyenne des ménages est de quatre personnes, le nombre de personnes échantillonnées dans une sous-population représentant 2,5% de la population totale (disons, les enfants âgés de 12-23 mois) serait d'environ 540, au lieu de 600, après avoir tenu compte des non réponses. Parmi ces personnes, la colonne 5 indique le nombre attendu de personnes échantillonnées qui auraient la caractéristique, r . Le nombre attendu de personnes échantillonnées serait 54 si r est de 10%, de 108 si r est de 20%, 162 si r est de 30%, et de 270 si r est de 50%.

Vous remarquerez que l'écart-type varie considérablement selon la taille de la sous-population et la taille de l'indicateur. Pour l'évaluation de vos résultats, l'intervalle de confiance – dans la colonne du Tableau 4.5 – est une mesure importante de leur fiabilité. L'intervalle de confiance, ou IC, indique l'étendue au sein de laquelle votre estimation devrait varier pour une valeur vraie dans la population, en prenant en compte l'écart-type. Il est calculé en ajoutant et en soustrayant le double de l'écart-type (pour 95% d'intervalle de confiance) à l'estimation de l'indicateur. En regardant la toute dernière ligne du Tableau 4.5, l'intervalle de confiance pour une estimation d'indicateur de 0,50 est de $|0,485 - 0,515|$. Cela signifie que si vous estimez à 0,50 la prévalence de votre indicateur, alors vous pouvez être sûr à 95% que la vraie valeur de votre indicateur au sein de la population est entre 48,5% et 51,5%.

EXEMPLES ILLUSTRANT LA MANIÈRE DE DECIDER DU NOMBRE D'UNITES PRIMAIRES DE SONDRAGE ET DES TAILLES DES GRAPPES

Au début de la section sur la taille de l'échantillon, nous avons discuté de la manière dont le nombre d'UPS et la taille des grappes jouent un rôle sur la détermination de la taille de l'échantillon. Nous avons mis en exergue le fait que plus le nombre d'UPS sera élevé et la taille des grappes sera petite, plus l'échantillonnage sera fiable. Nous concluons cette section par trois exemples en utilisant différents scénarios pour illustrer les interrelations entre la taille de l'échantillon, le nombre d'UPS et la taille de la grappe.

EXEMPLE 1 :

Groupe cible :	Enfants de 12-23 mois
Pourcentage de la population :	2,6%
Indicateur-clé :	Couverture de la vaccination du DTCoq
Prévalence (couverture) :	40%
<i>Deff</i> :	Pas d'information
Taille moyenne du ménage	6,0

Pour ce scénario, utilisez le Tableau 4.3 parce que le taux de couverture de l'indicateur-clé et la taille du ménage peuvent être trouvés dans le tableau. La population-cible, représentant 2,6%, est aussi très proche des 3% de l'exemple sur lequel est basé le tableau 4.3. Si vous ne possédez aucune information sur l'effet du plan de sondage, postulez une valeur de 1,5, et le facteur d'ajustement des non-réponses est fixé à 1,1, correspondant à un taux de non-réponse attendu de 10%. La taille de l'échantillon pour une taille moyenne de ménage de 6,0 personnes pour un taux de couverture de 49% sera alors de 4.583 ménages.

Supposez que votre pays soit relativement vaste et que de plus, il y ait un grand nombre de provinces, par exemple, 15. Votre équipe d'échantillonnage et vous-mêmes avez déjà conclu que vous avez besoin d'un minimum de 300 UPS pour couvrir le territoire et pour obtenir que votre échantillon soit représentatif au niveau de chaque province. De plus, vous avez décidé que le budget de l'enquête vous le permettait. La taille des grappes peut alors être calculée en divisant 4.583 par 300, ou par environ 15-16 ménages.

Au lieu de vous fixer sur 300 UPS, vous avez décidé, vous et le personnel de l'enquête et de l'échantillonnage, que vous vouliez des grappes d'une taille, disons, de 10 pour pouvoir répondre à des besoins opérationnels comme la répartition du travail aux enquêteurs. Dans ce cas, vous devrez diviser 4.583 par 10 pour obtenir le nombre d'UPS qui sera d'environ 458. Vous reverrez alors ces chiffres en fonction du coût et d'autres considérations ; ensuite, soit vous accepterez cette taille de grappe, soit vous l'ajusterez. Vous pouvez conclure que 425 est le nombre minimum d'UPS que vous pouvez enquêter ; dans ce cas, vous ajusterez la taille de la grappe à 11 (autrement dit, 4.583/425).

EXEMPLE 2 :

Groupe cible :	Enfants de 12-23 mois
Pourcentage de la population :	2,5%
Indicateur-clé :	Couverture de la polio
Prévalence (couverture) :	26%
<i>Deff</i> :	Aucune information
Taille moyenne du ménage :	6

Pour ce scénario, utilisez toujours le Tableau 4.3 parce que, mis à part le taux de couverture de l'indicateur-clé, tous les paramètres du tableau sont applicables, étant donné que nous pouvons encore supposer que l'effet du plan de sondage est de 1,5, et le facteur d'ajustement des non-réponses est de 1,1. Pour la couverture, r , nous pouvons utiliser la colonne pour 25%, puisque la valeur estimée - 26% - en est très proche. La taille de l'échantillon pour une taille moyenne de ménage de 6 personnes sera de 9.167 ménages.

À cause des considérations de coûts et de charges de travail sur le terrain, supposez que l'équipe d'enquête décide qu'elle veut des grappes de 30 ménages, si possible. En divisant 9.167 par 30, vous obtenez 306 UPS ; vous devez décider ensuite si ce nombre est acceptable en termes de charge de travail sur le terrain. Si, d'un autre côté, vous avez décidé que vous vouliez avoir environ 400 UPS pour la répartition géographique et aussi pour avoir suffisamment d'UPS permettant de disposer d'estimations au niveau sous-national dans cinq régions, vous diviserez 9.167 par 400, ce qui vous donnera 23 comme taille de grappe. Rappelez-vous que plus la taille de la grappe est petite et plus les estimations de l'indicateur seront fiables (pour tous les indicateurs, et pas seulement pour l'indicateur-clé). Vous pouvez décider par conséquent d'utiliser les 400 UPS avec une taille moyenne de grappe de 23 ménages, tout en étant conscient du fait que ce schéma sera plus onéreux que celui des 306 UPS, à cause du coût des déplacements.

EXEMPLE 3 :

Groupe cible :	Enfants de 0-11 mois
Pourcentage de population :	3,5%
Indicateur-clé :	Nourrissons alimentés de manière adéquate
Prévalence (couverture) :	24%
<i>Deff</i> :	1.4 (d'une enquête précédente)
Taille moyenne du ménage :	4
Taux de non-réponse attendu :	10%

Dans ce scénario, vous devrez calculer la taille de l'échantillon en vous servant de la formule fournie dans cette section, puisque plusieurs des paramètres diffèrent de ceux utilisés ou postulés dans le Tableau 4.3. Parmi ceux-ci figurent les valeurs de p , f et du facteur d'ajustement des non-réponses, ce dernier étant basé sur un taux de non-réponse attendu de 5%, au lieu de 10%, tiré d'enquêtes similaires dans votre pays. L'application de la formule donne un chiffre de 10.303 ménages.

Supposez que le personnel de l'enquête ait conclu que l'enquête ne peut se charger que de 300 UPS au maximum, à cause de contraintes de coûts. Dans ce cas, vous obtiendrez la taille de la grappe en divisant 10.303 par 300, ce qui vous donnera une taille de grappe de 34 ménages. Vous devriez évaluer ici si une taille de grappe aussi grande donnera des estimations fiables pour les indicateurs autres que ceux qui sont clés⁸. Si nous supposons que la taille de grappe maximale ne devrait pas excéder 30 ménages, le nombre d'UPS qui seraient nécessaires pour 10.303 ménages est 343. Ainsi, il s'agit ici d'opérer un choix entre deux options : utiliser 300 UPS mais en acceptant une fiabilité moindre, ou alors utiliser 343 UPS pour un coût plus élevé.

DETERMINER QUEL ÉCHANTILLON UTILISER

Une fois que vous avez choisi la taille de l'échantillon et que vous avez déterminé le nombre de vos UPS, l'étape suivante consiste à prendre une décision sur le type d'échantillon à utiliser pour l'enquête. Concevoir, sélectionner et réaliser du début à la fin un bon échantillon probabiliste est un processus qui nécessite beaucoup de temps et qui est coûteux (l'échantillonnage probabiliste est présenté à la section suivante). Pour la MICS3, il est nécessaire d'estimer des indicateurs dans un court laps de temps et vous n'aurez peut-être pas le temps de concevoir un nouvel échantillon pour l'enquête. Par conséquent, il y a deux étapes importantes à suivre pour déterminer le type échantillon à utiliser pour l'enquête :

Étape 1 : Déterminer si un échantillon existant peut être utilisé pour l'enquête.

Étape 2 : Si aucun échantillon existant ne convient, il vous faudra concevoir un échantillon spécifique pour l'enquête MICS3.

⁸ Alors que l'effet du plan de sondage est très faible pour l'indicateur-clé de cet exemple et que, par conséquent, la fiabilité de cette estimation devrait normalement être en accord avec les exigences de précision, les autres indicateurs ayant une plus grande corrélation intra-grappe que ceux des enfants de moins d'un an risqueraient d'avoir des marges d'erreur plus importantes dans le cas d'une taille de grappe supérieure à 30, comparativement à, disons, 20 ou 25.

Cette section présente l'étape 1. Si un échantillon approprié pour la MICS3 est disponible, vous n'avez pas besoin de passer en revue les différentes options présentées dans la section suivante, pour concevoir un plan de sondage lors de l'étape 2. Il serait néanmoins utile de parcourir la section suivante, afin de vous assurer que l'échantillon disponible que vous comptez utiliser est un bon échantillon probabiliste, avec une base de sondage actualisée et de taille raisonnable.

UTILISER UN ECHANTILLON EXISTANT – OPTION 1

La plupart des pays ont des programmes d'enquêtes bien développés, gérés par leurs bureaux nationaux des statistiques ou par le ministère de la santé. Par conséquent, il peut être possible dans votre pays d'utiliser un échantillon déjà existant, un qui a été conçu pour d'autres objectifs. *C'est l'option recommandée pour votre enquête si l'échantillon existant est un échantillon probabiliste valable et s'il est disponible.* L'échantillon existant doit être évalué pour s'assurer qu'il répond aux critères d'un échantillon probabiliste (critères qui ont été discutés ci-dessus).

Il y a différentes manières d'utiliser un échantillon déjà existant :

- Joindre les modules MICS3 aux questionnaires qui vont être utilisés dans une autre enquête ;
- Utiliser l'échantillon ou un sous-ensemble d'une enquête précédente ;
- Utiliser les listes de ménages des zones de dénombrement (ou grappes) d'une autre enquête ;
- Utiliser les zones de dénombrement (ZD) ou grappes d'une enquête précédente avec une nouvelle liste de ménages.

Chacune de ces options présente des avantages et des inconvénients. Des considérations de temps sont aussi très importantes. Par exemple, il ne sera pas possible d'utiliser la première option s'il n'est pas prévu qu'une enquête se réalise au même moment qu'une enquête MICS. Cette option – attacher les modules MICS3 aux questionnaires d'une autre enquête – consiste en fait à collecter des données simultanément. *C'est intéressant dans la mesure où l'échantillon a déjà été sélectionné, ce qui évite les coûts d'échantillonnage pour l'enquête MICS3.* Cependant, le fait que les personnes enquêtées soient soumises, à la fois, aux questionnaires MICS3 qui sont assez longs et aux questionnaires de l'autre enquête, qui peuvent aussi être assez longs, constitue un inconvénient majeur de cette option. Ces aspects doivent être soigneusement évalués et discutés avec les bailleurs et l'équipe de pilotage de l'enquête-mère.

Le deuxième choix, qui consiste à utiliser l'échantillon d'une enquête précédente, présente aussi l'avantage de fournir un échantillon déjà mis en place, ce qui évite les coûts d'échantillonnage. Si la taille de l'échantillon de l'enquête précédente est trop grande, le statisticien devrait facilement pouvoir découper l'échantillon original en sous-échantillons pour que sa taille soit en accord avec les nécessités de l'enquête MICS3. Cependant, à l'opposé, si la taille de l'échantillon est trop petite, l'agrandir est plus problématique. Cette option comporte aussi un autre inconvénient, puisqu'il faut revisiter les mêmes ménages que ceux de l'enquête précédente,

cela pouvant être une source d'ennuis pour les personnes précédemment enquêtées et/ou pouvant créer aussi un conditionnement de ces personnes. Finalement, l'enquête précédente doit être très récente pour que cette option soit valable.

La troisième option, qui consiste à utiliser la liste des ménages des ZD de l'échantillon d'une précédente enquête pour sélectionner l'échantillon de la MICS3 présente un double avantage : celui d'avoir : (1) les unités primaires de sondage déjà tirées ; et, (2) la liste des ménages disponible. Par conséquent, ici aussi, la plus grande partie des opérations d'échantillonnage et des coûts ont déjà été pris en charge. Un autre avantage est dû au fait que, dans cette option, des ménages différents pourraient être sélectionnés pour l'enquête MICS3, éliminant ainsi le problème de deuxième visite et donc de fatigue et de conditionnement des ménages. Par contre, le fait que la liste des ménages puisse ne plus être valable si l'enquête précédente date de plus d'une année ou deux – ce qui rendrait ce choix non viable – en constitue un inconvénient. En fait, quand la liste des ménages est trop ancienne, il faut alors se résoudre à envisager la quatrième option figurant sur la liste ci-dessus. Pour utiliser cette option, il faut établir une nouvelle liste des ménages des ZD de l'échantillon. Si cette option présente l'inconvénient de devoir mettre à jour une nouvelle liste des ménages et d'en payer le coût, elle a cependant comme avantage que les unités primaires ont déjà été sélectionnées et que le plan de sondage lui-même est pratiquement en place sans qu'il y ait besoin d'y apporter des modifications supplémentaires.

Tableau 4.6**Option 1- Échantillon Existant*****Pour***

- Réduit le temps et les coûts
- Probablement conçu de manière appropriée avec des méthodes probabilistes
- Les ajustements pour s'adapter à une MICS3 peuvent être simples

Contre

- Nécessite une mise à jour s'il n'est plus à jour
- Les personnes enquêtées peuvent être saturées
- Le questionnaire peut être trop long si des modules additionnels sont ajoutés à des questionnaires de base
- Les ajustements pour s'adapter à la MICS3 peuvent être complexes

Chacun de ces points doit être soigneusement évalué et il est souhaitable, avant de prendre la décision d'utiliser un échantillon existant, de se déterminer sur la faisabilité des modifications nécessaires.

L'échantillon utilisé dans les Enquêtes Démographiques et de Santé (EDS)⁹ peut parfaitement être utilisé dans une enquête MICS. De nombreux pays ont réalisé récemment de telles enquêtes et d'autres s'apprêtent à le faire dans les mois à venir¹⁰. Les objectifs des EDS sont assez voisins de ceux des enquêtes MICS. Pour cette raison, le plan de sondage utilisé dans les EDS est vraisemblablement le plus approprié dans votre situation.

