

Section 2 : Préparation de l'échantillon

Présentation générale

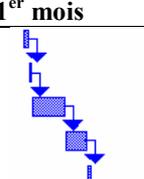
Introduction Cette section traite des principes et des méthodes à appliquer et des tâches à accomplir pour préparer, planifier et sélectionner l'échantillon pour votre enquête STEPS.

Public concerné Cette section s'adresse en premier lieu aux personnes remplissant les fonctions suivantes :

- conseiller en statistiques,
 - coordinateur de site STEPS,
 - membre d'un comité de coordination.
-

Tâches et délais La préparation de l'échantillon s'inscrit dans le processus de planification et de préparation de l'enquête. Elle doit prendre entre deux jours et une semaine en fonction des méthodes choisies et de la disponibilité des informations nécessaires pour déterminer l'échantillon.

Le tableau ci-dessous établit une liste des tâches principales dont il est question dans cette section et indique les délais prévus pour chacune.

Tâche	Durée	1 ^{er} mois
Définir la population cible	1 jour	
Définir la taille de l'échantillon	1 jour	
Définir la base de sondage et le plan d'échantillonnage	1 semaine	
Sélectionner l'échantillon des participants	3 jours	
Documenter la sélection de l'échantillon	1 jour	

Dans cette section Cette section traite des thèmes suivants

Thème	Voir page
Instructions pour l'échantillonnage	2-2-2
Définition de la taille de l'échantillon	2-2-3
Identification de la base de sondage	2-2-10
Choix du plan d'échantillonnage	2-2-12
Sélection de l'échantillon	2-2-20
La méthode Kish	2-2-25
Documentation du plan d'échantillonnage	2-2-27
Préparation des formulaires de collecte des données	2-2-28

Instructions pour l'échantillonnage

Introduction

Les meilleures techniques d'enquête peuvent donner un aperçu fiable des facteurs de risque des maladies chroniques au sein d'une population à partir d'un échantillon de cette population. Pour obtenir de tels résultats, l'échantillon doit être sélectionné de manière scientifique. S'il est correctement déterminé, il sera représentatif de la totalité de la population. Pour obtenir des résultats intéressants et utiles, il est primordial que l'échantillon sélectionné pour votre enquête STEPS soit de très bonne qualité.

Faire correspondre votre échantillon au champ de votre enquête

Pour obtenir un échantillon qui correspond au champ de l'enquête vous devez :

- définir une population cible,
 - sélectionner de manière scientifique un échantillon de population représentatif de la population cible,
 - prévoir de rapporter les résultats de l'enquête par sexe et tranches d'âge de 10 ans.
-

Définir la population cible

Chaque site doit définir la population cible de son enquête STEPS. Pour ce faire, vous devez tenir compte de l'objectif et de l'utilisation qui sera faite des données de l'enquête. Par exemple, avez-vous besoin d'une enquête représentative de toute la population ou de celle d'une région en particulier ?

Pour la surveillance SEPS des facteurs de risque de maladie chronique, il est souhaitable que la population cible soit constituée au moins de l'ensemble des personnes âgées de 25 à 64 ans résidant dans le secteur d'enquête. Il est possible de sélectionner davantage de tranches d'âge mais il est, en revanche, déconseillé de choisir une tranche d'âge plus réduite.

Personnes faisant partie de l'échantillon

Les personnes faisant partie de l'échantillon sont choisies de manière scientifique au sein de la population cible. Après avoir défini la population cible, vous y choisissez les participants.

Estimations par sexe et par tranche d'âge

La prévalence de la plupart des facteurs de risque de maladie chronique tend à augmenter avec l'âge et varie selon le sexe. Il est donc souhaitable que le rapport d'enquête indique des estimations par sexe et par tranche d'âge de 10 ans.

Afin que les données tirées de l'enquête permettent d'obtenir des estimations précises pour chaque tranche d'âge, par sexe, le calcul de la taille de l'échantillon doit tenir compte du nombre total de tranches d'âge pour chaque sexe. La section qui suit explique comment intégrer le nombre total d'estimations souhaitées dans le calcul de la taille de l'échantillon.

Définition de la taille de l'échantillon

Introduction

Pour que les résultats de l'enquête soient suffisamment précis, la taille de l'échantillon déterminé dans la population cible doit être assez importante. Pour calculer la taille de l'échantillon nécessaire, il faut tenir compte des facteurs suivants :

- niveau de confiance souhaité pour les résultats de l'enquête
- marge d'erreur acceptable pour les résultats de l'enquête
- effet du plan d'échantillonnage
- niveaux initiaux estimés des comportements ou des indicateurs à mesurer

En outre, la taille de l'échantillon doit être ajustée en fonction :

- du nombre d'estimations par sexe et par tranche d'âge
 - du nombre anticipé de non-réponses.
-

Terminologie utile

On trouvera dans le tableau ci-dessous la définition de plusieurs termes statistiques fondamentaux. Il est important de bien maîtriser cette terminologie avant de calculer la taille de l'échantillon.

Terme	Définition
Moyenne de l'échantillon / Prévalence dans l'échantillon	Moyenne ou prévalence estimée d'un paramètre donné de population (p. ex. : nombre moyen de jours où des fruits ont été consommés au cours d'une semaine donnée), calculée à partir des données tirées de l'enquête
Moyenne de la population / Prévalence dans la population	Moyenne de la population ou prévalence dans la population d'un paramètre donné pour l'ensemble de la population cible. La moyenne de l'échantillon est une estimation de la moyenne de la population.
Intervalles de confiance	Marge de valeurs autour de la moyenne de l'échantillon ou de la prévalence dans une population donnée. Par exemple, un intervalle de confiance de 95 % indique que, dans 95 enquêtes sur 100, la moyenne de la population sera située dans la marge de valeur autour de la moyenne de l'échantillon.

Suite à la page suivante

Instructions pour l'échantillonnage, suite

Variables utilisées pour calculer la taille de l'échantillon

On trouvera dans le tableau ci-dessous une définition des variables utilisées pour calculer la taille de l'échantillon, ainsi que les valeurs recommandées pour chaque variable.

Variable	Définition	Valeur recommandée
Niveau de confiance	<ul style="list-style-type: none"> • Valeur de probabilité associée à un intervalle de confiance donné. • Désigne le niveau d'incertitude pour la moyenne ou la prévalence de l'échantillon, sous la forme d'une estimation de la moyenne de la population ou de la prévalence dans la population. • Plus le niveau de confiance est élevé, plus l'échantillon doit être grand. 	<ul style="list-style-type: none"> • 1,96 • NB : 1,96 est la valeur de probabilité associée à un intervalle de confiance de 95 %.
Marge d'erreur	<ul style="list-style-type: none"> • Demi-intervalle de confiance attendu. • Plus la marge d'erreur est étroite, plus l'échantillon doit être grand. 	<ul style="list-style-type: none"> • 0,05
Effet du plan d'échantillonnage	<ul style="list-style-type: none"> • Désigne une perte d'efficacité de l'échantillonnage due à la complexité du plan d'échantillonnage. • Pour un échantillon aléatoire simple, l'effet du plan d'échantillonnage est de 1,00. Pour les plans d'échantillonnage plus complexes, il faut constituer un échantillon plus grand s'il on veut que les résultats de l'enquête soient aussi précis qu'avec un échantillon aléatoire simple. Donc, plus le plan d'échantillonnage est complexe, plus l'effet qu'il induit est important. 	<ul style="list-style-type: none"> • 1,50 • Note : 1,50 est la valeur recommandée pour la plupart des enquêtes STEPS dont le plan d'échantillonnage est complexe. Si l'on dispose d'informations sur l'effet du plan d'échantillonnage pour des enquêtes précédentes, dont le plan était similaire à celui proposé pour l'enquête STEPS, il est conseillé d'utiliser ces estimations.
Niveaux initiaux estimés des comportements ou des indicateurs à mesurer	<ul style="list-style-type: none"> • Prévalence estimée des facteurs de risque dans la population cible. • Les valeurs les plus proches de 50 % sont celles qui offrent la meilleure fiabilité, et exigent de disposer d'un échantillon le plus grand possible. 	<ul style="list-style-type: none"> • 0,50 si aucune donnée antérieure sur la population cible n'est disponible. • La valeur la plus proche de 0,50 si on dispose de données antérieures sur la population cible.