Dans quel cas est-il indiqué d'utiliser un échantillon de l'EDS ? Il sera nécessaire de déterminer s'il est disponible et à jour, et s'il convient à vos besoins. *Soit un échantillon d'une EDS très récente mais antérieure à 2003 pourra être utilisé pour réaliser l'enquête MICS3, soit une EDS à venir pourra être utilisée avec l'enquête MICS3 en supplément.* L'EDS sera nécessairement conçu comme un échantillon probabiliste. De ce fait, il vous suffit d'évaluer si (1) la taille de l'échantillon est suffisamment grande pour la MICS, et (2) le nombre d'UPS et les tailles des grappes sont situés dans les intervalles présentés dans ce manuel. En fin de compte, cela nécessitera, d'une part, l'autorisation de l'agence qui réalise ou qui finance l'EDS dans votre pays et, d'autre part, la coopération avec cette agence, sans oublier les contraintes dues à la saturation des personnes enquêtées, problème précédemment évoqué.

Il y a une autre enquête que de nombreux pays ont réalisée et dont l'échantillon peut être approprié : il s'agit des enquêtes sur l'emploi. Bien que les objectifs de telles enquêtes soient assez différents de ceux des enquêtes MICS3, les enquêtes sur l'emploi ont une conception assez proche des Enquêtes par Grappes à Indicateurs Multiples en termes de stratification, de taille d'échantillon et d'autres critères d'échantillonnage.

CONCEPTION D'UN PLAN DE SONDRAGE SPECIFIQUE POUR UN NOUVEL ECHANTILLON

Quand un échantillon existant ne peut être utilisé, il sera nécessaire de concevoir et/ou d'utiliser une nouvelle base de sondage des ménages, à partir de laquelle un nouvel échantillon sera sélectionné pour la MICS3. La construction de la base de sondage devra respecter rigoureusement les principes de l'échantillonnage probabiliste.

CARACTERISTIQUES D'UN BON TIRAGE PROBABILISTE ET D'UNE BONNE BASE DE SONDRAGE

Concevoir un bon échantillon aléatoire pour une enquête est tout aussi important que développer les différents modules des questionnaires, dans la mesure où un bon échantillon permettra de produire des résultats valables et comportant le moins de biais possible. Il y a plusieurs moyens de tirer un échantillon aléatoire, et chaque pays se caractérise par une situation particulière, des conditions et des besoins en données qui conditionnent le plan de sondage particulier qu'il

⁹ Les problèmes relatifs au sondage sont présentés dans "Demographic and Health Surveys: Sampling Manual, Basic Documentation - 8, Macro International Inc., Calverton, Maryland 1987".

¹⁰ Il y a lieu de signaler, néanmoins, qu'il n'est pas recommandé de mener une MICS3 si une EDS a été réalisée depuis 2003, ou si elle va être réalisée en 2005 ou en début 2006.

adopte. Cependant, il y a un certain nombre de conditions qui doivent être respectées par tous les pays qui veulent disposer d'un échantillon probabiliste scientifique.

- Utiliser des méthodes reconnues de sondage aléatoire à tous les stades du tirage de l'échantillon ;
- Sélectionner un échantillon représentatif sur le plan national ;
- S'assurer que la mise en œuvre de l'échantillon est conforme au plan de sondage ;
- S'assurer que la taille de l'échantillon est suffisante pour réaliser, de manière fiable, les objectifs fixés.

Outre ces quatre conditions, il y a aussi d'autres procédures nécessaires qu'il faut suivre pour tirer un échantillon et que nous vous recommandons fortement de suivre, bien que certaines puissent être modifiées en fonction des pays et de leurs besoins. Il s'agit, entre autres :

- Des procédures simples d'échantillonnage, opposées à des procédures complexes,
- L'utilisation du recensement de population le plus récent comme base de sondage,
- Un échantillon auto-pondéré si possible.

Des méthodes de sondage scientifiquement éprouvées ont été utilisées lors d'enquêtes dans la plupart des pays depuis des décennies. Si un échantillon ne représente pas exactement la population parce que des méthodes techniques ayant fait leur preuve n'ont pas été utilisées, les estimations de l'enquête seront biaisées. De plus, l'importance de ces biais sera inconnue. Il est donc très important de s'assurer que la méthodologie d'échantillonnage se base sur des techniques de sélection probabiliste à chaque stade du processus de sélection.

Pour éviter les biais, vous devez procéder à un tirage probabiliste pour sélectionner votre échantillon de personnes à enquêter. Les biais dépendent de la procédure de sélection et non de la taille de l'échantillon. Augmenter la taille de l'échantillon ne contribuera pas à éliminer les biais si le choix des techniques est mauvais.

Un tirage probabiliste garantit que tous les individus de la population-cible¹¹ ont une chance connue d'être sélectionné dans l'échantillon. De plus, cette chance doit être non-nulle et calculable. Le fait pour le statisticien en charge de l'échantillonnage de ne pas pouvoir calculer les probabilités de sélection de l'échantillon aura certainement pour conséquence un échantillon qui ne sera pas probabiliste.

Certaines méthodes de sondage ne sont pas basées sur des probabilités techniques : il s'agit par exemple, des échantillons au jugé, des échantillons choisis à dessein et des échantillons par

¹¹ La MICS3 a différentes populations-cibles, qui sont fonction de l'indicateur. Par exemple, les enfants de 0-11 mois, les enfants de 12-23 mois, les enfants de moins de 5 ans, les enfants de moins de 5 ans avec diarrhée, les femmes de 15-49 ans et la population totale.

quotas. La méthode du trajet aléatoire de sélection des enfants est une procédure simple par quotas. Il est important que vous n'utilisiez pas de telles procédures pour la MICS3.

Etre exigeant sur le choix de la procédure d'échantillonnage probabiliste est le meilleur moyen de limiter les biais d'échantillonnage. Il existe d'autres biais qui ne sont pas imputables à des problèmes d'échantillonnage, comme les non réponses, les réponses erronées ainsi que les erreurs commises par les enquêteurs ; ces erreurs se produisent partout à des degrés divers quelle que soit la méthode d'échantillonnage utilisée.

Dans le sondage probabiliste, chaque individu dans la population-cible a une chance d'être sélectionné, la chance d'être sélectionnée n'est pas nulle, et elle est mathématiquement calculable, et les techniques probabilistes sont utilisées à *chaque stade de sélection*.

Des mesures appropriées telles que la réalisation de pré-tests, l'organisation de bonnes sessions de formation des enquêteurs et l'instauration d'un contrôle de qualité du travail sur le terrain doivent être prises pour limiter ces biais non imputables à des erreurs d'échantillonnage.

Le deuxième critère du plan de sondage de la MICS3 est que l'échantillon doit avoir une couverture nationale. Cela est nécessaire parce que les indicateurs calculés doivent refléter la situation du pays dans son ensemble. Dans la mesure du possible, il est important d'inclure les groupes difficiles à recenser pour garantir une couverture nationale complète. Il peut s'agir de nomades, de sans-abri ou de personnes en situation provisoire ou encore de camps de réfugiés, de camps militaires ou de campements situés dans des zones isolées qui sont, pour une raison ou une autre, difficiles d'accès. Il est probable que l'état de santé des enfants notamment, vivant dans de telles situations, est différent de celui des enfants qui vivent dans un environnement plus stable ou plus traditionnel, et les exclure aurait pour conséquence des indicateurs biaisés.

L'une des plus importantes manières de garantir que l'échantillon a une couverture nationale et, donc, est conforme aux principes d'un échantillonnage probabiliste, c'est de s'assurer que la base de sondage couvre l'ensemble de la population du pays. Le plan de sondage est présenté plus en détail ci-dessous.

Pour qu'un échantillonnage probabiliste soit efficace, il est essentiel que la mise en œuvre du plan de sondage sur le terrain, y compris les procédures d'enquêtes, soit aussi fidèle que possible au plan de sondage. Dans de nombreux cas, un travail de terrain de mauvaise qualité a ruiné un plan de sondage parfaitement acceptable. Les superviseurs de terrain doivent veiller à ce que les procédures de sélection soient strictement suivies.

Pour qu'un échantillon probabiliste soit valable, le calcul de la taille de l'échantillon doit tenir compte des niveaux de précision nécessaires. Cette question a déjà été présentée dans la section précédente relative à la détermination de la taille de l'échantillon. Il a été recommandé que le niveau de précision de l'indicateur-clé soit fixé à une marge d'erreur relative de 12% à un niveau de confiance de 95%, et c'est sur la base de ces critères que la formule de calcul de la taille de l'échantillon se base. Si votre indicateur-clé a, par exemple, une couverture ou une prévalence de

20%, alors l'erreur relative de 12% correspondra à une marge d'erreur de 2,4 points de pourcentage, et l'intervalle de confiance pour votre estimation de 20% devrait être | 17,6 – 22,4 |.

Votre plan de sondage doit être aussi simple que possible. Il est bien connu que plus le plan de sondage sera complexe, plus sa mise en œuvre posera des problèmes. En particulier, les problèmes peuvent se poser au niveau du terrain si des procédures d'échantillonnage compliquées doivent être utilisées. De plus, la production des résultats de l'enquête en temps voulu pourrait ne pas se faire.

Un plan de sondage est dit *auto-pondéré* si chaque individu de la population-cible a une même probabilité d'ensemble d'être sélectionné. La probabilité d'ensemble est le produit des probabilités à chaque degré de sélection. Un sondage auto-pondéré est souhaitable parce que les différentes estimations – par exemple, les répartitions en pourcentage – pourront être calculées directement à partir des valeurs de l'échantillon sans qu'il soit nécessaire de les pondérer ou de les extrapoler. Si l'on tient à ce que le plan de sondage reste simple, il est préférable d'avoir un échantillon auto-pondéré plutôt qu'un échantillon plus compliqué, non pondéré. Cela dit, *l'auto-pondération ne doit pas être considérée comme un critère strict*, parce que pondérer les résultats de l'échantillon pour préparer les estimations des indicateurs peut facilement être effectué par les ordinateurs actuels. Il y a néanmoins des situations dans lesquelles l'échantillon ne peut être auto-pondéré.

EXEMPLE :

Supposer que dans votre pays, vous vouliez estimer des indicateurs séparés pour les milieux urbain et rural, et supposer aussi qu'ultérieurement, vous vouliez des estimations également fiables pour les deux milieux. Pour cela, il vous faudra sélectionner un échantillon de taille égale pour chacun des deux milieux - urbain et rural. A moins que les populations urbaines et rurales soient de même taille, le taux de sondage dans chaque milieu sera différent. Aussi, l'échantillon national d'ensemble devra t-il être pondéré pour corriger les résultats et, par conséquent, l'échantillon de l'enquête ne sera donc pas auto-pondéré.

UTILISATION DU RECENSEMENT COMME BASE DE SONDAGE ET QUAND FAIRE UNE MISE A JOUR

Il est fortement recommandé d'utiliser comme base de sondage le recensement de la population le plus récent, mis à jour si nécessaire. À l'heure actuelle, pratiquement tous les pays ont effectué un recensement récent de la population, c'est-à-dire un recensement réalisé au cours des dix dernières années. La base est essentiellement constituée d'un ensemble de matériaux à partir desquels l'échantillon de l'enquête est sélectionné. Une base de sondage parfaite est une base de sondage complète, précise et mise à jour, et étant donné qu'une base de sondage n'est jamais parfaite à 100%, le recensement de la population est ce qui s'en rapproche le plus dans la plupart des pays. *En ce qui concerne nos enquêtes, l'utilisation première d'un recensement est de fournir une liste complète des zones de dénombrement (ZD) y compris une mesure de la taille, comme le nombre d'habitants ou de ménages pour la sélection des unités de sondage au premier degré.* Dans la plupart des pays, les cartes font généralement parties du recensement de la population et

elles peuvent comprendre des croquis des zones de dénombrement. Les cartes sont très utiles parce que la taille des ZD sélectionnées devra probablement être mise à jour du point de vue du nombre de ménages qui y résident *actuellement*, en particulier, si le recensement date de plus d'une année ou deux.

Dans le cadre de la série des recensements de l'an 2000, certains pays ont réalisé le leur plus tôt – en 1999, alors que plusieurs autres ont effectué le leur au cours de la période 2000-2002. Cela soulève la question cruciale de savoir si la base de sondage du recensement devrait ou non être mise à jour pour la MICS. *En général, il est recommandé que la mise à jour ne soit pas effectuée si la base de sondage du recensement date de 2003 ou plus tard, avec une seule exception.* Dans les pays ayant enregistré d'importants déplacements de population depuis 2003, surtout dans les zones fortement urbanisées qui se sont fortement étendues dans des zones spécifiques du fait d'une croissance exponentielle de nouvelles constructions de zones résidentielles. Une opération de mise à jour de ces zones devrait être opérée. Toutefois, vous pouvez décider que cela ne soit pas nécessaire si votre recensement de population est récent, et précède votre enquête MICS3 de 12 mois ou moins.

Si la base de sondage de votre recensement national est antérieure à 2003, alors sa mise à jour s'impose.

La raison de la mise à jour semble évidente. Il est nécessaire de garantir que la couverture de l'ensemble de la population est aussi précise et exhaustive que possible. Les étapes recommandées pour la mise à jour de la base de sondage du recensement sont similaires quel que soit le scénario, autrement dit, que ce soit dans le cas d'une forte croissance urbaine depuis 2003 ou dans le cas d'une mise à jour globale d'une vieille base de sondage de recensement datant d'avant 2003. La différence se situe au niveau de l'échelle et de l'ampleur de l'opération de mise à jour. La mise à jour d'une vieille base de sondage de recensement – antérieure à 2003 – demande beaucoup plus de travail et est plus coûteuse que la mise à jour de bases de sondage récentes. Néanmoins, dans tous les deux cas, l'opération devrait couvrir l'ensemble de la base de sondage, et non pas seulement les zones de dénombrement – UPS – qui auront à être sélectionnées lors de l'échantillonnage. En fait, ce sont les informations recueillies lors de la mise à jour qui seront utilisées pour sélectionner l'échantillon.