Suite à la page suivante

Instructions pour l'échantillonnage, suite

Équation permettant de calculer la taille de l'échantillon

L'équation permettant de calculer la taille de l'échantillon est la suivante :

$$n = Z^2 \frac{P(1-P)}{e^2}$$

où :

- Z est le niveau de confiance
 - P est le niveau initial des indicateurs
 - e est la marge d'erreur
-

Exemple de calcul

Si l'on suit les recommandations ci-dessus pour chaque paramètre, le calcul **initial** de la taille de l'échantillon est le suivant :

$$n = 1,96^2 \frac{0,5(1-0,5)}{0,05^2} = 384$$

Le résultat **doit** toutefois être ajusté en fonction de l'effet du plan d'échantillonnage, du nombre d'estimations par tranche d'âge et par sexe à rapporter et du nombre attendu de non-réponses.

Ajustement par rapport à l'effet du plan d'échantillonnage

Pour tenir compte de l'effet du plan d'échantillonnage, il suffit de **multiplier** la taille de l'échantillon par l'effet du plan d'échantillonnage.

Ajustement par rapport au nombre d'estimations par tranche d'âge et par sexe

Comme indiqué dans la section précédente, il est recommandé de ventiler les résultats de l'enquête par tranche d'âge de 10 ans, pour chaque sexe. Pour que les estimations par tranche d'âge et par sexe soient suffisamment précises, il faut **multiplier** la taille de l'échantillon par le nombre de tranches d'âge pour chaque sexe.

Le nombre d'estimations par tranche d'âge et par sexe variera selon la tranche d'âge cible de l'enquête. Pour les enquêtes couvrant les 25-64 ans, le nombre d'estimations par tranche d'âge et par sexe est **8**.

Pour chaque tranche d'âge de 10 ans supplémentaire, il faut ajouter 2 estimations par âge et par sexe. Par exemple, si on veut ajouter la tranche d'âge 15-24 ans à l'enquête, le nombre total d'estimations sera de 10.

Suite à la page suivante

Instructions pour l'échantillonnage, suite

Ajustement par rapport au taux de non-réponse attendu

Pour tenir compte du nombre attendu de non réponses, il faut **diviser** la taille de l'échantillon par le **taux de non réponse** attendu.

Il est conseillé de prévoir un taux de non réponse de 20 %. C'est une estimation raisonnable basée sur les taux de réponse aux enquêtes STEPS précédentes. Si, sur votre site, les taux de réponse ont été constamment plus élevés pour des enquêtes similaires auprès des ménages, vous pouvez prévoir un taux de non réponse plus faible, par exemple de 10 %.

Exemple : Si le taux de non réponse prévu est de 20 %, diviser la taille de l'échantillon par 0,80.

Synthèse du calcul de la taille de l'échantillon

Le tableau ci-dessous récapitule les différentes étapes du calcul de la taille de l'échantillon.

Étape	Description
1	Déterminer la valeur de toutes les variables nécessaires pour calculer la taille de l'échantillon.
2	Poser l'équation indiquée ci-dessus en utilisant le niveau de confiance, la marge d'erreur et le niveau initial des indicateurs afin d'obtenir une estimation de n (taille de l'échantillon).
3	Multiplier n par l'effet du plan d'échantillonnage et par le nombre d'estimations par tranche d'âge et par sexe.
4	Pour connaître la taille définitive de l'échantillon, diviser le résultat obtenu à l'étape 3 par le taux attendu de non réponse.

Exemple de calcul de la taille d'un échantillon

On utilisera dans cet exemple les valeurs recommandées pour tous les paramètres de l'équation servant à calculer la taille de l'échantillon. Le calcul initial est donc le suivant :

$$n = 1,96^2 * \frac{0,5 (1-0,5)}{0,05^2} = 384$$

On multiplie ensuite ce n initial par l'effet du plan d'échantillonnage (1,5) et les 8 estimations par âge et par sexe souhaitées pour les résultats de l'enquête :

$$n = 384 * 1,5 * 8 = 4\ 608$$

Enfin, on divise n par 0,80 pour tenir d'un taux prévu de non réponse de 20 % :

$$n = 4\ 608 \div 0,80 = 5\ 760$$

La taille définitive de l'échantillon est de 5 760 personnes.

Suite à la page suivante

Instructions pour l'échantillonnage, suite

Échantillonnage de populations très petites

Lorsque la population cible est très petite (<50 000 personnes), on peut réduire la taille de l'échantillon à l'aide d'une correction d'échantillonnage pour population finie. La procédure décrite ci-dessous permet de vérifier que cette correction est possible sur votre site et de l'appliquer pour réduire la taille de votre échantillon.

Étape	Description										
1	Afin d'obtenir le n pour chaque estimation, n'effectuez que les étapes 1 et 2 indiquées au tableau précédent.										
2	<p>Calculez la taille de la population cible pour chaque estimation en vous basant sur les données démographiques disponibles ou une source similaire fiable.</p> <p>Exemple : Si l'on veut obtenir 8 estimations par tranche d'âge de dix ans, par sexe, il faut calculer le nombre d'individus dans chacune de ces tranches (p. ex. : le nombre d'hommes dans la tranche des 25-34 ans).</p>										
3	<p>N'utiliser la correction d'échantillonnage pour population finie que lorsque l'échantillon à déterminer représente plus de 10 % de la population cible. Donc, pour chaque estimation, il faut diviser le n calculé à la première étape par la population cible afin de vérifier que la correction d'échantillonnage pour population finie peut être appliquée.</p> <p>Exemple : On souhaite obtenir huit estimations par tranches d'âge de 10 ans pour une population n de 384 personnes. Le tableau ci-dessous indique les données recueillies pour les quatre premières estimations.</p> <table border="1" data-bbox="539 1317 1315 1509"> <thead> <tr> <th>Estimations souhaitées</th> <th>Taille de la population cible</th> </tr> </thead> <tbody> <tr> <td>Hommes, 25-34 ans</td> <td>2548</td> </tr> <tr> <td>Femmes, 25-34 ans</td> <td>2641</td> </tr> <tr> <td>Hommes, 35-44 ans</td> <td>3465</td> </tr> <tr> <td>Femmes, 45-54 ans</td> <td>3356</td> </tr> </tbody> </table> <p>Diviser n par la population cible pour chaque estimation :</p> <p>$384/2548 = 0,15$ $384/2641 = 0,15$ $384/3465 = 0,11$ $384/3356 = 0,11$</p>	Estimations souhaitées	Taille de la population cible	Hommes, 25-34 ans	2548	Femmes, 25-34 ans	2641	Hommes, 35-44 ans	3465	Femmes, 45-54 ans	3356
Estimations souhaitées	Taille de la population cible										
Hommes, 25-34 ans	2548										
Femmes, 25-34 ans	2641										
Hommes, 35-44 ans	3465										
Femmes, 45-54 ans	3356										
4	<p>Si la plupart ou la totalité des quotients obtenus à l'étape 3 sont égaux ou supérieurs à 0,10, on peut appliquer la correction d'échantillonnage pour population finie (passer à l'étape suivante). Dans le cas contraire, revenir à l'étape 3 du tableau précédent et déterminer la taille totale de l'échantillon à partir du n déjà calculé.</p>										

Suite à la page suivante

Instructions pour l'échantillonnage, suite

Échantillonnage de populations très petites (suite)

Étape	Description
5	<p>Pour chaque estimation, appliquer la correction d'échantillonnage pour population finie à n en posant l'équation suivante :</p> $\text{nouveau } n = \frac{n}{1 + \frac{n}{\text{population}}}$ <p>où « population » s'entend de la population cible pour une estimation donnée et pas de l'ensemble de la population cible.</p>
6	Additionner tous les « nouveaux n » et multiplier la somme par l'effet du plan d'échantillonnage.
7	Pour calculer la taille définitive de l'échantillon, diviser le résultat obtenu à l'étape 6 par le taux de réponse attendu.

Modifications supplémentaires de la taille de l'échantillon

Dans plusieurs situations, on peut être amené à ajuster la taille de l'échantillon déterminée grâce aux calculs indiqués ci-dessus. Le tableau ci-dessous décrit certaines de ces situations et donne des indications pour ajuster la taille de l'échantillon. Si votre situation n'est pas mentionnée ici ou si vous avez besoin d'aide, veuillez contacter l'équipe STEPS.

Si ...	Alors ...						
On a besoin de données sur des sous-groupes spécifiques (p. ex. groupes ethniques, populations urbaines ou rurales).	Il y a deux manières de procéder selon les informations souhaitées :						
	<table border="1"> <thead> <tr> <th>Si ...</th> <th>Alors ...</th> </tr> </thead> <tbody> <tr> <td>Les données doivent être rapportées pour l'ensemble des individus de chaque sous-groupe.</td> <td>Prendre comme nombre d'estimations le plus grand des deux chiffres suivants : <ul style="list-style-type: none"> le nombre d'estimations par tranche d'âge et par sexe souhaitées le nombre de nouveaux sous-groupes </td> </tr> <tr> <td>Dans chaque sous-groupe, les données doivent être rapportées par tranche d'âge et par sexe.</td> <td>Pour déterminer le nombre total d'estimations, multiplier le nombre de tranches d'âge (par sexe) par le nombre total de nouveaux sous-groupes (p. ex. de groupes ethniques).</td> </tr> </tbody> </table>	Si ...	Alors ...	Les données doivent être rapportées pour l'ensemble des individus de chaque sous-groupe.	Prendre comme nombre d'estimations le plus grand des deux chiffres suivants : <ul style="list-style-type: none"> le nombre d'estimations par tranche d'âge et par sexe souhaitées le nombre de nouveaux sous-groupes 	Dans chaque sous-groupe, les données doivent être rapportées par tranche d'âge et par sexe .	Pour déterminer le nombre total d'estimations, multiplier le nombre de tranches d'âge (par sexe) par le nombre total de nouveaux sous-groupes (p. ex. de groupes ethniques).
	Si ...	Alors ...					
Les données doivent être rapportées pour l'ensemble des individus de chaque sous-groupe.	Prendre comme nombre d'estimations le plus grand des deux chiffres suivants : <ul style="list-style-type: none"> le nombre d'estimations par tranche d'âge et par sexe souhaitées le nombre de nouveaux sous-groupes 						
Dans chaque sous-groupe, les données doivent être rapportées par tranche d'âge et par sexe .	Pour déterminer le nombre total d'estimations, multiplier le nombre de tranches d'âge (par sexe) par le nombre total de nouveaux sous-groupes (p. ex. de groupes ethniques).						
Note : Lorsqu'on répartit l'échantillon, il est important de tenir compte de ces sous-groupes afin de disposer d'un nombre suffisant de participants dans chaque sous-groupe (voir section suivante).							

Suite à la page suivante

Instructions pour l'échantillonnage, suite

Modifications supplémentaires de la taille de l'échantillon (suite)

Si ...	Alors ...
Une surreprésentation des sous-populations très petites est souhaitée.	Accroître le n global en augmentant de 10 % le n pour les estimations concernées.
Une surreprésentation de certaines sous-populations où le taux de non réponse est plus élevé que la moyenne est souhaitée.	Accroître le n global en augmentant de 10 % à 20 % le n pour les estimations concernées.
Une surreprésentation de la tranche d'âge 55-64 ans est souhaitée car on pense qu'il sera difficile d'obtenir un nombre suffisant de personnes interrogées dans cette tranche d'âge en raison d'un taux élevé de non réponse et/ou de la petite taille de cette sous-population.	Accroître le n global en augmentant de 10 % à 20 % les estimations concernées pour les hommes et les femmes de cette tranche d'âge. Pour savoir comment surreprésenter les 55-64 ans dans les ménages, voir la méthode Kish, à partir de la p. 2-2-25.

Note : Si l'on souhaite recourir à la surreprésentation, il faut généralement procéder aussi à des ajustements au moment de la répartition de l'échantillon (voir section suivante) : bien souvent, il faut non seulement augmenter l'échantillon mais aussi tenir compte des groupes difficiles à atteindre, dont la proportion dans l'échantillon doit être plus importante.

Échantillons plus petits

Il est tentant d'envisager d'interviewer moins de gens pour réduire les coûts. Aux Steps 1 et 2, la réduction de la taille de l'échantillon **n'est pas** recommandée car elle rendrait beaucoup moins précises les estimations en population pour les divers facteurs de risque.

Si, au Step 3, il est économiquement impossible de travailler sur l'ensemble de l'échantillon, vous pouvez prendre un sous-échantillon de personnes ayant participé aux Steps 1 et 2, tout en sachant que cela diminuera la précision des estimations en population. En cas de sous-échantillonnage au Step 3, le sous-échantillon doit être au moins égal à 20 % de l'échantillon total.

Identification de la base de sondage

Introduction

La base de sondage est une liste d'unités ou d'éléments qui définissent la population cible. C'est à partir de cette liste que l'échantillon est déterminé. Une base de sondage est capitale pour toutes les enquêtes.

Trouver des bases de sondage disponibles

Pour trouver les bases de sondage disponibles et déterminer laquelle correspond le mieux à votre site, cherchez des listes, des bases de données, des registres ou d'autres sources à jour qui couvrent correctement la population que vous voulez étudier. (Recherchez par exemple les registres de population ou les listes de recensement).

Il convient de contacter plusieurs services gouvernementaux et organes nationaux pour établir quelles bases existent dans votre pays et, le cas échéant, s'il est possible d'y avoir accès pour STEPS.

Secteurs de dénombrement (SD)

La base de sondage s'appuie le plus souvent sur des secteurs de dénombrement (SD), c'est-à-dire des zones géographiques – de taille petite à moyenne – définies lors d'un recensement préalable. La plupart des pays disposent de ces informations, qu'il est généralement préférable d'incorporer dans la base de sondage.

Facteurs à prendre en compte

Une base de sondage (ou plusieurs bases de sondage) devrait couvrir l'ensemble de la population (ou des éléments, ou des unités) dans le site étudié. Si cette population est correctement couverte, cela signifie que chaque membre éligible a une chance de faire partie de l'échantillon.

Il faut tenir compte de la représentativité de tous les sous-groupes au moment de décider quelles bases seront utilisées. Il est important de prendre en compte qu'un groupe d'âge, un des sexes ou un groupe ethnique, voire une zone géographique, aient des chances plus ou moins grandes d'être inclus dans la base de sondage. Des biais peuvent apparaître si certains groupes sont moins bien représentés.

Bases de sondage multiples

En raison de contraintes financières et logistiques, on recourt, pour la plupart des enquêtes nationales, à un échantillonnage à plusieurs degrés, technique exposée en détail dans la section suivante. Pour déterminer un échantillon à plusieurs degrés, il faut une base de sondage pour chaque degré.