Il est important de prendre conscience que la mise à jour d'une base de sondage est une opération purement statistique. S'il s'avère nécessaire de procéder à une mise à jour, cela devrait être pris en compte dans l'algorithme de coûts lors de la préparation de votre budget. De plus, si la mise à jour est incontournable, il est impératif que vous sollicitiez les services de la Direction Nationale de la Statistique de votre pays. Les étapes spécifiques à suivre sont les suivantes :

- Identifier les zones, plus particulièrement dans les grandes villes, ayant connue une forte croissance de la construction résidentielle depuis le recensement, indépendamment du fait que le recensement de la population ait eu lieu avant ou après 2003.
- Identifier les nouvelles zones, telles que les bidonvilles qui ont connu une forte croissance de la population depuis le recensement. Celles-ci incluent les zones qui étaient "vides" ou peu peuplées au moment du recensement.
- Ignorer les zones résidentielles vieilles et stables ayant enregistrées peu de changements dans le temps.
- Comparer les zones identifiées aux étapes 1 et 2 avec leurs zones de dénombrement du recensement, en tenant compte des limites qui se superposent.
- Dans les zones de dénombrement affectées, effectuer un sondage léger et un décompte rapide des logements. Notez que le décompte rapide suppose seulement un comptage brut des habitations, sans pour autant procéder au dénombrement de leurs occupants. Le

décompte rapide n'occasionnerait pas de devoir frapper aux portes pour parler aux enquêtés, excepté dans les cas d'immeubles d'habitation où il est difficile, à partir de la rue, de connaître le nombre d'appartements qui y sont.

Utilisez le nouveau décompte rapide de logements¹² en remplacement du décompte original de ménages figurant dans la base de sondage du recensement. Ce sera sa nouvelle 'mesure de la taille', une valeur nécessaire à l'établissement des probabilités de sélection de l'échantillon des zones de dénombrement.

Il est indéniable que la mise à jour de la base de sondage comme préalable à la sélection de l'échantillon n'est pas une opération triviale, mais plutôt une opération demandant beaucoup de temps et des ressources financières. C'est là l'une des raisons pour lesquelles il est recommandé, à chaque fois que cela est possible, d'utiliser un échantillon déjà existant.

UTILISATION D'UN NOUVEL ÉCHANTILLON ET CHOIX D'UNE OPTION POUR SA CONCEPTION

Quand un échantillon existant adapté n'est pas disponible pour être utilisé pour réaliser une MICS3, soit pour une enquête indépendante, soit comme supplément à une autre enquête, un nouvel échantillon devra être conçu et sélectionné, en commençant par l'élaboration de la base de sondage (tel que présenté précédemment dans ce chapitre).

Cette section porte sur les recommandations concernant les principales caractéristiques qu'un échantillon d'une enquête MICS3 doit posséder. Ci-dessous, deux options sont présentées mais, en premier lieu, les caractères généraux sont résumés.

De manière générale, l'échantillon de votre enquête doit être un échantillon probabiliste à tous les stades de la sélection ; il doit avoir une couverture nationale et il doit être conçu de la manière la plus simple possible pour que le terrain puisse se dérouler facilement avec un minimum de déviation par rapport à la conception. En gardant comme but principal la simplicité, la stratification et le nombre de stades de sélection doivent être minimaux.

En ce qui concerne la stratification, son objectif premier consiste à augmenter la précision des estimations de l'enquête ; en outre, elle permet de suréchantillonner des zones à un niveau sous-national, quand ces zones présentent un intérêt particulier. Une stratification implicite est un type de stratification très efficace et simple à mettre en œuvre quand des estimations au niveau national constituent l'objectif principal. C'est une forme de stratification géographique qui, quand elle est utilisée ensemble avec un échantillonnage systématique avec *ppt*¹³ (voir

¹² Il est communément admis que le nombre de logements pourrait ne pas être égal au nombre de ménages. Toutefois, ce qui importe ici c'est d'obtenir une estimation grossière afin de fixer la mesure de la taille. Par exemple, si lors du décompte rapide, ce sont 120 logements qui ont été identifiés dans une zone de dénombrement sélectionnée pour l'échantillon, et qu'ultérieurement, on se rend compte que ce sont 132 ménages qui occupent ces logements, ni la validité ni la fiabilité des résultats de l'échantillon n'en seront affectées de manière significative.

¹³ C'est une probabilité proportionnelle à la taille (*ppt*) et qui se réfère à une technique de sélection des zones de l'échantillon proportionnelle à leur taille en population ; par conséquent, une zone contenant 600 personnes aura une probabilité d'être sélectionnée deux fois supérieure à une zone ne contenant que 300 personnes.

explications vers la fin de ce chapitre) répartit automatiquement de manière proportionnelle l'échantillon dans chaque subdivision administrative du pays, comme dans les secteurs urbain et rural. Une stratification implicite est réalisée en suivant un ordre géographique en serpent, en séparant l'urbain du rural, avant d'appliquer un sondage systématique avec *ppt*.

De plus, la conception doit être celle d'un échantillon à trois degrés. Les unités de sondage au premier degré, ou unités primaires de sondage (UPS) doivent être définies, si possible, comme des zones de dénombrement (ZD) et elles doivent être sélectionnées avec une probabilité proportionnelle à la taille (*ppt*). La ZD est recommandée parce que l'UPS doit être une zone dans laquelle le terrain doit pouvoir être organisé de manière pratique ; elle doit être suffisamment petite pour permettre la cartographie, la segmentation ou l'énumération des ménages, mais elle doit être suffisamment grande pour être facilement repérable sur le terrain.

Le second degré consiste en la sélection des segments (grappes), et le troisième comprend la sélection des ménages à l'intérieur du segment qui seront retenus pour l'interview lors de l'enquête. Ces ménages peuvent être sélectionnés de plusieurs manières – à partir de sous-échantillons de listes de ménages existantes dans chaque segment ou à partir de listes nouvelles.

Cette conception laisse une certaine flexibilité, cela dépendant des conditions et des besoins du pays. La conception va probablement beaucoup varier d'un pays à un autre concernant le nombre d'UPS, le nombre de segments ou de grappes par UPS, le nombre de ménages par segment et, donc, la taille d'ensemble de l'échantillon.

Comme règle générale :

- Le nombre d'UPS doit se situer entre 250 et 350 ;
- La taille des grappes (c'est-à-dire le nombre de ménages à interviewer dans chaque segment) doit se situer entre 10 et 30, cela dépendant de l'option retenue parmi les deux options décrites ci-dessous ;
- La taille de l'échantillon d'ensemble doit être de l'ordre de 2.500 à 10.500 ménages.

Un pays peut décider, pour ses propres objectifs, qu'il a besoin d'estimer des indicateurs pour certaines sous-régions, en plus du niveau national. Dans ce cas, il faudra obligatoirement inclure dans le plan de sondage un niveau différent de stratification et un nombre plus important d'UPS pour garantir une représentation géographique adéquate des zones de l'échantillon dans chaque sous-région. De plus, la taille de l'échantillon devra être augmentée de manière sensible pour pouvoir fournir des estimations fiables pour les sous-régions ou pour d'autres domaines sous-nationaux (présenté ci-après de manière plus détaillée dans ce chapitre).

METHODOLOGIE DES SEGMENTS DE MEME TAILLE, FIXEE A L'AVANCE - OPTION 2

Nous avons déjà mentionné le fait que les enquêtes EDS fournissent un échantillon existant approprié pour la MICS3 (souvenez-vous que nous parlons de l'utilisation d'un échantillon existant comme l'Option 1). Le plan de sondage standard des EDS est, en fait, un bon modèle

pour la MICS3, si vous décidez qu'un nouvel échantillon est nécessaire. Le modèle d'échantillon EDS a aussi été utilisé dans d'autres enquêtes qui collectent des données sur la santé comme les enquêtes du programme PAPCHILD dans les pays arabes¹⁴.

Les modèles d'échantillons des enquêtes EDS et PAPCHILD sont basés sur ce qui est communément appelé un *plan de sondage utilisant des segments de même taille fixée à l'avance*, qui possède les qualités de la méthodologie probabiliste, la simplicité et de plus est proche des objectifs de la MICS3, à la fois sur le plan substantif et statistique. Dans les manuels de sondage des enquêtes EDS et PAPCHILD, on peut lire que la plupart des pays disposent de bases de sondage pratiques qui se présentent sous la forme de liste de ZD établies à partir du recensement de population le plus récent. Des croquis sont généralement disponibles pour les ZD, ainsi que des décomptes de la population et/ou des ménages. Les zones de dénombrement du recensement sont habituellement assez uniformes en ce qui concerne la taille. De nombreux pays ne disposent pas de listes satisfaisantes de ménages, ni d'un système d'adresses adéquat, en particulier dans les zones rurales. Par conséquent, il est nécessaire de préparer de nouvelles listes de ménages pour mettre la base de sondage à jour.

Pour pouvoir utiliser ce type de sondage (utilisant des segments de même taille fixée à l'avance) avec la MICS, il faut en premier lieu organiser la base de sondage des ZD fournie par le recensement selon un ordre géographique pour obtenir une stratification implicite. Certaines ZD sont si grandes que, si elles sont sélectionnées, il serait trop coûteux d'établir une nouvelle liste de tous les ménages. Il est plus efficace d'utiliser des segments. Cela consiste à attribuer à chaque zone de dénombrement une taille égale au nombre désiré de "segments-types" qu'elle contient. Dans les manuels de sondage de l'EDS et de PAPCHILD, il est recommandé que le nombre de segments-types soit défini (et calculé) en divisant la population de la ZD d'après le recensement par 500, en arrondissant au nombre entier le plus proche. Notez que dans les cas où vous procédez à une mise à jour de la base de sondage du recensement, c'est le *décompte des habitations (multiplié par 5)* que vous obtenez à la dernière étape de l'opération de mise à jour (décrite dans la section précédente sur les bases de sondage) qui devrait être utilisé, plutôt que l'effectif de population issu du recensement. Le facteur de multiplication 5 est nécessaire pour approximer l'effectif de population actuel dans les ZD mises à jour, de sorte que sa taille soit définie de la même manière que celle des ZD qui ne sont pas mises à jour. *Si vous décidez d'utiliser l'Option 2, c'est cette taille de segment-type qui est recommandée pour la MICS3.*

La prochaine étape consiste à sélectionner les ZD avec tirage proportionnel à la taille. Notez que la taille est aussi le nombre de segments. Dans de nombreux cas, vous pouvez trouver que la taille moyenne d'une ZD est d'environ 500 (équivalent à 100 ménages quand la taille moyenne d'un ménage est égale à 5) ; par conséquent, la mesure de taille typique sera 1.

La segmentation, utilisant les cartes disponibles, constitue la phase suivante. *Quand le nombre de segments dans une ZD est égal à 1, aucune segmentation n'est nécessaire, parce que le segment*

¹⁴ Voir The Arab Maternal and Child Health Survey, Basic Documentation 5: Sampling Manual, Ligue Arabe, Le Caire, 1990.

et la ZD sont un et identiques. Si le nombre de segments est supérieur à 1, alors la segmentation sera nécessaire. Cela a pour conséquence une subdivision des ZD en parts (égales au nombre de segments), avec chaque part contenant à peu près le même nombre de ménages. La segmentation peut être faite au bureau si les cartes sont suffisamment exactes. Si ce n'est pas le cas, une visite sur le terrain est nécessaire, en particulier dans les cas où des limites internes identifiables à l'intérieur de la ZD ne sont pas clairement délimitées (voir Chapitre 6 pour les détails concernant la cartographie et la segmentation).

Tableau 4.7

Option 2 – Résumé de la méthodologie des segments de même taille fixée à l'avance

Caractéristiques

- Sondage à trois degrés avec stratification implicite
- Sélection des ZD avec *ppt*
- Cartographie et segmentation des ZD ayant plus d'un segment-type
- Sélection, de façon aléatoire, d'un segment dans chaque ZD
- Enumération des ménages dans les segments sélectionnés
- Sélection systématique des ménages échantillon dans les segments

Paramètres

- Généralement, échantillon de 250 à 400 ZD (UPS)
- Segments-type de 500 habitants (environ 100 ménages)
- Grappes non-compactes de 10 à 35 ménages (différent de l'Option 3 ci-dessous)
- Taille d'échantillon de, généralement, 4.000 à 14.000 ménages*

** Notez qu'en général, nous ne recommanderions pas le nombre minimum d'UPS multiplié par la taille de grappe minimum (250 multiplié par 10), car 2.500 est vraisemblablement une trop petite taille d'échantillon pour, dans la plupart des pays, mesurer de manière fiable d'importants indicateurs.*

Après la segmentation, un segment est sélectionné au hasard dans chaque ZD. Dans chaque segment sélectionné, *on établit une nouvelle liste des ménages*. Rappelons qu'elle devrait comprendre environ 100 ménages. Ensuite, à partir des listes, un échantillon de ménages à enquêter est sélectionné systématiquement utilisant un intervalle de tirage fixe.

Tableau 4.8**Option 2 – Segments de même taille fixée à l'avance****Pour**

- Échantillon probabiliste
- Cartographie et segmentation minimales
- L'énumération est limitée
- Un peu plus fiable que l'Option 3 (ci-dessous)
- Corrige en partie l'ancienne base de sondage
- Plan auto-pondéré

Contre

- Une énumération, bien que minimale, est nécessaire dans chaque segment
- Peut donner des tailles de segments très variables, surtout si la base est ancienne et n'est pas mise à jour

EXEMPLE :

On peut avoir décidé de sélectionner un cinquième des ménages nouvellement listés dans chaque segment. Par conséquent, s'il y a, disons, 300 segments, alors le nombre de ménages sélectionnés dans chaque segment sera approximativement de 20 (bien qu'il puisse varier par UPS) et la taille de l'échantillon d'ensemble sera, approximativement, de 6.000 ménages.

La méthode des segments de même taille est pratique. Dans un pays typique, c'est-à-dire un pays dans lequel les ZD comptent, en moyenne, environ 100 ménages, il y aura très peu de segmentation à effectuer. De plus, la quantité de ménages à lister sera aussi limitée.

Les ménages sélectionnés selon l'Option 2 proviennent de grappes non-compactes¹⁵ et l'échantillon est auto-pondéré. Le nombre de ménages sélectionnés dans chaque UPS variera quelque peu parce que les UPS sont sélectionnées sur la base de leur taille au recensement (hormis celles qui ont été mises à jour), taille qui sera sans aucun doute différente des tailles actuelles quand la nouvelle liste des ménages est établie.

EXEMPLE :

Supposez que l'on a calculé que le taux de sélection à l'intérieur du segment est de 1 ménage listé sur 5. Si un segment est sélectionné sur la base d'une taille de 98 ménages selon le recensement, mais si selon l'énumération il y a maintenant 112 ménages, alors un ménage sélectionné sur cinq donnera 22 ou 23 ménages (le nombre correct) au lieu des 19 ou 20 prévus. La procédure reflète non seulement de manière correcte l'augmentation de la population, mais aussi elle maintient le caractère auto-pondéré de l'échantillon. La déviation dans la taille moyenne du segment ne doit pas être importante, à moins que la base de sondage ne soit ancienne¹⁶.

¹⁵ Une grappe non compacte est une grappe dans laquelle les ménages sélectionnés sont éparpillés systématiquement à travers toute la zone sélectionnée. Une grappe compacte est une grappe dans laquelle chaque ménage dans un segment donné est contigu au ménage voisin. Les grappes non compactes donnent des résultats plus fiables que les grappes compactes à cause de leur effet de sondage plus petit

¹⁶ Une autre procédure est possible quand on pense que la population a changé de manière significative et que la

METHODOLOGIE MODIFIEE DES SEGMENTS DE MEME TAILLE, FIXEE A L'AVANCE - OPTION 3

Dans le cas où un échantillon bien conçu est disponible, nous avons montré que l'utilisation de cet échantillon existant est l'option souhaitable pour la MICS3. Dans le cas où votre pays doit concevoir l'échantillon à partir de rien, nous avons aussi montré que l'utilisation des plans de sondage des enquêtes EDS et PAPCHILD, la méthodologie des segments de même taille fixée à l'avance, est la seconde meilleure option.