Suite à la page suivante

Identification de la base de sondage, suite

Caractéristiques d'une base de sondage fiable

Voici quelques caractéristiques d'une base de sondage fiable :

- elle ne contient pas de doubles, ou, dans le cas contraire, ils peuvent être facilement repérés et supprimés ;
- elle ne contient pas de vides, par exemple des maisons inoccupées ou des individus décédés ;
- elle contient des informations qui permettent à toutes les unités d'être distinguées les unes des autres et facilement localisables (par exemple par une adresse complète) ;
- elle contient au moins des informations sur le nombre de ménages ou le nombre total d'individus ;
- elle peut être mise à disposition de l'équipe STEPS dans un délai et à un prix raisonnables.

Note : Les bases de sondage doivent être évaluées au regard de toutes les caractéristiques ci-dessus, mais surtout au regard de l'**exhaustivité** et du **biais potentiel**.

Choix du plan d'échantillonnage

Introduction Le choix du plan d'échantillonnage dépend de plusieurs facteurs, dont il faut tenir compte, et surtout de la taille de la population, de la géographie de la région à couvrir et des ressources disponibles pour effectuer l'enquête.

Stratification La stratification consiste à diviser la base de sondage en sous-groupes ou strates s'excluant mutuellement. L'échantillon est ensuite déterminé proportionnellement ou non à partir de **toutes** les strates. La population cible est stratifiée en fonction des informations disponibles concernant la base de sondage et des renseignements que l'on souhaite obtenir à partir des résultats de l'enquête.

Les strates correspondent souvent à la localisation géographique des unités d'échantillonnage, par exemple :

- secteurs de dénombrement (SD) ou autres régions géographiques bien définies
- régions urbaines ou rurales.

Les strates correspondent parfois aux caractéristiques des individus inclus dans la base de sondage. Ceci est moins fréquent dans les grandes enquêtes nationales en raison de l'absence de données précises sur l'ensemble des individus composant la population cible et de la difficulté de constituer une base de sondage pour chaque strate. Les strates peuvent, par exemple correspondre à :

- l'origine ethnique
- le statut socioéconomique
- le sexe.

La stratification n'est pas obligatoire mais est recommandée pour les raisons suivantes :

- plus grande précision des estimations de l'enquête
- couverture assurée de toutes les strates
- facilité de gestion.

La stratification peut être appliquée parallèlement à d'autres stratégies d'échantillonnage, parmi lesquelles l'échantillonnage aléatoire simple et l'échantillonnage en grappes, abordés plus loin dans cette section.

Stratification et répartition de l'échantillon Si on a décidé de stratifier la population, il faut opter soit pour l'échantillonnage proportionnel à partir de toutes les strates soit pour l'échantillonnage d'une proportion importante d'individus à partir de quelques strates et une proportion plus réduite d'individus à partir d'autres strates (répartition non proportionnelle).

Suite à la page suivante

Choix du plan d'échantillonnage, suite

Stratification et répartition de l'échantillon (suite)

La répartition proportionnelle consiste à échantillonner la même proportion d'individus dans chaque strate afin que la répartition soit similaire dans l'échantillon et dans la population cible. Ce type de répartition est approprié pour les enquêtes dont les données ne seront rapportées que toutes strates confondues.

La répartition non proportionnelle consiste à échantillonner certaines strates plus que d'autres, souvent en déterminant un échantillon de même taille dans toutes les strates. Ce type de répartition est approprié pour les enquêtes lorsqu'on souhaite obtenir des résultats pour chaque strate. Dans cette situation, il faut en général disposer d'un échantillon plus grand pour que les estimations par strate soient suffisamment précises. Le principal écueil de cette méthode est une perte d'efficacité des estimations toutes strates confondues.

Note : Lorsque les strates sont très petites, on peut choisir la répartition proportionnelle mais il faut recourir à la surreprésentation.

Exemple de répartition proportionnelle

Vous trouverez ci-dessous un exemple de répartition proportionnelle, car c'est plus probablement à cette formule que l'on aura recours pour une enquête STEPS.

Dans cet exemple, l'échantillon calculé est de 3000 personnes. La population cible a été divisée suivant les 4 districts administratifs du pays, qui serviront de strates. Le tableau ci-dessous indique la population cible à l'intérieur de chaque strate et la proportion de chaque strate dans la population cible totale.

Strate	Population cible	Proportion de la population
District 1	25 955	0,24 ←
District 2	30 568	0,28
District 3	32 578	0,30
District 4	19 054	0,18
Total	108 155	1,00

$= 25\,955 \div 108\,155$

Pour calculer le nombre d'individus à prendre dans chaque strate, multiplier la taille de l'échantillon par la proportion de chaque strate.

Strate	Population cible	Proportion de la population	Échantillon
District 1	25 955	0,24 ←	720 ←
District 2	30 568	0,28	840
District 3	32 578	0,30	900
District 4	19 054	0,18	540
Total	108 155	1,00	3 000

$= 0,24 \times 3\,000$

Choix du plan d'échantillonnage, suite

Échantillonnage aléatoire simple

Dans de rares cas, il est possible de procéder à un échantillonnage aléatoire simple. Pour les enquêtes auprès des ménages, il faut généralement respecter les conditions suivantes :

- population cible de petite taille,
- secteur d'enquête peu étendu, que les ressources disponibles permettent de couvrir entièrement,
- existence d'une base de sondage détaillée, énumérant au moins l'ensemble des ménages du secteur d'enquête ou, mieux encore, tous les individus éligibles dans le secteur d'enquête.

On peut combiner échantillonnage aléatoire simple et stratification. En cas d'échantillonnage aléatoire stratifié, on commence par stratifier la population puis on détermine un échantillon aléatoire dans chaque strate.

Note : Si vous estimez qu'il est possible de procéder à un échantillonnage aléatoire simple ou stratifié sur votre site, vous pouvez utiliser un échantillon plus petit. Lors du calcul de la taille de l'échantillon, il faut appliquer un effet du plan d'échantillonnage de 1.

Échantillonnage à plusieurs degrés en grappes

L'échantillonnage à plusieurs degrés en grappes est l'une des techniques d'échantillonnage les plus courantes pour les enquêtes nationales et la méthode recommandée pour la plupart des enquêtes STEPS.

« À plusieurs degrés » signifie que l'échantillonnage est réalisé en plusieurs étapes : on choisit d'abord des unités d'échantillonnage puis, à l'intérieur de celles-ci, des unités plus petites. « En grappes » signifie que les unités d'échantillonnage sont subdivisées en groupes s'excluant mutuellement et, à la différence de la stratification, on ne sélectionne pour l'enquête qu'un **échantillon** de ces groupes.

Pourquoi recourir à l'échantillonnage à plusieurs degrés en grappes ?

Le tableau ci-dessous présente les deux principaux problèmes très couramment rencontrés dans les enquêtes nationales, que l'échantillonnage à plusieurs degrés en grappes permet de résoudre.

Problème	Solution
On ne dispose pas d'informations détaillées sur l'ensemble des ménages ou des individus composant la population de l'échantillon et il est impossible d'établir une base de sondage détaillée pour tout le secteur d'enquête.	L'échantillonnage à plusieurs degrés en grappes permet de choisir des unités d'échantillonnage plus grandes (p. ex. des villages) pour lesquelles on a besoin d'informations moins détaillées sur la population cible. C'est seulement au dernier stade de l'échantillonnage (le plus souvent au moment de la sélection des ménages) qu'on a besoin d'informations détaillées. Cependant, comme on ne sélectionne que certaines grappes à chaque étape, on n'a besoin d'une base détaillée que pour un sous-ensemble de la population cible.