L'option 3 consiste en la modification de la méthodologie des segments. La *méthodologie modifiée des segments* est similaire à la méthodologie des segments-types, mais il y a néanmoins des différences importantes¹⁷. Plutôt que de créer des segments-types dont la taille est de 500 individus dans chaque ZD, on subdivise la ZD en un nombre de segments prédéterminé. Ce nombre prédéterminé est égal au nombre de ménages du recensement (ou du comptage des logements actuels) dans la ZD divisé par la taille souhaitée de la grappe, arrondi à l'entier près. Vous noterez qu'ici, on utilise les ménages (ou les habitations pour les zones actualisées), plutôt que la population, qui a été utilisée pour l'Option 2. Ainsi, il n'est pas nécessaire de multiplier le nombre d'habitations des zones actualisées par 5.

EXEMPLE :

Si la taille désirée de la grappe est de 20 ménages et s'il y a 155 ménages dans la ZD, alors il faudra créer 8 segments.

De même qu'avec l'Option 2, les ZD sont sélectionnées avec une probabilité proportionnelle au nombre de segments qu'elles contiennent. Chaque ZD sélectionnée est alors segmentée en un nombre prédéterminé de segments en utilisant des cartes, de même qu'un rapide comptage des logements actuels. Les limites soigneusement tracées doivent être identifiées dans la segmentation et le nombre de logements dans chaque segment doit être à peu près égal, bien qu'il ne soit pas nécessaire qu'il soit exact. Notez que ce comptage rapide peut être basé sur les logements plutôt que sur les ménages, ainsi que cela a été fait pour la mise à jour de la base (voir cette section pour de plus amples détails).

Après la segmentation, un (et seulement un) segment est sélectionné au hasard à l'intérieur de chaque ZD. *Tous les ménages dans les limites du segment sélectionné sont alors interviewés pour l'enquête*, le segment formant donc une grappe compacte de ménages.

taille moyenne des segments peut être trop variable pour permettre un travail sur le terrain efficace. On peut fixer la taille du segment plutôt que la fraction de ménages à sélectionner ; dans ce cas, un intervalle de tirage différent devra être calculé et appliqué à chaque segment sélectionné. Chaque segment aura alors un poids différent et ceci devra être pris en compte dans le calcul des indicateurs.

¹⁷ Pour une description complète de la méthodologie modifiée des segments (ou grappes), voir Turner A., Magnani, R. et Shuaib, M. (1996) – “A Not Quite as Quick but Much Cleaner Alternative to the Expanded Programme on Immunization (EPI) Cluster Survey Design”. *International Journal of Epidemiology*, Vol. 25, No.1.

Les autres caractéristiques de la méthode des segments modifiés sont, pour l'essentiel, les mêmes que celles de la méthode des segments de même taille – échantillonnage à trois degrés, stratification implicite, sélection des ZD avec *ppt*.

La méthodologie modifiée du segment présente un avantage sur celle du segment de même taille : en effet, il n'est pas nécessaire d'établir des listings de ménages, ce qui réduit le coût de l'enquête. Cependant, le rapide comptage et le tracé des cartes constituent des dépenses supplémentaires mais le coût du comptage rapide est peu élevé puisqu'il consiste en une inspection visuelle au lieu de devoir frapper aux portes pour parler aux enquêtés. De plus, la procédure permet de compenser l'utilisation d'une base de sondage éventuellement trop ancienne, en interviewant tous les ménages actuels dans un segment sélectionné, quel que soit le nombre de ménages qui se trouvaient là au moment du recensement.

Tableau 4.9

Option 3 – Résumé de la méthodologie modifiée des segments de même taille fixée à l'avance

Caractéristiques

- Sondage à trois degrés avec stratification implicite
- Prédétermination du nombre de segments par UPS
- Sélection des ZD avec *ppt*
- Cartographie et segmentation dans toutes les ZD
- Sélection, de façon aléatoire, d'un segment dans chaque ZD
- Interview de tous les ménages dans le segment sélectionné

Paramètres

- Généralement, échantillon de 250 à 400 ZD (UPS)
- Grappes compactes de 20 à 30 ménages (taille minimum 20)
- Taille de l'échantillon de, généralement, 5 000 à 12.000 ménages*
- La *taille du segment* et la *taille de la grappe* sont synonymes (*différent de l'Option 2*)

**Notez que l'intervalle des tailles d'échantillons est différent de ceux de l'Option 2 dans le Tableau 4.7, parce que les tailles de grappes compactes recommandées sont différentes.*

Un inconvénient de cette méthodologie est dû au fait que les segments (les grappes) sont compacts. Par conséquent, pour une même taille d'échantillon, la fiabilité de l'échantillon selon ce plan sera quelque peu moindre que celle d'un plan conçu selon la méthodologie des segments de même taille où les grappes ne sont pas compactes. Cependant, cela peut être compensé en échantillonnant davantage de ZD avec un échantillon plus petit prélevé à l'intérieur des ZD. Il y a un autre inconvénient qui est dû au fait que la segmentation elle-même nécessite comparativement de petits segments pour être délimitée, ce qui dans certains pays peut ne pas se révéler pratique. Cela peut être problématique dans de petites zones où il n'y a pas suffisamment de limites naturelles comme des routes, des passages, des fleuves, etc., pour que la segmentation soit précise ou même correcte. *Pour cette raison, il est recommandé que la taille du segment dans cette option soit d'au moins 20 ménages ; pour compenser la perte de fiabilité avec le*

segment compact, il ne devrait pas être plus grand que 30 ménages. La délimitation est extrêmement importante quand on forme les segments, en termes de contrôle de biais d'échantillonnage.

PLAN DE SONDAGE SIMPLIFIE – NON RECOMMANDE

La première série de MICS, en 1995, s'est particulièrement intéressée à la méthode du trajet aléatoire, qui est utilisée dans le Programme Élargi de Vaccination (PEV). Le fait que la sélection des ménages n'est pas basée sur des méthodes d'échantillonnage probabiliste, mais plutôt sur une procédure fournissant un échantillonnage par quota, constitue l'objection principale à l'utilisation de la méthode du trajet aléatoire pour la MICS3.

Comme la MICS3 a des tailles d'échantillon relativement grandes, la méthode du trajet aléatoire est inadaptée. On prétend parfois que les enquêtes PEV à petite échelle, avec leur faible taille d'échantillon, sont plus affectées par la variance d'échantillonnage que par des biais, ce qui par conséquent justifie en partie

Des procédures simplifiées – telles que celle du trajet aléatoire – qui dévient d'une conception probabiliste ne sont pas recommandées pour la MICS3 et ne doivent pas être utilisées.

l'utilisation de la méthode du trajet aléatoire. Cependant, pour la MICS3, ce même argument conduit à la conclusion inverse – à savoir que le biais est un problème beaucoup plus important que la variance d'échantillonnage, du fait des tailles d'échantillon beaucoup plus grandes – et, par conséquent, des méthodologies d'échantillonnage probabiliste plus strictes doivent être utilisées à tous les degrés de sélection.

Tableau 4.10

Résumé de la liste de contrôle pour la taille de l'échantillon et le plan

- Déterminer le groupe-cible qui représente la plus petite proportion de la population
- Déterminer le niveau de couverture pour le même groupe-cible
- Si la situation de votre pays cadre avec les hypothèses et les conditions du Tableau 4.3, choisissez dans ce dernier la taille d'échantillon qui correspond ; OU BIEN, calculer la taille d'échantillon avec la formule donnée dans ce chapitre
- Décider de la taille de la grappe, généralement de l'ordre de 10 à 35 ménages
- Diviser la taille de l'échantillon par la taille de la grappe pour obtenir le nombre d'UPS (zones d'échantillonnage).
- Revoyez vos choix de n, taille de la grappe, et le nombre d'UPS, DANS LE BUT DE choisir parmi les Options 1, 2 ou 3 pour le plan de sondage.

POINTS IMPORTANTS POUR L'ÉCHANTILLON DE LA MICS3

Cette section présente quelques-uns des autres points importants qui devraient être pris en compte lors de la planification des questions d'échantillonnage de la MICS3 dans votre pays. Ces points concernent les estimations au niveau sous-national, les analyses des tendances des sous-groupes ainsi que les estimations relatives à l'eau et à l'assainissement.

ESTIMATIONS AU NIVEAU SOUS-NATIONAL

Jusqu'à présent, on s'est intéressé aux tailles d'échantillon nécessaires pour estimer des indicateurs au niveau national. Cependant, de nombreux pays voudront également utiliser la MICS3 pour calculer des indicateurs à un niveau sous-national – par exemple, au niveau urbain-rural, pour des régions, des états ou des provinces ou éventuellement des districts. De telles données peuvent être utilisées pour identifier les zones dans lesquelles il est nécessaire de déployer des efforts importants. De telles données peuvent également être utilisées à des fins de programmation et d'évaluation.

La taille de l'échantillon est un facteur essentiel limitant les possibilités d'estimer des indicateurs fiables pour des sous-populations. Pour que les résultats aient un degré de précision acceptable pour chaque *domaine d'étude* (c'est-à-dire, chaque domaine sous-national tel que la région ou les milieux urbain ou rural), la taille d'ensemble de l'échantillon doit être augmentée de façon importante. Si l'on désire obtenir des résultats ayant un degré de précision *identique* pour chaque domaine, on augmente généralement la taille de l'échantillon *national*, n , par un facteur proche du nombre de domaines d'étude, et en sélectionnant ensuite n cas *dans chaque domaine*. Ainsi, si l'on désire obtenir des résultats ayant un degré de précision *identique* pour cinq régions du pays, la taille de l'échantillon qui a été calculée pour les estimations nationales sur la base du Tableau 4.3 ou directement à partir de la formule de calcul, devra être *multipliée par un facteur d'environ 5* pour obtenir des estimations régionales. Ceci, bien évidemment, accroît de manière sensible la taille globale de l'échantillon (et son coût), et peut s'avérer impossible à réaliser dans la plupart des pays¹⁸.

Différentes options devront être étudiées, en particulier si le nombre de domaines est important. Une option possible consiste à limiter les domaines d'études, comme les provinces, à ceux qui dépassent une certaine taille de population. Les zones restantes peuvent alors être combinées en groupe de régions. Une autre possibilité consiste à permettre que les niveaux de précision pour les estimations au niveau des différents domaines soient moins rigoureux que ceux des estimations nationales. Par exemple, si la marge d'erreur de l'indicateur-clé pour l'estimation nationale a été fixée à 12% de r (voir section sur la taille de l'échantillon), les domaines d'études

¹⁸ Au lieu de cela, Vijay Verma suggère d'augmenter la taille de l'échantillon national d'un facteur D^{65} , où D est le nombre de domaines. En suivant cette approche, le niveau de précision des estimations pour chaque domaine est légèrement inférieur aux estimations nationales. Voir "A Critical Review of MICS Sampling Methodology", rapport rédigé par Vijay Verma pour l'UNICEF, avril 1995.

pourraient voir leurs marges d'erreur considérablement réduites – allant même jusqu'à 25 à 30% de r . De plus, ces deux options possibles peuvent être combinées.

ANALYSE DES TENDANCES ET ANALYSE DE SOUS-POPULATIONS

Certains des objectifs internationaux sont exprimés en termes de réduction attendue, comme une baisse de 20% de la prévalence de la malnutrition au cours d'une période de 5 ans. Vous pouvez également avoir une enquête MICS ou EDS remontant à quelques années et qui fournissent des estimations des indicateurs couverts par la MICS3, et il peut arriver que vous désiriez apprécier les changements intervenus depuis lors. Ce type d'évaluation nécessite deux enquêtes, l'une au début, l'autre à la fin de la période. La taille de l'échantillon nécessaire pour mesurer des changements entre deux périodes est fortement dépendante de l'importance du changement, de même que de la valeur des estimations à chaque période. Il est très difficile et pratiquement impossible de fournir des règles simples et générales pour la mesure des changements. Si vous planifiez de mesurer des tendances, il est recommandé d'obtenir l'assistance du Bureau National des Statistiques ou de spécialistes en sondage.

Concernant les analyses de sous-populations, comme des indicateurs par sexe ou par groupe socio-économique, les estimations seront moins précises que celles concernant l'ensemble de l'échantillon.

L'exemple ci-dessous montre comment les marges d'erreur augmentent avec la diminution de la taille des sous-populations.

EXEMPLE :

Basé sur un échantillon d'ensemble (national) avec une marge d'erreur (précision), par exemple, de plus ou moins 5 points de pourcentage pour un taux de couverture de 50%, la marge d'erreur sera approximativement plus ou moins :

- 6,3 points de pourcentage pour les indicateurs pour chaque sexe, s'il y a 50% de garçons et 50% de filles dans l'échantillon,
- 8,6 points de pourcentage pour une sous-population correspondant à 20% de l'ensemble de l'échantillon¹⁹.

Par conséquent, des résultats ayant un niveau de précision acceptable peuvent être obtenus pour des indicateurs selon le sexe, ainsi que pour des sous-populations correspondant à un cinquième ou plus de l'ensemble de l'échantillon.

¹⁹ Voir la note non publiée destinée à l'UNICEF – “Some Proposed Modifications for the WHO Simplified Cluster Sampling Method for Estimating Immunization Coverage” – rédigée par Graham Kalton, Septembre 1988, 10 pages.

INDICATEURS SUR L'EAU ET L'ASSAINISSEMENT

La collecte de données sur l'eau et l'assainissement est une composante très importante de la MICS3. Cependant, l'utilisation et l'évaluation de ces données soulèvent des questions statistiques qui doivent être pris en compte.

Comme nous l'avons déjà signalé, le plan de sondage des MICS est basé sur la collecte de variables personnelles, opposées aux variables ménage. L'accès à l'eau et à l'assainissement constituent des caractéristiques du ménage, dans le sens où l'ensemble du ménage a le même type d'accès. De plus, dans une même zone de grappe, il arrive souvent que tous les ménages de la grappe aient le même type d'accès. Pour ces raisons, l'effet du plan de sondage, *deff*, dont nous avons déjà parlé, est particulièrement plus élevé pour les indicateurs d'accès à l'eau et à l'assainissement que la valeur 1,5, que nous avons retenu pour les calculs de la taille de l'échantillon. Comme corollaire, les erreurs de sondage pour les indicateurs sur l'eau et l'assainissement seront plus grandes que celles basées sur les individus.

Nous devons mettre en exergue le fait que si le but principal de la MICS3 avait été de recueillir des données sur les types d'accès à l'eau et à l'assainissement, le plan de sondage aurait été très différent. Il est probable que le plan d'échantillonnage occasionnerait, plus simplement, une enquête communautaire dans laquelle un seul informateur clé dans les communautés échantillonnées (zones de grappes) serait interrogé sur l'eau et l'assainissement, contrairement à l'approche de la MICS3, où ces questions sont posées à chaque ménage de l'échantillon.

Malgré l'ampleur des erreurs de sondage attendues pour les indicateurs d'accès à l'eau et à l'assainissement dans la MICS3, les résultats obtenus seront tout de même utiles, ne serait ce que pour la détermination des *tendances*. Cela parce que des données auront été collectées sur ces indicateurs dans d'autres enquêtes ménages, telles que les EDS, dont le plan est semblable à celui de la MICS3. La comparaison des résultats de la MICS3 avec ceux d'autres enquêtes serait d'une grande utilité dans la détermination des tendances, parce que les effets du plan de sondage, quoiqu'élevés dans toutes les deux enquêtes, tendent à s'annuler lorsque les estimations changent.

DETERMINATION DES ESTIMATIONS ET CALCUL DES ERREURS DE SONDRAGE

Cette section porte sur les différentes options de pondération pour déterminer les estimations et sur le calcul des erreurs de sondage.

Quand cela est nécessaire, deux types de pondérations peuvent être appliqués à la suite pour produire les estimations des indicateurs. À moins que l'échantillon de ménages ait été sélectionné à probabilité d'ensemble uniforme (c'est-à-dire, un échantillon auto-pondéré), toutes les données de l'échantillon devront être pondérées en utilisant l'inverse des probabilités d'ensemble de sélection, ou *poids selon le plan de sondage*. Cependant, les poids selon le plan de sondage devront être ajustés pour tenir compte des non-réponses, même si l'échantillon est auto-pondéré. Cela peut être fait de différentes façons, y compris en pondérant les enquêtés de chaque

UPS (ou grappe) pour qu'ils représentent les personnes non-enquêtées de cette UPS. Le principal avantage de cette approche est qu'elle ne nécessite aucune donnée externe. Ces deux étapes, à savoir appliquer les poids selon le plan de sondage et les ajustements pour les non-réponses, peuvent constituer l'ensemble des pondérations nécessaires à votre enquête.

Cependant, des pondérations additionnelles peuvent être appliquées par ajustement des poids selon le plan de sondage, de façon à rendre la distribution de l'échantillon pondéré selon certaines variables de base, comme la distribution urbain-rural ou par région, conforme à une distribution externe de la population, telle que le recensement le plus récent. Ce type de pondération après stratification doit être pris en compte quand il y a eu des déviations importantes du plan de sondage au moment de sa mise en œuvre, quand des procédures approximatives ont été utilisées du fait de déficiences de la base de sondage ou quand l'échantillon dévie du critère strict de probabilité.

Les formules et calculs véritables des poids selon le plan de sondage, des facteurs d'ajustement des non réponses et, éventuellement, des ajustements post-stratification sont fortement dépendants du plan de sondage utilisé – le nombre de degrés de tirage, le fait d'utiliser des plans auto-pondérés ou non auto-pondérés, le fait que les non réponses soient uniformes ou très variables à travers les sous-groupes de population ou les zones, la disponibilité de données externes pour la stratification à posteriori, etc. Par conséquent, il ne serait pas pratique de présenter de manière détaillée toutes les étapes de la procédure de calcul des poids pour votre MICS3 (cependant, un exemple de la méthode de calcul des poids pour une MICS spécifique figure dans l'exemple pays du Liban, dans la dernière section de ce chapitre). Le statisticien expert en sondage qui concevra votre échantillon devrait être parfaitement capable et entièrement responsable de la conception de l'approche de pondération, ainsi que de la préparation de la documentation s'y rapportant.

CALCULER LES ERREURS DE SONNAGE

Tel que cela a été suffisamment mis en exergue tout au long de ce chapitre, la taille d'échantillon de votre enquête est basée sur un indicateur-clé, pour lequel le niveau de précision attendu a été prédéfini. Cependant, l'enquête produira des centaines d'estimations, chacune d'elles ayant sa propre précision (marge d'erreur), et celles-ci vont varier selon qu'elles sont de niveau national ou sous-national, ainsi qu'en fonction de la taille de p , r et du *deff* associé à chacune d'elles (voir formule de la taille de l'échantillon). Tel que noté plus haut, il est attendu que les estimations sur l'eau et l'assainissement, par exemple, affichent des erreurs de sondage plus importantes que les autres indicateurs. C'est la raison pour laquelle il est important de calculer les erreurs de sondage pour certains (et non tous) indicateurs seulement.

La préparation des résultats de l'enquête doit comprendre l'estimation des erreurs de sondage et des variables associées, comme les effets du plan de sondage (*deffs*), pour les principaux indicateurs – peut-être 30 à 50 estimations différentes. Cela est essentiel pour évaluer la fiabilité des indicateurs estimés. À cette fin, *l'intervalle de confiance comprenant la marge d'erreur autour des estimations de l'enquête ne peut pas être déterminé, à moins que les erreurs de*

sondage n'aient été estimées. Sinon, l'interprétation des estimations sera sérieusement compromise.

Le calcul des erreurs de sondage ou *erreurs-types*, peut constituer une partie assez complexe de l'enquête. Les erreurs-types doivent être calculées de façon à prendre en compte la complexité du plan de sondage (effet de grappe, stratification et pondération). En règle générale, l'application inadaptée de formules d'un sondage aléatoire simple sous-estimera sérieusement les erreurs-types.

Comme certainement, différents types de plan de sondage seront utilisés dans les pays participants, y compris certains basés sur des échantillons existants, il n'est pas possible de présenter une procédure générale pour l'estimation des erreurs de sondage de la MICS3. Cependant, il y a une technique, connue sous le nom de *méthode de la grappe finale*, qui peut s'adapter à la plupart des plans de sondage utilisés dans la MICS3. Une simple feuille de calcul Excel et un programme SPSS - accompagnés d'instructions sur la méthode - ont été conçus à cet effet. Tous ces éléments se trouvent sur le site Web de la MICS3, www.childinfo.org. En général, la méthode de la grappe finale est appropriée si votre échantillon est auto-pondéré ou à peu près auto-pondéré.

Si la feuille de calcul ne peut être utilisée, il existe de nombreux logiciels qui ont été développés et qui peuvent être facilement adaptés pour l'estimation de la variance. Ces logiciels comprennent le programme CLUSTERS développé à l'origine pour l'Enquête Mondiale de Fécondité et disponible auprès de l'Université d'Essex ; CENVAR, un logiciel disponible gratuitement auprès du Bureau du Recensement des USA ; et WesVar, un programme conçu par WESTAT pour être utilisé avec SPSS. Certains logiciels tels que SAS, SUDAAN et Epi Info peuvent aussi prendre en charge les plans de sondage complexes. Les versions récentes de SPSS ont la capacité de calculer les erreurs de sondage pour les plans de sondage complexes. Certains de ces logiciels sont gratuits et peuvent même être récupérés sur l'Internet, alors que d'autres sont commercialisés²⁰.

Dans tous les cas, que ce soit la feuille de calcul ou un logiciel qui soit utilisé, la procédure sera facilitée si l'on veille à ce que l'enregistrement des données contienne les identifiants des UPS. Si des strates ont été utilisées, les strates doivent être identifiées pour chaque UPS.

²⁰ Une revue complète de ces programmes a été publiée, intitulée "Sampling Error Software for Personal Computers", par Jim Lepkowski et Judy Bowles de l'Université du Michigan. L'article est paru dans *The Survey Statistician*, No. 35, Décembre 1996, pages 10-17 (voir le site Web : www.fas.harvard.edu/~stats/survey-soft/iass.html). Une revue plus récente est disponible dans une publication de la Division des Statistiques des Nations Unies, "Household Sample Surveys in Developing and Transition Countries", Mars 2005, et plus particulièrement le Chapitre 21, sur "Sampling Error Estimation for Survey Data", préparée par Donna Brogan (http://unstats.un.org/unsd/HHsurveys/pdf/Household_surveys.pdf).

TECHNIQUES DETAILLEES DE SELECTION D'ECHANTILLONS

Cette section fournit des exemples détaillés de la manière d'appliquer les techniques de sélection d'échantillons qui ont été présentés dans ce chapitre.

PROCEDURES D'ECHANTILLONNAGE AVEC PPT – OPTION 2

La première technique qui est illustrée ici est la façon de tirer des unités au premier degré en utilisant les *ppt*. Cette illustration montre également la façon de combiner un échantillonnage systématique avec *ppt* avec un ordre géographique de la base de sondage pour réaliser la stratification *implicite*.

Pour l'illustration, nous prenons l'Option 2 de ce chapitre, la méthodologie des segments-types, et nous sélectionnons un échantillon national. Supposez que : (1) la taille du segment-type sous l'Option 2 soit de 500 personnes, ou environ 100 ménages ; (2) les zones de dénombrement (ZD) du recensement soient la base de sondage ; et (3) le nombre d'unités primaires de sondage (UPS) à choisir soit de 300. Les étapes du tirage au premier degré qui suivent *devraient être effectuées sur un ordinateur*, bien qu'il soit possible de les effectuer manuellement.

- Étape 1 :** Triez le fichier des ZD par urbain et rural.
- Étape 2 :** Dans la catégorie urbaine, triez davantage le fichier dans un ordre géographique serpentin selon les subdivisions administratives de votre pays (par exemple, province ou état, district, commune, etc.).
- Étape 3 :** Répétez l'étape 2 pour la catégorie rurale.
- Étape 4 :** Dans une colonne, montrez la population des ZD du recensement.
- Étape 5 :** Dans la colonne suivante, calculez le nombre de segments ou grappes-types, qui est égal au dénombrement de la population divisé par 500, et arrondi au nombre entier le plus proche. C'est la mesure des tailles des ZD.
- Étape 6 :** Faites le cumul des mesures des tailles des ZD dans la colonne suivante.
- Étape 7 :** Calculez l'intervalle de tirage, I , en divisant le total cumulé par 300, avec un chiffre après la virgule. Dans cette illustration, supposez que le total cumulé est de 5 281. Dans ce cas, l'intervalle de tirage, I , serait égal à $5.281/300$ ou à 17,6.
- Étape 8 :** Choisissez un nombre aléatoire compris entre 0 et 17,6. La façon de le faire, dans la pratique, est d'utiliser une table des nombres aléatoires, et de choisir un nombre de trois chiffres entre 001 et 176 et d'insérer la décimale après. Supposez que vous choisissiez.042 ; votre nombre aléatoire sera donc 4,2. Par conséquent, votre première UPS de l'échantillon serait celle pour laquelle la mesure cumulée de la taille est la plus petite valeur supérieure ou égale à 4,2²¹.

²¹ Kish recommande d'arrondir au chiffre inférieur quand l'intervalle de tirage est très petit. Voir Kish, L. (1965) *Survey Sampling*, Wiley, New York, page 116.

- Étape 9 :** Additionnez 4,2 à I, soit $4,2 + 17,6 = 21,8$; ainsi, votre UPS suivante de l'échantillon serait celle dont le cumul correspond à la plus petite valeur supérieure ou égale à 21,8.
- Étape 10 :** Additionnez 21,8 à I, soit $21,8 + 17,6 = 39,4$; l'UPS suivante de l'échantillon est celle dont le nouveau cumul correspond à la plus petite valeur supérieure ou égale à 39,4.
- Étape 11 :** Continuez comme ci-dessus, pour les ZD urbaines suivies des ZD rurales, jusqu'à ce que toutes les 300 UPS soit sélectionnées.

La procédure est présentée plus amplement au Tableau 4.11.

Les deux UPS de l'échantillon qui sont présentées dans l'illustration sont celles dans la ZD 003 de la commune 01 et dans la ZD 002 de la commune 03, toutes dans le district 01 et dans la province 01. Dans le cas de la première ZD, sa taille est de 3, ce qui signifierait la création de trois segments, chacune comprenant approximativement 540 personnes (1.630 divisées par 3), et donc un des segments serait choisi au hasard pour le listing et la sélection des ménages. Dans la deuxième ZD de l'échantillon, deux segments seraient créés, chacun contenant environ 590 personnes, avant de choisir l'un d'eux au hasard.

L'illustration démontre les nombreux avantages de la stratification implicite. Premièrement, il est très facile de la réaliser, simplement on exige que la base des zones de dénombrement soit triée géographiquement avant de sélectionner l'échantillon systématiquement avec *ppt*. Deuxièmement, elle fournit automatiquement un échantillon d'UPS qui est proportionnellement répartie entre milieux urbain et rural et par province (ou d'autres subdivisions géographiques). Par exemple, si 10% de votre population sont situés dans la province 12, alors 10% de votre échantillon seront également choisis dans cette province. Troisièmement, elle peut être facilement mise en application sur ordinateur.

Tableau 4.11
Illustration d'un échantillonnage systématique avec *ppt* et Stratification Implicite –
Option d'échantillon 2

Urbain	Population	Mesure de la taille (segments d'une population de 500)	Cumul
Province 01			
District 01			
Commune 01			
ZD 001	1 470	3	3
ZD 002	562	1	4
ZD 003	1 630	3	7 tirés
ZD 004	1 006	2	9
Commune 02			
ZD 001	412	1	10
ZD 002	1 537	3	13
ZD 003	1 312	3	16
ZD 004	397	1	17
Commune 03			
ZD 001	1 540	3	20
ZD 002	1 181	2	22 tirés
ZD 003	1 025	2	24
District 02			
Commune 01			
ZD 001	567	1	25
ZD 002	1 111	2	27
ZD 003	409	1	28
*			
*			
etc.			
Rural			
Province 12			
District 05			
Commune 05			
ZD 001	512	1	5 280
ZD 002	493	1	5 281

Une fois que les UPS ont été choisies, sous l'Option 2, la segmentation devra être effectuée dans les UPS dont la taille (nombre de segments) est de deux ou plus, suivie d'un segment sélectionné aléatoirement dans chaque UPS. Ensuite, un nouveau listing des ménages devra être fait dans les segments sélectionnés en plus des UPS à un seul segment. L'étape finale de la procédure de sélection pour l'Option 2 est de sélectionner les ménages de l'échantillon dans les segments choisis. Cette procédure est décrite au Tableau 4.12 avec une illustration.

Tableau 4.12**Sélection des ménages – Option 2**

Supposez que la taille d'un segment type soit 500 personnes, et que la taille désirée de la grappe de votre échantillon soit désignée comme ménages \bar{n} .

1. Calculez les ménages moyens par segment en divisant 500 par la taille moyenne de ménages dans votre pays. Ceci est désigné par s_h .
2. Divisez s_h par \bar{n} . C'est l'intervalle de tirage, I , pour sélectionner les ménages dans chaque segment de l'échantillon.

(Remarque : si la taille type de votre segment est différente de 500, cette autre valeur doit être naturellement utilisée).