Suite à la page suivante.

Choix du plan d'échantillonnage, suite

Pourquoi recourir à l'échantillonnage à plusieurs degrés en grappes ? (suite)

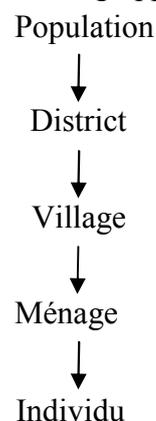
Problème	Solution
Le secteur d'enquête est trop vaste et/ou le coût des voyages est trop élevé pour déterminer un échantillon dans l'ensemble du pays ou des régions objet de l'enquête.	L'échantillon n'étant prélevé que dans certaines grappes, l'échantillonnage à plusieurs degrés en grappes permet d'enquêter dans un secteur limité tout en préservant la représentativité de l'échantillon au niveau national (ou local). Note : L'échantillonnage à plusieurs degrés en grappes, s'il n'est pas fait correctement, ne garantit pas la représentativité de l'échantillon. À chaque degré, les grappes doivent être déterminées et choisies avec soin et de manière cohérente. Cette procédure doit être consignée par écrit en détail.

Préparation d'un échantillon à plusieurs degrés en grappes

Pour procéder à un échantillonnage à plusieurs degrés en grappes, il faut diviser la population en grappes dont chacune sera composée de grappes plus petites ou, au dernier stade, de ménages ou d'individus.

Le schéma ci-contre montre les degrés que l'on pourrait définir pour un site.

Le premier degré comprend le plus souvent les secteurs de dénombrement (SD) définis à partir d'informations tirées du recensement. Les degrés intermédiaires, lorsqu'il y en a, peuvent comprendre des unités administratives existantes (p. ex. des villages) ou des unités créées artificiellement (par exemple une série de pâtés de maisons).



Important : Au premier degré, le nombre d'unités d'échantillonnage doit être assez important (>100) afin de pouvoir en choisir au moins 50 à 100. Si le nombre d'unités d'échantillonnage choisies au premier degré est moindre, les données seront davantage concentrées et les estimations tirées de l'enquête seront moins précises.

Au premier degré, il faudra constituer, pour chaque grappe, une base de sondage contenant au moins l'ensemble des ménages ou l'ensemble des individus cibles de la grappe.

Aux degrés d'échantillonnage suivants, une base de sondage ne sera nécessaire que pour **certaines** grappes et on n'aura besoin d'informations détaillées (liste des ménages ou des individus éligibles) qu'au dernier degré de l'échantillonnage.

Suite à la page suivante

Choix du plan d'échantillonnage, suite

Terminologie relative à l'échantillonnage à plusieurs degrés en grappes

On trouvera dans le tableau ci-dessous certains termes importants concernant l'échantillonnage à plusieurs degrés en grappes.

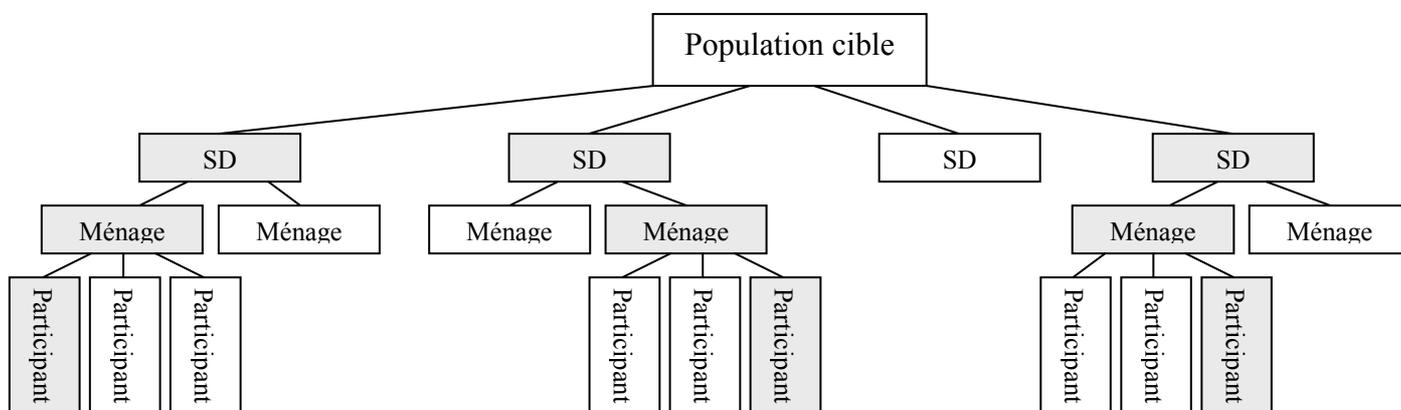
Terme	Définition
Unité primaire d'échantillonnage (UPE)	Grappes choisies en premier lieu. Les UPE sont le plus souvent des secteurs de dénombrement (SD) déterminés lors d'un recensement récent
Unité secondaire d'échantillonnage (USE)	Grappes choisies en second lieu, dans chaque UPE.
Unité tertiaire d'échantillonnage (UTE)	Grappes choisies dans chaque USE.

On peut définir, sur ce modèle, autant de degrés de sondage que nécessaire.

Exemple 1

Dans l'exemple suivant, il y a trois degrés d'échantillonnage. Les SD servent d'UPE. Pour chaque UPE choisie, une base de sondage comprenant la liste des ménages du SD a été créée. On a ensuite sélectionné des ménages dans chaque UPE puis un participant dans chaque ménage.

Les encadrés grisés indiquent que la grappe ou le participant a été choisi(e).



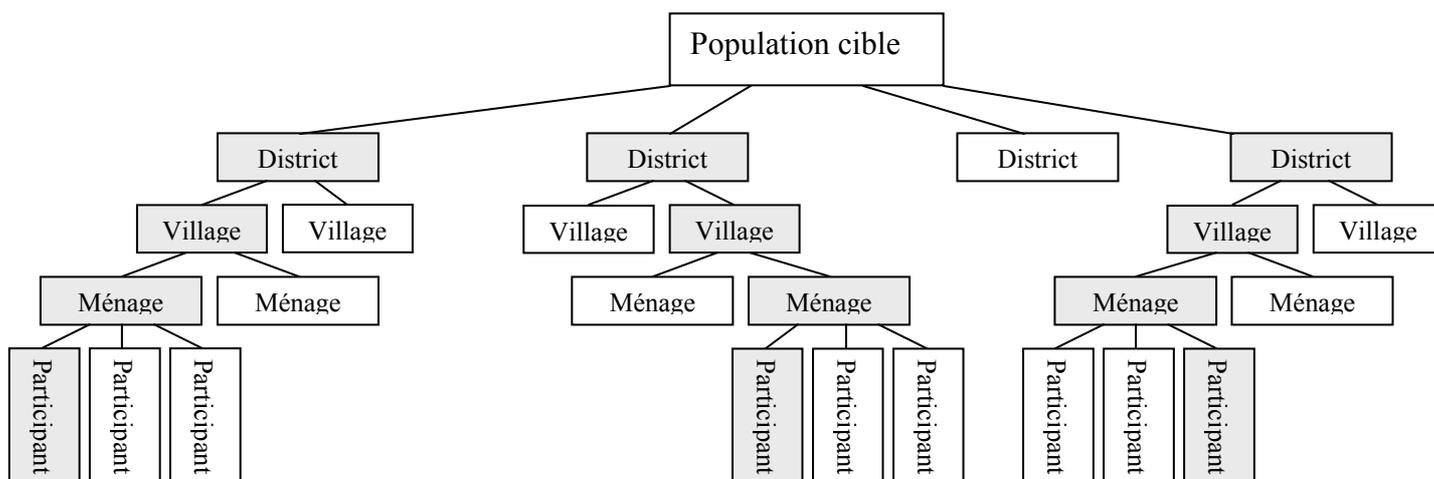
Suite à la page suivante

Choix du plan d'échantillonnage, suite

Exemple 2

Dans cet exemple, il y a quatre degrés d'échantillonnage. Les districts servent d'UPE. Pour chaque UPE choisie, une base de sondage comprenant la liste de tous les villages (les USE) et la population cible de chaque village a été créée. On a aussi établi, pour chaque village, une base de sondage comprenant la liste de tous les ménages. Si une liste détaillée de tous les individus éligibles est disponible pour l'un quelconque des villages sélectionnés, on peut utiliser cette liste à la place de celle des ménages et on peut sélectionner directement les participants.

Les encadrés grisés indiquent que la grappe ou le participant a été choisi(e).



Qualités d'un bon plan d'échantillonnage à plusieurs degrés en grappes

Pour un plan d'échantillonnage à plusieurs degrés en grappes, il est très important de vérifier que chaque individu de la population cible ne figure que dans une seule unité d'échantillonnage par degré. Cela signifie qu'à chaque degré d'échantillonnage, les grappes doivent couvrir l'ensemble de la population cible et s'exclure mutuellement (il ne doit pas y avoir de chevauchement).

Il est aussi important de vérifier les caractéristiques des UPE. Les deux premières rubriques de la liste ci-dessous peuvent aussi servir à vérifier les USE et les UTE mais, étant donné la nature des plans d'échantillonnage à plusieurs degrés en grappes, ces vérifications sont surtout importantes pour les UPE.

Si ...	Alors ...
Certaines UPE sont très petites.	Fusionner ces UPE avec une UPE voisine avant de sélectionner l'échantillon.
Certaines UPE sont très grandes.	Diviser ces UPE en deux ou plus, pour obtenir une taille plus proche de celle des autres UPE.
Il n'y a pas assez d'UPE (<100).	Commencer à échantillonner au niveau des USE (qui deviendront alors des UPE) ou subdiviser les UPE existantes pour pouvoir sélectionner au moins 50 à 100 UPE.

Suite à la page suivante

Choix du plan d'échantillonnage, suite

Répartition de l'échantillon et plan d'échantillonnage à plusieurs degrés en grappes

Une fois que les unités d'échantillonnage du premier et du second degré ont été définies, il faut déterminer la répartition de l'échantillon, c'est-à-dire le nombre total d'UPE, et le nombre total d'USE par UPE, à sélectionner.

Le tableau ci-dessous décrit la marche à suivre pour déterminer la répartition de l'échantillon.

Étape	Description
1	Calculer la taille totale de l'échantillon.
2	Évaluer les ressources disponibles et déterminer le nombre total d'UPE à échantillonner, sans oublier qu'il faut en sélectionner au moins 50 à 100.
3	Diviser la taille totale de l'échantillon par le nombre d'UPE à échantillonner pour calculer le nombre d'individus à sonder par UPE.
4	Continuer à subdiviser l'échantillon à chaque degré de l'échantillonnage selon le nombre d'unités à sélectionner à chaque degré.

Note : Comme il a déjà été indiqué, il est possible de combiner stratification et plan d'échantillonnage à plusieurs degrés en grappes. Le nombre total d'UPE sera réparti proportionnellement ou non (selon les résultats que l'on souhaite obtenir de l'enquête) dans toutes les strates, en suivant les étapes indiquées ci-dessus.

Exemple

Dans cet exemple, supposons que l'échantillon est de 3200 personnes. On a décidé que les UPE seraient les régions, les USE les villages et que des ménages seraient sélectionnés dans tous les villages. Les ressources disponibles permettent de sélectionner 80 UPE, ce qui signifie que 40 individus ($=3200/80$) seront sélectionnés dans chaque UPE.

Les 40 individus de chaque UPE peuvent être répartis avec une certaine souplesse. Il est alors souhaitable d'envisager quelques scénarios et d'en choisir un qui soit réaliste et permette de bien répartir les individus dans l'UPE (il ne faut pas qu'il y ait trop ou pas assez d'individus du même village). Deux scénarios sont proposés ci-dessous :

Suite à la page suivante

Choix du plan d'échantillonnage, suite

Exemple (suite)

Scénario	Description
1	10 individus seront sélectionnés dans chaque village, ce qui signifie qu'il faut choisir 4 villages (= 40/10) par UPE. <u>Répartition de l'échantillon :</u> 80 régions x 4 villages/région x 10 individus/village = 3200
2	5 individus seront sélectionnés dans chaque village, ce qui signifie qu'il faut choisir 8 villages (= 40/5) par UPE. <u>Répartition de l'échantillon :</u> 80 régions x 8 villages/région x 5 individus/village = 3200

En termes de ressources, la principale différence entre les scénarios ci-dessus est le nombre de villages dans lesquels il faudrait se rendre dans chaque UPE, ce qui sera probablement un critère décisif dans la répartition de l'échantillon. Il ne faut toutefois pas oublier que s'il on sélectionne un grand nombre d'individus dans quelques villages seulement, les données de l'enquête seront davantage concentrées et les estimations seront peut-être moins précises.

Exemple avec stratification

Dans cet exemple, supposons encore que l'échantillon est de 3200 personnes et que les UPE seront les régions, les USE les villages, et que des ménages seront sélectionnés dans tous les villages. Exactement comme dans l'exemple précédent, les ressources disponibles permettent de sélectionner 80 UPE. Ici, cependant, les concepteurs de l'enquête souhaitent que l'échantillon soit déterminé proportionnellement sur les 4 îles du pays.

Le tableau ci-dessous indique la part de la population totale vivant sur chaque île. La colonne de droite indique la répartition proportionnelle des UPE sur les 4 îles ou strates.

Île	Proportion de la pop. totale	UPE
A	0,50	40
B	0,175	14
C	0,125	10
D	0,20	16
Total	1,00	80

40 régions (UPE) seront donc choisies parmi les régions de l'île A, 14 régions parmi celles de l'île B, etc... Une fois que les UPE sont sélectionnées sur chaque île, la répartition se poursuit comme dans l'exemple précédent. Il faut sélectionner le même nombre de villages dans chaque UPE, quelle que soit l'île où celle-ci est située.

Sélection de l'échantillon

Introduction

Une fois que le plan d'échantillonnage a été choisi et que la base de sondage a été établie, vous pouvez procéder à la sélection de l'échantillon. Cette section donne des indications pour les différentes étapes d'échantillonnage.

Outils disponibles

STEPS Sampling.xls est un classeur Excel qui comporte des feuilles de calcul pour chaque étape de la sélection de l'échantillon. Il vous permettra :

- d'obtenir un échantillonnage avec probabilité proportionnelle à la taille (PPT) (voir description ci-dessous) pour les unités primaire et secondaire d'échantillonnage selon les besoins ;
- de sélectionner aléatoirement des ménages ou des individus
- d'avoir des informations sur la pondération.

Ce classeur est disponible sur le site Web de STEPS (www.who.int/chp/steps) et sur le cédérom.

Échantillonnage avec probabilité proportionnelle à la taille (PPT)

L'échantillonnage avec probabilité proportionnelle à la taille (PPT) est une méthode par laquelle la probabilité de sélectionner une unité d'échantillonnage donnée est proportionnelle à sa taille (c'est-à-dire, le plus souvent, le nombre d'individus ou de ménages dans l'unité d'échantillonnage).