Illustration :

Supposez que votre taille moyenne de ménage soit de 5,5. Donc s_h est $500/5,5$ ou 90,9. Supposez que vous vouliez que la taille de vos grappes, \bar{n} , soit de 25. Divisez 90,9 par 25 (1 chiffre après la virgule) = $90,9/25$ ou 3,6. Puis, sélectionnez les ménages dans chaque segment au pas de 1 sur 3,6, en commençant par un nombre aléatoire entre 01 et 36 (en insérant la décimale après le nombre choisi).

PROCEDURES D'ECHANTILLONNAGE AVEC PPT – OPTION 3

Si l'Option 3 - la méthodologie modifiée des segments décrite dans ce chapitre - est utilisée au lieu de l'Option 2, la stratification implicite est faite de la même manière, *bien que la mesure de la taille soit différente*. Sous l'Option 3, si nous supposons, comme exemple, que la taille de notre segment va être de 20 ménages (en moyenne), donc la mesure de la taille serait calculée en divisant le comptage des ménages du recensement par 20, arrondi au nombre entier le plus proche. Notez que sous l'Option 3, la deuxième colonne du Tableau 4.13 doit être le nombre de ménages plutôt que la population. Vous calculeriez l'intervalle de tirage, I , en divisant le total cumulé - supposez qu'il soit de 26.425 - par le nombre désiré d'UPS, qui est fixé encore à 300. Ainsi, vous auriez $26.425/300 = 88,1$. Si le nombre aléatoire choisi est de 19,4, les deux premières UPS choisies, comme l'indique le Tableau 4.13, seraient celles correspondant aux plus petits cumuls excédant les valeurs 19,4 et 107,5 ($88,1 + 19,4$) respectivement. Elles sont la ZD 002 dans la commune 01 et la ZD 002 dans la commune 03 de la province 01 et du district 01.

Tableau 4.13

Illustration d'un échantillonnage systématique avec ppt et Stratification Implicite – Option d'échantillon 3

Urbain	Population	Mesure de la taille (segments d'une population de 500)	Cumul
Province 01			
District 01			
Commune 01			
ZD 001	290	14	14
ZD 002	120	6	20 tirés
ZD 003	325	16	36
ZD 004	200	10	46
Commune 02			
ZD 001	81	4	50
ZD 002	307	15	65
ZD 003	261	13	78
ZD 004	80	4	82
Commune 03			
ZD 001	308	15	97
ZD 002	236	12	109 tirés
ZD 003	205	10	119
*			
*			
*			
etc.			
Rural			
Province 12			
District 05			
Commune 05			
ZD 001	102	5	26 400
ZD 002	99	5	26 405

Rappelez-vous que, sous l'Option 3, la mesure de la taille est équivalente au nombre de segments de taille prédéfinie qui doivent être créés (dans notre illustration, c'est 20). Ainsi, pour les UPS ou ZD tirées de l'échantillon, 6 segments d'une taille approximative de 20 ménages chacun doivent être créés dans la première UPS, et 12 segments dans la deuxième. De nouveau, un des segments serait donc choisi au hasard dans chaque UPS de l'échantillon, *et tous les ménages dans ce segment seraient interviewés pour l'enquête*, même si le nombre réel de ménages dans le segment s'écarte de manière significative de sa taille prévue.

Le chapitre 6 présente les détails des procédures pour créer des segments pour l'Option 2 et l'Option 3.

EXEMPLES-PAYS DE LA SÉRIE DES MICS DE L'AN 2000

Dans la dernière section de ce chapitre, nous présentons les plans de sondage qui ont été utilisés dans trois pays ayant participé à la série des MICS de l'an 2000. Chacun d'eux illustre un aspect différent des questions d'échantillonnage qui a été examiné dans ce chapitre. Les exemples en question sont ceux du Liban, de la Papouasie Nouvelle-Guinée et de l'Angola.

LIBAN

Nous commençons par le Liban, et faisons une description particulièrement détaillée, afin de comparer le plan de sondage avec celui du Manuel MICS de l'an 2000. De plus, nous montrons comment les poids ont été calculés pour illustrer comment cela se fait dans une enquête spécifique.

Plan de Sondage

L'enquête MICS du Liban a été réalisée par le Bureau National de la Statistique (*Central Administration of Statistics*). Les principaux objectifs de mesure de la MICS étaient : (1) de recueillir les indicateurs appropriés du Sommet Mondial pour l'Enfance de la fin de la décennie, en vue de cerner la situation des femmes et des enfants au Liban, et (2) de fournir ces estimations pour le niveau national et les cinq zones (domaines) du niveau sous-national, à savoir les quatre plus grands mohafazats (gouvernorats) en plus de la combinaison du Sud et de Nabatiyeh. La méthodologie d'échantillonnage de la MICS du Liban était de sélectionner un échantillon probabiliste de 1.625 ménages dans chacun de ces cinq domaines, en utilisant un plan de sondage stratifié par grappes. C'était une procédure de sélection à deux degrés utilisant le Recensement de l'Habitation de 1995-1996 comme base de sondage. Dans l'ensemble, 8.125 ménages devaient être sélectionnés sur l'ensemble du territoire.

Les grappes à tirer au premier degré (autrement dit, les unités primaires de sondage – UPS) avaient été définies comme des ilots, et ceux-ci étaient identiques en termes de conception et de construction aux zones de dénombrement du recensement. Elles avaient été sélectionnées avec probabilité proportionnelle à leur taille (ou *pps*), cette dernière étant celle des décomptes des logements occupés en 1995-1996 tirés du Recensement de l'Habitation. Un total de 65 ilots fut sélectionné dans chaque domaine, soit un total de 325. Les ilots sélectionnés au premier degré ont du être listés de nouveau, c'est-à-dire une nouvelle énumération des ménages préparée lors d'une opération de terrain. Les ilots qui étaient très larges furent divisés en segments géographiques, dont un seul fut sélectionné de manière aléatoire à partir de la fiche d'énumération. C'est cette nouvelle fiche d'énumération qui a constitué la base de sondage pour la sélection au second degré.

Une taille d'échantillon fixe de 25 ménages des nouvelles fiches d'énumération fut sélectionnée de manière systématique dans chaque ilot échantillonné for les interviews de la MICS. Dans quelques uns des ilots échantillonnés, autrement dit ceux qui abritaient moins de 25 ménages, tous les ménages furent enquêtés pour la MICS.

Il avait été prévu que, lors de l'enquête, les non-réponses s'élèveraient à 10 ou 12%, cas dans lequel le nombre de ménages à interviewer varierait entre 1,430 et 1,465 dans chaque domaine, soit entre 7,150-7,325 sur l'ensemble du territoire.

Discussion du Plan de Sondage – Philosophie de la MICS

Il y a lieu de noter que le plan de sondage décrit ci-dessus concorde avec la philosophie d'ensemble de la MICS en ce qui concerne la conception et la mise en œuvre de l'échantillonnage, selon l'édition précédente du manuel de la MICS. Cette édition, à l'instar de celle-ci, promeut un certain nombre de conditions pour la conception du plan de sondage, et toutes celles-ci ont été rigoureusement respectées par le plan réalisé par la Direction Nationale de la Statistique du Liban. Celles-ci comprennent :

- L'utilisation d'une méthodologie d'échantillonnage simple ;
- L'utilisation du dernier recensement comme base de sondage ;
- Un échantillonnage à deux degrés ;
- L'utilisation de techniques probabilistes à chaque degré de sélection ;
- L'utilisation de grappes de tailles moyenne ;
- L'utilisation d'une taille d'échantillon adéquate.

La méthodologie d'échantillonnage est simple et directe, plutôt que complexe ; pour cette raison, il était attendu que la marge d'erreur serait réduite au minimum. Le Recensement de l'Habitation de 1995-1996 était la meilleure base de sondage à utiliser, car (1) c'était la plus récente disponible, et (2) des cartes étaient disponibles, ce qui permettaient de localiser facilement les ilots sélectionnés pour la MICS, et d'identifier leurs limites internes. Cet aspect cartographique est également extrêmement pratique pour l'opération de segmentation qui était nécessaire pour les ilots trop grands. L'âge de la base de sondage impliquait que des mouvements naturels et des variations de la population pourraient réduire la précision des estimations, et pour cette raison, une nouvelle énumération des ménages au second degré était nécessaire, afin de mettre à jour la base dans les zones sélectionnées.

Des méthodes strictes de sélection probabiliste furent utilisées à tous les deux degrés. Cela devait permettre de garantir que les résultats seraient représentatifs des différentes populations ciblées, y compris les ménages, les femmes en âge de reproduction et les enfants du Liban. La taille de grappe de 25 ménages est acceptable, ce qui permet de garantir que l'effet du plan de sondage ne serait pas grand au point de remettre en question la fiabilité des résultats pour des indicateurs clés.

La taille de l'échantillon national de 8.125 ménages était en conformité avec les recommandations de l'édition 2000 du manuel de la MICS relatives à cet important paramètre de l'enquête. Comme le stipulait ce manuel, la taille de l'échantillon devait être suffisamment grande pour permettre la réalisation de plusieurs objectifs de mesure. Au nombre de ceux-ci figurent la possibilité de mesurer les indicateurs clés des cibles avec une marge d'erreur de 3-5 points de pourcentage, cela en fonction de l'indicateur. Un second objectif important était d'être

à même d'obtenir des estimations raisonnablement fiables des indicateurs pour les domaines concernés (tel que mentionnés dans le résumé, ceux-ci incluent les quatre grands gouvernorats de Beyrouth, du Mont Liban, de la Beqaa et du Nord, en plus de la combinaison Sud/Nabatiyeh, ainsi que le niveau national).

A l'instar de la présente édition du manuel de la MICS, le manuel de 2000 suggérait que le calcul de la taille de l'échantillon aurait besoin d'être ajusté à la hausse pour prendre en compte les pertes de cas dues aux non-réponses. Cette condition fut respectée.

Taille de l'Echantillon

L'indicateur retenu pour la détermination de la taille de l'échantillon au Liban était la proportion des enfants de moins de cinq ans ayant eu la diarrhée au cours des deux semaines précédentes. Il était attendu que cette proportion tournerait autour de 15%, niveau pour lequel la marge d'erreur, ou la précision, était fixée à plus ou moins 3%. Alors, l'intervalle de confiance, à 95% de niveau de confiance, devrait conduire à une valeur de l'enquête située entre 12 et 18%, si la vraie valeur au sein de la population était de 15%.

La population cible pour l'indicateur sur la diarrhée est constituée des enfants de moins de cinq ans, et ils représentent à peine un peu moins de 10% de la population totale. L'effet du plan de sondage a été estimé à environ 1,5. C'était le facteur à utiliser pour augmenter la taille de l'échantillon, afin de contrebalancer le fait qu'un sondage par grappe ait été utilisé, en lieu et place d'un sondage aléatoire simple, eu égard à ses effets sur la fiabilité de l'échantillonnage.

Les domaines étaient au nombre de cinq, comme susmentionné. Le taux de non-réponse attendu était estimé à 12,5%. Ce qui est un chiffre prudent, conduisant à l'élaboration d'un échantillon plus grand que nécessaire. On espérait que le taux effectif de réponse de l'enquête serait supérieur à 87,5%. La taille moyenne des ménages au Liban en ce moment là était de 4,8. Le facteur requis pour atteindre un niveau de confiance de 95% est 4.

Compte tenu de tous les paramètres qui précèdent, la formule ci-dessous de calcul de la taille de l'échantillon (voir Annexe 7 du Manuel de la MICS 2000) donne le nombre de ménages dont on a besoin pour l'échantillon d'un domaine. Il a fallu le modifier pour cinq domaines, tel que présenté ci-dessous.

$$n = \frac{4r(1-r)(1,125)f}{e^2(p)(m)} = \frac{4(0,15)(0,85)(1,125)(1,5)}{(0,03)(0,03)(0,1)(4,8)} = 1992$$

Le manuel de la MICS 2000 suggère deux alternatives pour l'ajustement de la taille des domaines. L'une consiste à multiplier le n dont on a besoin pour un domaine par le nombre de domaines. Cela donne 5 x 1.992, soit 9.961. Etant donné que le budget de l'enquête ne peut prendre en charge une taille d'échantillon aussi grande, une autre approche qui donnerait des marges d'erreurs plus élevées consiste à multiplier n par le facteur D à la puissance 0,65, où D est le nombre de domaine. Cela équivaut approximativement à prendre la racine cubique du carré

de D, ce qui, dans notre cas, est la racine cubique de 25, soit environ 2,92. Ce nombre, multiplié par 1.992, donne une taille d'échantillon de 5.817 pour les cinq domaines. Il a été retenu d'utiliser la moyenne de ces deux alternatives, 7.889. Ce chiffre a ensuite été ajusté à la hausse afin de couvrir 65 grappes par domaine (7.889 exigerait environ 63 grappes, autrement dit $7.889/25$, soit environ 315 ou 316 au total, puis $315/5$, ou 63 dans chaque domaine).

Il est utile de voir l'effet de la marge d'erreur avec cette taille d'échantillon. La valeur calculée de n pour la diarrhée est de 1.992. Il a été signalé plus haut que cela donnerait une marge d'erreur de plus ou moins 3 points de pourcentage. Cependant, la taille d'échantillon finale est de $8.125/5$ ou 1.625 par domaine. Ainsi, la marge d'erreur est quelque peu plus élevée. L'accroissement est calculé comme la racine carrée du ratio des tailles d'échantillon, ou la racine carrée de $1.992/1.625$, qui est d'environ 1,033. En définitive, la marge d'erreur attendue pour cet indicateur spécifique au niveau gouvernorat serait d'environ 3,3 points de pourcentage, si le taux de prévalence pour la diarrhée était approximativement de 15%.

Distribution de l'Echantillon

Les objectifs de mesure doivent être pris en compte lors de l'allocation optimale de l'échantillon. Si la priorité est donnée aux estimations du niveau national, l'échantillon devrait être réparti de manière proportionnelle entre les différents gouvernorats. Par exemple, si un gouvernorat abrite 35% de la population du pays, alors sous l'approche de distribution proportionnelle, 35% de l'échantillon sera sélectionné dans ce gouvernorat. Le problème avec cette approche réside dans le fait que les petits gouvernorats (disons, ceux qui abritent seulement 10% de la population du pays) auraient des tailles d'échantillon trop petites pour donner des estimations fiables. Cette dernière constitue un problème majeur si la priorité est donnée au niveau gouvernorat, contrairement au niveau national.

En fait, la priorité a été donnée au niveau gouvernorat par le gouvernement et l'UNICEF/Liban. Cela se justifiait par le fait que la planification programmatique et l'exécution des projets pour l'amélioration de la situation des femmes et des enfants devaient se faire au niveau local. Pour cette raison, les estimations les plus fiables possibles étaient recherchées au niveau de chacun des gouvernorats. Pour y parvenir, l'échantillon devait être distribué de manière équitable entre les gouvernorats (les quatre grands plus la combinaison du sud).

Le Tableau 4.14 présente la distribution de l'échantillon et des autres paramètres par gouvernorat.

Tableau 4.14
Paramètres de l'échantillon – MICS 2000 du Liban

<i>Mohafazat (gouvernorat)</i>	<i>Taille d'échantillon – nombre de ménages (hhs)</i>	<i>Taille d'échantillon – nombre de grappes (ilots)</i>	<i>Taille de grappe – nombre de hhs échantillonnés par grappe</i>	<i>Nombre total d'unités primaires de la base (logements occupés)</i>	<i>Intervalle d'échantillon pour la sélection des ilots ppt</i>
Beyrouth	1.625	65	25	101.707	1 sur 1.564,7
Mont Liban	1.625	65	25	356.517	1 sur 5.484,9
Nord	1.625	65	25	150.018	1 sur 2.308,0
Beqaa	1.625	65	25	97.638	1 sur 1.502,1
Sud/Nabatiyeh	1.625	65	25	149.093	1 sur 2.293,7
Total Liban	8.125	325	25	854.973	-

Mise en œuvre – Procédures de Sélection de l'échantillon

Le premier degré de sélection a consisté en un échantillonnage systématique des ilots du recensement avec probabilité proportionnelle à la taille (*pps*), et la mesure de la taille était constituée du décompte de logements occupés. Comme préalable à cette sélection, les ilots ont été triés par gouvernorat et, à l'intérieur de chaque gouvernorat, par caza (district), grande zone et ilot, cette dernière ayant été classée selon un ordre géographique. Toute cette démarche a été faite en utilisant le code d'identification géographique à sept chiffres de la base de données géographiques. De cette façon, la sélection systématique qui a été opérée a permis de garantir la stratification implicite de l'échantillon pour chacun des domaines, ou gouvernorats. Soixante cinq ilots furent ainsi sélectionnés par ordinateur pour chaque domaine.

Pour la mise en œuvre du second degré de sélection, une nouvelle opération d'énumération des ménages fut réalisée. La plupart des ilots échantillonnés étaient beaucoup trop grands, et il fut nécessaire de les segmenter avant de procéder à l'énumération. Environ 180 des 325 ilots furent segmentés, car – selon la base de sondage de 1995-1996 – elles contenaient chacune plus de 150 ménages.

La segmentation a été effectuée selon le Tableau 4.15.

Tableau 4.15
Nombre de Segments à Créer

Nombre de ménages (selon la base de sondage)	Nombre de segments
Moins de 150	Pas de segmentation
150-249	2
250-349	3
350-449	4
450-549	5
550-649	6
650-749	7
Etc.	Etc.

Certains des ilots échantillonnés, environ 12 d'entre eux, avaient moins de 25 ménages, selon la base de sondage. Aucune opération d'énumération ne fut nécessaire pour ces ilots. En lieu et place, tous les ménages situés à l'intérieur des limites de l'ilot furent interviewés pour la MICS, même si le nombre effectif de ces ménages au moment de la réalisation de la MICS en 2000 était supérieur à 25.

Pour la segmentation, le Tableau 4.15 fut utilisé pour calculer le nombre de segments à créer. Par exemple, si la base de sondage révélait que l'ilot échantillonné contenait 580 ménages, six segments furent créés. Toutes les procédures de construction des segments présentées dans le Chapitre 6 de l'édition 2000 du manuel de la MICS furent suivies. Après la segmentation, un segment fut sélectionné de façon aléatoire dans chaque ilot segmenté.

Pour tous les ilots (moins de 150 ménages plus les segments sélectionnés pour ceux qui ont plus de 150 ménages), une nouvelle descente sur le terrain fut effectuée en vue d'effectuer une énumération exhaustive des ménages actuels. Ensuite, l'échantillonnage des ménages fut systématique au sein de chaque ilot, ou segment, au taux de 1 sur I, où I était l'intervalle de tirage de l'échantillon, égal à $N_i \div \eta$, où N_i était le nombre réel de ménages issu de l'énumération pour la grappe échantillonnée i^{th} , et η était de 25, la taille désirée de la grappe. La valeur de l'intervalle de tirage de l'échantillon, I, était variable pour chaque ilot ou segment échantillonné et il était calculé avec un chiffre après la virgule.

Pondération et Estimation

Pour la préparation des estimations de l'enquête, la pondération devait nécessairement être utilisée, car l'échantillon n'était pas auto-pondéré ; en d'autres termes, les différentes unités de l'échantillon n'avaient pas la même probabilité de sélection. La pondération a nécessité une multiplication des données brutes de l'échantillon par le poids de sondage de l'échantillon. Les poids étaient des variables définies par grappe, puisque le poids des ménages dans chaque grappe est fonction de la mesure de la taille et du nombre de ménages issu de l'opération d'énumération pour cette grappe.

Les poids sont égaux à l'*inverse* des probabilités de sélection. La probabilité pour les ménages, les individus, les enfants, les femmes ou tout autre groupe de population-cible dans une grappe spécifique est donnée par la formule suivante :

$$P_{ih} = \frac{(65)(25)m_{ih}}{M_h N_{ih} S_{ih}} \text{ Où}$$

P_{ih} est la probabilité de sélectionner un ménage (ou une personne) dans la grappe i^{th} du domaine hth ;

65 est le nombre de grappes dans chaque domaine ;

25 est la taille de la grappe ;

m_{ih} est la mesure de la taille dans la grappe i^{th} du domaine hth, autrement dit, le nombre de logements occupés selon la base de sondage ;

- N_{ih} est le nombre de ménages énumérés dans la grappe i^{th} du domaine h ;
 s_{ih} est le nombre de segments créés dans un ilot (la valeur de s_{ih} est 1 pour les ilots qui n'ont pas été segmentés) ;
 M_h est la mesure de la taille dans le domaine h , autrement dit, la somme des valeurs de m_{ih} , ou $\sum m_{ih}$.

Notez que la valeur de N_{ih} est le nombre de ménages énumérés dans la grappe, ce qui dans les ilots non segmentés est l'ilot tout entier, tandis que dans les ilots segmentés c'est la partie segmentée (voir les exemples qui suivent).

Comme susmentionné, le poids, W_{ih} , est l'inverse de la probabilité, ou $1/P_{ih}$.

Les valeurs M_h sont celles figurant dans la 4^{ème} colonne du Tableau 4.14 pour chacun des domaines. Le produit de 65 par 25 donne 1.625. Les poids pour les ménages ou les individus pour les cinq domaines sont présentés dans le Tableau 4.16.

Tableau 4.16
Poids de l'Enquête par Domaine (pour des Ilots de 25 ménages ou plus)

Gouvernorat	Poids
Beyrouth	$(62.6N_i s_i)/m_i$
Mont Liban	$(219.4N_i s_i)/m_i$
Nord	$(92.3N_i s_i)/m_i$
Beqaa	$(60.1N_i s_i)/m_i$
Sud/Nabatiyeh	$(91.7N_i s_i)/m_i$

EXEMPLE :

Supposons qu'à Beyrouth la mesure de la taille, m_i , dans la base de sondage pour la première grappe d'échantillon sélectionnée soit de 612. Alors, elle sera segmentée en six segments, qui seraient la valeur de s_i . Supposons ensuite que le nombre de ménages énumérés dans la grappe soit de 110. Alors, le poids pour tous les ménages et les individus de l'échantillon de cette grappe serait égal à $(62,6)(110)(6)/(612)$, ou 67,5.

EXEMPLE :

Supposons qu'au Mont Liban, la mesure de la taille, m_i , dans la base de sondage pour la première grappe d'échantillon sélectionnée soit de 106. Donc, elle ne sera pas segmentée et la valeur de s_i serait de 1. Supposons ensuite que le nombre de ménages énumérés dans la grappe soit de 98. Alors, le poids pour tous les ménages et les individus de l'échantillon de cette grappe serait égal à $(219,4)(98)(1)/(106)$, ou 202,8.

On pensait qu'il serait nécessaire d'ajuster les poids *selon le plan de sondage* présentés ci-dessus par un facteur supplémentaire pour tenir compte des non-réponses. Il s'agissait simplement de multiplier le poids, séparément pour chaque domaine, par le ratio des ménages échantillonnés et des ménages interviewés. Le produit du poids selon le plan de sondage et du facteur de non-interview devient alors le poids final, qui sera multiplié par les données brutes de l'échantillon pour obtenir les estimations.

PAPOUASIE NOUVELLE GUINEE

L'enquête MICS en Papouasie Nouvelle Guinée fut réalisée par le Bureau National de la Statistique (National Statistics Office). Sa conception illustre les aspects relatifs à la mise à jour et à la couverture de la base de sondage, de même que le fait d'avoir à choisir entre l'échantillon d'une enquête existante (EDS) et en concevoir un nouveau. La méthodologie de calcul des poids pour la MICS de la Papouasie Nouvelle Guinée est aussi très instructive. Mais elle est quelque peu semblable à celle du Liban, et ne sera donc pas présentée ici.

Modification de la Base de Sondage et EDS comme Alternative

L'échantillonnage a suivi un processus de sélection à deux degrés utilisant le Recensement de la Population de 1991, partiellement mis à jour, comme base de sondage. Les unités primaires de sondage (UPS) avaient été définies comme des zones ou unités de dénombrement du recensement (ZD). Un total de 175 UPS fut sélectionné avec probabilité proportionnelle à leur taille (*ppt*), cette dernière étant le décompte des ménages ruraux de 1991 et le décompte à jour des logements urbains. Les ZD du recensement sélectionnées au premier degré furent de nouveau énumérées ; autrement dit, une nouvelle énumération des ménages fut préparée au cours d'une opération de terrain. La nouvelle liste d'énumération servit comme base de sondage pour le tirage au second degré. Un échantillon de taille fixe de 25 ménages tiré des listes d'énumération fut sélectionné systématiquement dans chaque ZD échantillonnée pour la MICS. Au total, 4.375 ménages constituèrent la taille visée de l'échantillon.

Le calcul du nombre de grappes échantillonnées a donné 175, en divisant 4.375 par 25. Environ 43% d'entre elles furent affectées aux strates urbaines (75 grappes) et 57% aux strates rurales (100 grappes). Avant de procéder au tirage au premier degré, la base a subi une procédure de stratification à posteriori, qui a consisté à trier, à un premier niveau, le Recensement de Population par milieu – urbain et rural, et par région, province, district, sous-district et ZD aux niveaux suivants.

Modification de la Base de Sondage et EDS comme Alternative

Le recensement le plus récent à utiliser comme base pour l'échantillonnage de la MICS 2000 MICS en Papouasie Nouvelle Guinée était le Recensement de la Population de 1991 – or, il datait de 9 ans en ce moment. Il n'était plus à jour à différents degrés dans différentes zones du pays, mais surtout dans les zones urbaines, notamment à Port Moresby. Au moment de la MICS, le Bureau National de la Statistique était également entrain de préparer le Recensement de la Population de l'an 2000, et il comptait se servir des mêmes unités administratives du recensement, ZD, qui avaient été utilisées lors du Recensement de 1991. Le processus de mise à jour de ces ZD était en cours, à travers une nouvelle énumération des ménages. Au moment de la MICS, la mise à jour n'était effective que dans les zones urbaines, qui, compte tenu des objectifs de la MICS, constituaient le secteur le plus important en termes d'utilisation des données comme base de sondage.

Il y avait deux alternatives : (1) soit reporter la MICS jusqu'à ce que les opérations de mise à jour des zones urbaines soient terminées, (2) soit utiliser l'ancien recensement comme base de sondage pour les zones urbaines et la nouvelle énumération pour les zones rurales. Ce qui est le

plus important pour toute base est d'avoir des mesures de la taille – c'est-à-dire le décompte des ménages à utiliser pour définir le taux de sélection – qui soient relativement précis. Cependant, un point important à ne pas perdre de vue est qu'il n'est pas nécessaire que les mesures de la taille soient parfaites à 100%, pour être valablement utilisées pour le tirage de l'échantillon au premier degré. Par conséquent, il était prévu que les ZD rurales issues de l'ancien recensement auraient un décompte des ménages qui serait différent de la situation en 2000 de quelques points de pourcentage seulement ; donc, elles resteraient encore assez précises. A contrario, les ZD urbaines seraient probablement différentes, peut-être par des facteurs de 2 ou 3, ou même plus grands, comparativement à la situation d'il y avait 9 ans. Ce phénomène s'explique par le fait que, alors que relativement peu de personnes migrent de chaque ZD rurale, toutes celles qui migrent se dirigent vers les mêmes ZD urbaines, qui sont relativement peu nombreuses. Ainsi, il est beaucoup plus pertinent d'avoir des décomptes à jour des ZD urbaines que des ZD rurales.

Par conséquent, pour la MICS, il fut décidé que deux bases de sondage seraient utilisées pour le plan de sondage. Pour les ZD rurales, le Recensement de la Population de 1991 fut utilisé comme base. Pour les ZD urbaines, les listes à jour des ZD, avec leurs décomptes de logements révisés – et préparés pour le Recensement de la Population de l'an 2000 qui servit comme base de sondage.

On a pensé à utiliser l'Enquête Démographique et de Santé (EDS) de 1996 comme base de sondage alternative au Recensement de la Population de 1991. Cette enquête avait des objectifs semblables à ceux de la MICS et l'échantillon d'UPS avait été mis à jour en 1996. Ainsi, les UPS rurales étaient beaucoup plus actualisées en termes de mesures de la taille que le recensement de 1991. Toutefois, l'EDS avait pour objectifs de mesure de fournir des données au niveau national, ce qui signifie que les UPS échantillonnées étaient distribuées proportionnellement à travers le pays. De ce fait, il n'y avait que très peu d'UPS urbaines disponibles à partir de l'EDS, et ne pouvaient donc pas être utilisées pour la MICS, dont un objectif majeur était de fournir des données fiables pour le milieu urbain comme une estimation clé de domaine. En conséquence, l'EDS fut écartée des options possibles.

Couverture de l'Echantillon

Certaines zones de la Papouasie Nouvelle Guinée, dont la plupart situées à l'intérieur du pays, ne sont accessibles que par hélicoptère. Le Gouvernement se rendit compte que la couverture par l'enquête des populations de ces zones serait excessivement onéreuse. Par conséquent, le Gouvernement décida d'exclure les zones inaccessibles de la MICS, même s'il était reconnu que la situation sanitaire des femmes et des enfants dans ces zones pouvait être assez différente de celle du reste du pays.

En vue de réaliser une méthodologie d'enquête valable, il est important d'éliminer les zones exclues de la base de sondage avant la sélection de l'échantillon. De cette façon, l'échantillon qui est tiré peut encore être un échantillon probabiliste de la population *couverte*, même si cette dernière n'a pas une couverture nationale. La procédure alternative consistant à sélectionner d'abord les zones échantillonnées, ensuite à procéder à des remplacements à chaque fois qu'une

zone exclue est tirée est une procédure statistiquement biaisée, et elle ne fut pas utilisée. En conséquence, une liste des districts et des sous-districts inaccessibles qui devaient être exclus de la couverture de l'enquête fut élaborée, et toutes les ZD composant ces zones furent éliminées de la base de sondage.