Il est intéressant d'utiliser l'échantillonnage PPT lorsque les unités d'échantillonnage sont de tailles très différentes. En effet, dans ces cas, si on sélectionnait les unités d'échantillonnage de manière aléatoire, la probabilité de sélectionner les individus serait beaucoup plus faible dans les grandes unités que dans les petites. L'échantillonnage PPT permet de remédier à ce problème et donc de réduire le biais dans les estimations tirées de l'enquête.

Instructions pour l'échantillonnage PPT

Le tableau ci-dessous présente les différentes étapes de l'échantillonnage PPT de plusieurs unités. Avant de commencer, il faut dresser une liste des unités d'échantillonnage en indiquant leur taille (exprimée en nombre de ménages ou d'individus). Il est conseillé d'organiser cette liste par zones géographiques. Enfin, il faut déterminer le nombre d'unités d'échantillonnage (de grappes) à sélectionner.

Les feuilles de calcul du classeur **STEPSsampling.xls** permettent d'effectuer automatiquement les opérations des étapes 3 à 8 indiquées dans le tableau ci-dessous. La feuille d'instructions qui se trouve dans le fichier explique comment procéder à un échantillonnage PPT à partir de la feuille de calcul UPE ou USE.

Suite à la page suivante

Sélection de l'échantillon, suite

Instructions pour l'échantillonnage PPT (suite)

Étape	Action
1	Dressez la liste de toutes les unités d'échantillonnage, en indiquant leur taille (exprimée en nombre de foyers ou d'individus). Si possible, organisez la liste géographiquement en indiquant à la suite les unités adjacentes.
2	Déterminez le nombre d'unités d'échantillonnage à sélectionner sur la liste.
3	Créez une nouvelle colonne et indiquez-y la taille cumulée des unités d'échantillonnage. Le total doit être égal à la population de l'ensemble des unités d'échantillonnage.
4	Divisez la taille totale de la population cumulée (N) par le nombre d'unités d'échantillonnage à sélectionner (n) pour obtenir l'intervalle d'échantillonnage (k). $k = N/n$
5	Choisissez un nombre aléatoire (r) compris entre 1 et l'intervalle d'échantillonnage (k) $1 < r < k$
6	Commencez en haut de la liste et sélectionnez la première unité d'échantillonnage dont la taille de la population cumulée inclut le nombre aléatoire (r).
7	Sélectionnez une deuxième grappe en ajoutant l'intervalle d'échantillonnage au nombre aléatoire (r). Commencez à compter à partir de la grappe précédente jusqu'à ce que la taille de la population cumulée inclue la somme de r+ k.
8	Sélectionnez les grappes restantes en ajoutant l'intervalle d'échantillonnage, multiplié par 2, puis 3, etc., au nombre aléatoire. Commencez toujours à compter à partir de la grappe précédente et non au début de la liste. $r+(k \times 2)$ $r+(k \times 3)$ etc.
9	Continuez jusqu'à atteindre la fin de la liste. Ne vous arrêtez pas dès que votre quota est atteint. Pour éviter les biais, toutes les unités sélectionnées doivent être utilisées dans l'enquête, même si le nombre est légèrement supérieur à n.

Application de l'échantillonnage PPT avec un plan à plusieurs degrés en grappes

Il est possible d'appliquer l'échantillonnage PPT à tous les stades de l'échantillonnage à plusieurs degrés en grappes sauf au dernier, au moment de la sélection des ménages ou des individus.

Le classeur **STEPSsampling.xls** contient des feuilles de calcul qui vous permettent de sélectionner vos UPE et vos USE en appliquant l'échantillonnage PPT. La feuille de calcul intitulée UPE permet de sélectionner jusqu'à 100 UPE sur une liste et la feuille de calcul intitulée USE permet de sélectionner des USE dans chaque UPE. Il faut donc faire une copie de la feuille USE pour chaque UPE sélectionnée.

Suite à la page suivante

Sélection de l'échantillon, suite

Sélection des ménages et/ou des individus

La sélection des ménages et/ou des individus, dernière étape de l'échantillonnage, dépendra du type d'informations disponibles. Le tableau ci-dessous présente les différents scénarios possibles et la procédure de sélection de l'échantillon correspondante

Si ...	Alors ...
Il existe une liste des individus éligibles dans l'unité d'échantillonnage sélectionnée (p. ex. le village).	<p>Vérifiez d'abord que la liste des individus éligibles remplit les conditions suivantes :</p> <ul style="list-style-type: none">• elle est à jour. Par exemple, les personnes qui ont déménagé ou qui sont décédées n'y figurent pas ;• elle comporte des informations précises permettant aux enquêteurs de localiser tous les individus sélectionnés. <p>Si ces deux conditions sont réunies, on peut sélectionner aléatoirement des individus sur cette liste.</p>
On ne dispose d'aucune information ou presque sur les individus de l'unité d'échantillonnage sélectionnée mais il existe une liste des ménages pour cette unité.	<p>Vérifiez d'abord que la liste des ménages remplit les conditions suivantes :</p> <ul style="list-style-type: none">• elle est à jour et chaque ménage qui y figure correspond à un seul logement ;• elle comporte des informations précises permettant aux enquêteurs de localiser tous les ménages sélectionnés. <p>Si ces deux conditions sont réunies, on peut sélectionner aléatoirement des ménages sur cette liste. Pour sélectionner aléatoirement des participants de certains ménages, on peut appliquer la méthode Kish, objet de la section suivante.</p> <p>Si on craint que la liste ne soit périmée, l'équipe d'enquêteurs doit d'abord procéder à une vérification rapide de l'unité d'échantillonnage pour actualiser la liste, en notant les logements abandonnés ou détruits, les nouveaux logements ou les logements qui sont occupés par plusieurs familles alors qu'ils ne l'étaient que par une seule auparavant.</p>

Suite à la page suivante

Sélection de l'échantillon, suite

Sélection des ménages et/ou des individus (suite)

Si ...	Alors ...
Il existe une liste des individus éligibles dans l'unité d'échantillonnage sélectionnée (p. ex. le village).	<p>Vérifiez d'abord que la liste des individus éligibles remplit les conditions suivantes :</p> <ul style="list-style-type: none">• elle est à jour. Par exemple, les personnes qui ont déménagé ou qui sont décédées n'y figurent pas ;• elle comporte des informations précises permettant aux enquêteurs de localiser tous les individus sélectionnés. <p>Si ces deux conditions sont réunies, on peut sélectionner aléatoirement des individus sur cette liste.</p>
On ne dispose d'aucune information ou presque sur les individus de l'unité d'échantillonnage sélectionnée mais il existe une liste des ménages pour cette unité.	<p>Vérifiez d'abord que la liste des ménages remplit les conditions suivantes :</p> <ul style="list-style-type: none">• elle est à jour et chaque ménage qui y figure correspond à un seul logement ;• elle comporte des informations précises permettant aux enquêteurs de localiser tous les ménages sélectionnés. <p>Si ces deux conditions sont réunies, on peut sélectionner aléatoirement des ménages sur cette liste. Pour sélectionner aléatoirement des participants de certains ménages, on peut appliquer la méthode Kish, objet de la section suivante.</p> <p>Si on craint que la liste ne soit périmée, l'équipe d'enquêteurs doit d'abord procéder à une vérification rapide de l'unité d'échantillonnage pour actualiser la liste, en notant les logements abandonnés ou détruits, les nouveaux logements ou les logements qui sont occupés par plusieurs familles alors qu'ils ne l'étaient que par une seule auparavant.</p>

Suite à la page suivante

Sélection de l'échantillon, suite

Sélection des ménages et/ou des individus (suite)

Si ...	Alors ...
On connaît le nombre mais pas la localisation des ménages de l'unité d'échantillonnage.	Il faut cartographier l'unité d'échantillonnage pour déterminer où se trouvent les ménages. Pour plus d'informations sur cette méthode ou d'autres solutions envisageables, veuillez contacter l'équipe STEPS.