Taille et Allocation de l'Echantillon

Le Gouvernement souhaitait réaliser la MICS avec un échantillon d'environ 2.500 ménages. Le manuel de la MICS 2000 exigeait des pays de trouver une taille d'échantillon qui leur permettrait de mesurer leurs indicateurs les plus importants, avec une marge d'erreur de 3-5 points de pourcentage (95% de niveau de confiance). Au niveau national, il était effectivement attendu que 2.500 ménages seraient suffisants pour remplir ces exigences. Cependant, le Gouvernement désirait également disposer d'estimations distinctes et fiables pour certains domaines, autrement dit, des zones sous-nationales de grande importance pour la planification stratégique. Il s'agit des zones urbaines et rurales, en plus de quatre régions : le sud, le nord, les "Highlands" et les "Islands".

Pour les estimations régionales, en particulier, une taille d'échantillon de 2.500 fut jugée insuffisante, car il n'y aurait eu en moyenne qu'environ 625 ménages dans chaque échantillon par région – ce qui n'était pas assez pour tirer des estimations très fiables pour les indicateurs clés. Sur ce point, la distribution de la population dans les milieux urbain et rural suggère qu'un échantillonnage disproportionné devait être utilisé afin d'obtenir suffisamment de cas pour la partie urbaine. Cela parce que ce n'est qu'environ 15% de la population de la Papouasie Nouvelle Guinée qui réside dans les zones urbaines. Ainsi, si l'échantillon était sélectionné proportionnellement à la taille, moins de 400 des ménages seraient urbains, et dans seulement 15 UPS, éléments qui sont tous deux trop petits pour permettre l'obtention de résultats fiables.

Les éléments ci-dessus suggèrent deux choses : (1) un suréchantillonnage devrait être utilisé pour le milieu urbain et (2) la taille de l'échantillon devrait être augmentée, globalement, pour produire suffisamment de cas permettant de tirer des estimations régionales fiables.

Le nombre de grappes – autrement dit, les unités de sondage au premier degré (UPS) – est également un facteur à prendre en compte pour les estimations des domaines. Avec 100 UPS, le nombre moyen pour une région donnée ne serait que de 25. Ce qui est trop léger pour donner une "dispersion" adéquate de l'échantillon au niveau régional. Il aurait dû y avoir environ 50 UPS pour chaque domaine, mais le minimum absolu ne devrait pas être en dessous de 30. En conséquence, il fut décidé que la taille globale de l'échantillon serait augmentée – passant de 2.500 à 4.000 ménages. Il fut ensuite décidé de prévoir un supplément de 10% pour les non-réponses, ce qui fit passer la taille de l'échantillon à environ 4,375 ménages. Le nombre de grappes fut alors augmenté – passant de 100 à 175 ($4.375 \div 25$).

En fin de compte, il fut décidé que les strates urbaines seraient suréchantillonnées, par un facteur d'environ 3, autrement dit, une augmentation de la proportion de l'ensemble de l'échantillon urbain de 15% à 43%. Ce qui donnait environ 1.875 ménages dans 75 UPS pour les strates

urbaines et 2.500 ménages dans 100 UPS pour les strates rurales. Ce suréchantillonnage dans les zones urbaines était nécessaire compte tenu de l'impératif de fournir suffisamment d'unités d'échantillon pour disposer d'estimations fiables. Cependant, l'échantillon disproportionné dans les strates urbaines avait néanmoins laissé environ 2.500 ménages dans l'échantillon des strates rurales, ce qui était plus qu'approprié pour tirer des chiffres fiables.

Mise en œuvre – Procédures de Sélection de l'Echantillon

En Papouasie Nouvelle Guinée, les procédures de sélection de l'échantillon étaient très semblables de celles du Liban – mais avec tout de même une différence de taille. Aussi, nous ne présenterons ici que les points sur lesquels elles diffèrent. Deux options pourraient être considérées, selon le manuel de la MICS (toutes les deux éditions), lorsque l'on décide de la méthodologie à utiliser pour la sélection des ménages au second (ou au troisième) degré. L'une est de procéder à une nouvelle énumération des ménages dans les grappes échantillonnées et de sélectionner les ménages à partir de ces nouvelles fiches d'énumération. Tel que stipulé dans le manuel, c'est l'approche préférée pour la MICS, à la fois parce qu'elle est sensiblement plus fiable et quelque peu moins compliquée que l'autre alternative.

En ce qui concerne l'autre alternative, cependant, il s'agit d'une approche qui a l'avantage de faire l'économie d'une nouvelle énumération, qui peut être très onéreuse, principale raison pour laquelle elle est présentée dans le manuel comme une option pratique. Cette option entraîne la construction de segments géographiques de taille approximativement égale dans chaque grappe échantillonnée (les ZD pour la Papouasie Nouvelle Guinée), la sélection aléatoire de l'un d'eux et l'interview de tous les ménages localisés dans les segments choisis pour la MICS. Comme cela a été signalé ci-dessus, toutefois, cette alternative, quoique valable, n'est pas aussi fiable que l'approche d'énumération, car les ménages de l'échantillon sont "grappés" ensemble et l'effet du plan de sondage est donc quelque peu plus important. Deuxièmement, la segmentation est moins compliquée que l'énumération et peut donc introduire une plus grande non-erreur de sondage, surtout si le personnel n'est pas bien entraîné à l'utilisation correcte des cartes, y compris les croquis des segments. Le Bureau National de la Statistique décida d'opter pour l'approche préférée – l'énumération plutôt que la segmentation.

ANGOLA

L'Enquête par Grappes à Indicateurs Multiples fut mise en œuvre par l'Institut National de la Statistique (*Instituto Nacional de Estatística*) de l'Angola. Le plan de sondage pour la MICS dans ce pays illustre une exécution d'enquêtes faite en deux phases, séparées de plusieurs mois l'une de l'autre, ainsi que l'association des aspects liés à la couverture de la population, aux défaillances de la base et à la présentation des résultats aux utilisateurs.

Résumé du Plan de Sondage

Les principaux objectifs de mesure de la MICS étaient : (1) de recueillir les indicateurs appropriés du Sommet Mondial pour l'Enfance de la fin de la décennie, en vue de cerner la situation des femmes et des enfants en Angola, et (2) de fournir ces estimations pour le niveau national et pour trois zones écologiques (domaines) – le littoral, le planalto et l'intérieur.

La méthodologie d'échantillonnage pour la MICS 2000 en Angola a utilisé des procédures de sélection probabilistes pour choisir environ 6.660 ménages selon un plan stratifié par grappe. L'enquête a été réalisée en deux phases, la première utilisant une base de sondage développée pour l'Enquête sur les Revenus et les Dépenses de 2000-2001 (*Inquerito sobre Despesas e Receitas, ou IDR*). Quant à la deuxième phase, qui fut réalisée approximativement 2-3 mois plus tard, elle eut pour champ d'intervention l'environnement habituel de l'enquête MICS, autrement dit toutes les zones qui n'avaient pas été prises en compte lors de l'IDR. Cette dernière phase fut retardée parce que les documents de travail devant servir comme base de sondage devaient être élaborés au préalable.

Sélection de l'Echantillon

La sélection de l'échantillon a suivi un processus de sélection à deux degrés dans les zones urbaines et à trois degrés dans les zones rurales. Les unités de sondage au premier degré (autrement dit les unités primaires de sondage – UPS) ont été définies comme des '*seccas censitarias*' dans les zones urbaines et elles consistaient en des segments définies géographiquement – habituellement des blocs municipaux – comprenant en moyenne environ 100 ménages. Dans les zones rurales, les grappes étaient les communes, et les unités secondaires étaient les villages. Ainsi, les grappes de la MICS elles-mêmes constituaient les unités de sondage au premier degré dans les zones urbaines, mais les unités de sondage au second degré dans les zones rurales.

Les grappes furent sélectionnées dans les zones dites "couvertes" du pays. Celles-ci étaient définies comme des zones contrôlées par le gouvernement et accessibles, et il s'agissait généralement des capitales provinciales et/ou municipales, en plus de quelques zones rurales. La couverture de l'échantillon a été effectuée en retirant au préalable des bases de sondage toutes les zones qui avaient été exclues du fait de problèmes de sécurité interne et/ou d'accessibilité, y compris les zones qui étaient supposées avoir été minées. De plus, les personnes déplacées vivant dans des camps ne furent pas couvertes par l'enquête principale, bien qu'il y ait un plan pour les échantillonner et les interviewer dans le cadre d'une opération distincte.

La sélection de l'échantillon des unités de sondage au premier degré fut faite avec probabilité proportionnelle à leur taille (ou *ppt*). Les mesures de la taille, m.d.t., furent alors les *décomptes* actuels des maisons ou des autres logements dans les zones urbaines, et les *estimations* actuelles de la population rurale dans les zones rurales. Les décomptes furent obtenus par un sondage effectif sur le terrain, alors que les estimations furent tirées des informations locales fournies par le niveau provincial.

Au total, 408 grappes furent sélectionnées – 300 lors de la première phase, tirées à partir de la base de sondage de l'IDR et 108 lors de la deuxième phase, et issues des zones restantes. Les procédures d'échantillonnage ne furent pas les mêmes lors de ces deux phases. Dans la première phase, les grappes de la MICS furent un sous-ensemble systématique de celles déjà sélectionnées pour l'IDR. Ces grappes avaient déjà été énumérées de nouveau, autrement dit, une nouvelle énumération des ménages avait été préparée lors d'une opération de terrain pour l'IDR. Cette énumération a constituée la base de sondage pour la dernière étape de sélection. Une sélection systématique de 15 ménages fut effectuée dans chacune des 300 grappes de la MICS lors de la première phase, indépendamment des ménages

qui avaient été choisis pour l'IDR, afin d'éviter des chevauchements entre les interviews des deux enquêtes. Vingt ménages furent choisis dans chacune des 108 grappes lors de la deuxième phase²².

Dans la deuxième phase, la base de sondage devait être développée de la même manière qu'elle l'avait été pour l'IDR, à la seule différence qu'elle fut étendue aux provinces et aux autres zones MICS qui n'avaient pas été couvertes par l'IDR. C'est cette base de la deuxième phase qui a constitué la source pour la sélection des 108 grappes. Elles furent sélectionnées avec probabilité proportionnelle à la taille. Les grappes sélectionnées furent ensuite énumérées sur le terrain, afin que la sélection au second degré de 20 ménages par grappe puisse être effectuée. Celle-ci fut faite en utilisant la sélection systématique.

L'échantillon des 6.660 ménages fut distribué de manière proportionnelle entre les trois zones écologiques ou domaines. Chacun des domaines était suffisamment large, en termes de population, de sorte que des résultats fiables pouvaient être tirés au niveau des domaines, sur la base d'une allocation proportionnelle de l'échantillon. Les non-réponses avaient été estimés à environ 10% dans cette enquête, cas de figure qui aurait donné environ 6.000 comme nombre de ménages effectivement interviewés dans l'enquête.

Il n'était pas possible d'estimer avec précision le pourcentage de la population à couvrir par l'enquête, à cause de l'indisponibilité de données démographiques actuelles. On pensait, néanmoins, que probablement 10% de la population résidaient dans les camps de personnes déplacées, et qu'entre 15 et 25% résidaient dans des zones inaccessibles. Ainsi, il fut estimé qu'environ 65-75% de couverture de la population nationale serait atteinte par la MICS (voir section suivante).

Base de Sondage et Couverture

En pratique, les bases de sondage parfaites sont rarement trouvées, mais il fut particulièrement difficile pour l'Angola de satisfaire les trois critères d'exhaustivité (ou *complétude*), de précision et d'actualisation. Cela s'explique par plusieurs raisons. L'une est que c'est le dernier recensement qui est la base habituellement utilisée pour les échantillons d'enquêtes nationales, or celui-ci datait de 1970 dans le cas de l'Angola, ce qui était trop ancien pour être applicable. Aussi, d'autres bases plus récentes devaient être utilisées – celles-ci seraient probablement quelque peu inférieures à une base de recensement actuelle (voir la présentation détaillée ci-dessous).

Deuxièmement, tel que susmentionné, les zones minées avaient été exclues de la couverture, ainsi que les camps de personnes déplacées et toute autre zone inaccessible. Il en a résulté une violation du critère d'exhaustivité, et cela sans que l'on en connaisse l'ampleur, bien que l'on ait supposé que la couverture nationale serait d'environ 65-75%. En troisième lieu, les bases disponibles comportaient des inexactitudes dont on ignorait l'importance, surtout dans les zones rurales où l'on pensait que la qualité des estimations de population (qui avaient servi à établir les probabilités de sélection au premier degré) était très variable d'une province à une autre.

Les implications pour la MICS 2000 de l'Angola étaient que les résultats de l'enquête allaient être biaisés, et cela à un degré inconnu, *si les résultats avaient été présentés comme*

²² Les tailles des grappes entre les phases une et deux sont différentes parce que le nombre de ménages par grappe recommandé pour la MICS est de 20 (voir l'édition 2000 du manuel de la MICS), mais il était beaucoup plus pratique sur le plan opérationnel de sélectionner 15 ménages dans la base de l'IDR.

représentatifs du pays tout entier. Cela du fait que la 'représentativité' ne peut être statistiquement justifiée pour les sous-populations exclues de la couverture – sous-populations ayant probablement des caractéristiques quelque peu différentes de celles qui avaient été incluses. Une autre implication majeure est que les estimations de la taille de la population nationale ne pouvaient être tirées de la MICS, en raison de l'absence de couverture des sous-populations exclues.

Néanmoins, des résultats valables et utiles furent produits à partir des populations *couvertes* par la MICS, en appliquant avec précaution des méthodes probabilistes de sélection de l'échantillon des ménages à partir des bases qui furent utilisées, même si ces bases n'étaient pas parfaites. En termes de couverture d'enquête, la base qui fut élaborée par l'Instituto Nacional de Estatística pour l'IDR était excellente dans une certaine mesure. C'était une source valable et dont l'utilisation était défendable pour la phase une. Son seul problème était qu'elle ne couvrait que sept provinces et, à l'intérieur de ces provinces, seulement les espaces urbains habituels. Ce fut la raison pour la raison pour laquelle il s'est avéré nécessaire de développer une nouvelle base pour la phase deux – en vue de couvrir les 11 autres provinces, en plus du maximum de zones rurales qui semblaient accessibles pour l'enquête.

La MICS de l'Angola nous apprend une leçon importante sur la publication des données au public. La publication des résultats de l'enquête à travers des rapports devrait toujours être faite en précisant, de la manière la plus claire et précise possible, la définition de la population effectivement couverte par l'enquête. Dans le cas de l'Angola, il était important de signaler aux utilisateurs que la MICS présentait des résultats scientifiquement défendables pour approximativement 65-75% de la population de l'Angola, mais que ces résultats n'étaient pas représentatifs des personnes vivant en dehors des zones contrôlées par le gouvernement, celles résidant à proximité de certaines zones exclues parce que minées, et les personnes déplacées vivant dans les camps.