La feuille de calcul « RandHold » du classeur **STEPSSampling.xls** peut servir à sélectionner aléatoirement le nombre souhaité de participants sur une liste d'individus ou le nombre souhaité de ménages sur une liste de ménages.

Il est possible que l'on dispose d'informations plus détaillées pour certaines unités d'échantillonnage que pour d'autres. Dans ce cas, les cas de figure ci-dessus ne peuvent être appliqués qu'au cas par cas : dans les unités pour lesquelles on dispose d'informations plus détaillées, les individus peuvent être sélectionnés directement alors que dans les autres, il faut d'abord sélectionner les ménages.

Note : Tous les plans d'enquête STEPS prévoient qu'il est impossible de remplacer une unité ou une personne une fois que celle-ci a été sélectionnée. Si vous remplacez les personnes qui ne répondent pas ou celles qui ne sont pas à leur domicile au moment de l'entrevue, votre échantillon sera biaisé, il ne sera plus représentatif de la population cible.

La méthode Kish

Introduction

La méthode Kish est une technique qui permet de sélectionner aléatoirement un individu dans chaque ménage.

La méthode Kish peut être utilisée pour sélectionner des individus à l'intérieur des ménages quelle que soit la méthode d'échantillonnage choisie pour sélectionner ces ménages.

Documents

Pour appliquer la méthode Kish, vous aurez besoin de la **feuille de couverture Kish des ménages**, qui se trouve à la partie 6, section 2 de ce manuel.

Procédure

Le tableau ci-dessous fournit des instructions détaillées pour l'application de la méthode Kish dans chaque ménage. Vous en trouverez aussi un résumé au début de la feuille de couverture Kish des ménages.

Étape	Description																				
1	Demandez l'âge et le sexe de toutes les personnes du ménage âgées de 25 à 64 ans. Inscrivez ces informations dans le tableau vide qui se trouve sur la feuille de couverture.																				
2	<p>Attribuez un rang à chaque personne inscrite dans le tableau. Les rangs doivent être consécutifs et le premier porter le numéro 1. Attribuez les rangs selon les règles suivantes :</p> <ul style="list-style-type: none">• Attribuez d'abord un rang aux personnes de sexe masculin, par ordre décroissant d'âge (de la plus âgée à la plus jeune).• Attribuez ensuite un rang aux personnes de sexe féminin, par ordre décroissant d'âge. <p>Exemple :</p> <table border="1"><thead><tr><th>Sexe</th><th>Âge</th><th></th><th>Rang</th></tr></thead><tbody><tr><td>F</td><td>45</td><td></td><td>2</td></tr><tr><td>M</td><td>45</td><td></td><td>1</td></tr><tr><td>F</td><td>29</td><td></td><td>4</td></tr><tr><td>F</td><td>32</td><td></td><td>3</td></tr></tbody></table>	Sexe	Âge		Rang	F	45		2	M	45		1	F	29		4	F	32		3
Sexe	Âge		Rang																		
F	45		2																		
M	45		1																		
F	29		4																		
F	32		3																		
3	Dans le tableau de sélection de Kish (en bas de la feuille de couverture Kish des ménages), cherchez la colonne dont l'intitulé correspond au dernier chiffre du numéro d'identification du ménage. Dans cette colonne, cherchez la ligne dont l'intitulé correspond au nombre total de personnes éligibles dans le ménage. Le chiffre qui figure dans la case au croisement de cette ligne et de cette colonne correspond au rang de la personne à interroger.																				

Suite à la page suivante

La méthode Kish, suite

Préparation des documents Il faut préparer les feuilles de couverture Kish des ménages avant de commencer à recueillir des données. Les instructions concernant le remplissage des feuilles de couverture, et notamment l'attribution des numéros d'identification des ménages, se trouvent à la page 2-2-28, intitulée « Préparation des formulaires de collecte des données ».

Surreprésentation des 55-64 ans Selon la structure de la population sur votre site, il peut se révéler difficile d'interroger assez de personnes âgées de 55 à 64 ans pour obtenir des estimations précises pour cette classe d'âge. Afin de remédier à ce problème, on peut choisir de la surreprésenter au niveau des ménages.

On sélectionnera deux personnes âgées de 55 à 64 ans dans tous les ménages où il y en a et on aura donc besoin de deux feuilles de couverture Kish des ménages. L'une servira à sélectionner une personne dans le groupe qui n'est pas surreprésenté et l'autre à sélectionner une personne dans le groupe surreprésenté. Donc, sur l'une des feuilles on inscrira toutes les personnes âgées de 25 à 54 ans et sur l'autre toutes celles âgées de 55 à 64 ans.

Note : En cas de surreprésentation, assurez-vous d'ajuster la taille du ménage. Si dans un ménage de cinq personnes, une est âgée de 55 à 64 ans, la taille à prendre en compte pour échantillonner la classe d'âge 25-54 ans n'est que de quatre personnes et d'une personne seulement pour la classe d'âge 55-64 ans.

Documentation du plan d'échantillonnage

Introduction Une fois que le plan et la méthodologie d'échantillonnage ont été choisis, tous les aspects de l'échantillonnage doivent être clairement documentés.

Objectif Documenter le plan d'échantillonnage a pour objectif de permettre à l'analyste de données de comprendre comment l'échantillon a été déterminé afin d'ajuster correctement les résultats à la population cible. En outre, la présentation des données de l'enquête doit toujours être accompagnée d'une version résumée de la documentation afin d'expliquer comment les données ont été recueillies.

Enregistrement des informations pendant la collecte de données Suffisamment d'informations doivent être relevées **pendant la collecte de données** afin que l'analyste de données puisse procéder à tous les ajustements nécessaires pour que les résultats soient représentatifs de la population cible. L'analyste doit avant tout connaître :

- la probabilité de sélection de chaque unité à chaque degré d'échantillonnage (c'est-à-dire la probabilité de sélectionner chaque UPE, USE, ménage et individu) ;
- l'âge et le sexe des non-répondants.

Il est donc primordial de relever :

- l'ensemble des bases de sondage utilisées à chaque degré d'échantillonnage ;
 - la méthode de sélection de l'échantillon appliquée à chaque degré d'échantillonnage ;
 - le mode de stratification (en cas de stratification) ;
 - l'UPE, l'USE, etc... où chaque personne interrogée a été sélectionnée.
-

Enquêtes à venir Documenter les plans et la méthodologie d'échantillonnage est également important pour les enquêtes à venir lorsqu'on étudie l'évolution des facteurs de risque dans le temps, car les méthodes choisies pourront être différentes, ce qui aura une incidence sur la comparabilité

Archivage des documents Il est important d'archiver tous les documents pertinents relatifs à l'échantillonnage, notamment les formulaires présentés dans la partie suivante de cette section, les formulaires de préparation de recueil des données et tous les documents contenant des informations utilisées pour planifier et déterminer l'échantillon.

Si l'échantillon est déterminé par un organisme public tiers (p. ex. le bureau de la statistique), veillez à ce que celui-ci vous communique tous les documents et toutes les informations utilisées pour déterminer l'échantillon.

Préparation des formulaires de collecte des données

Introduction

Une fois que l'échantillon a été déterminé, le Formulaire de suivi des entretiens, les Feuilles de couverture Kish des ménages et, le cas échéant, les formulaires de recueil des données pour le Step 3 peuvent être préparés pour l'équipe de collecte des données. Il est souhaitable que la personne chargée de superviser le recueil des données et le conseiller en statistique participe à cette tâche afin de veiller à ce que les formulaires soient correctement remplis et organisés pour le recueil des données.

Attribution de numéros d'identification uniques

Avant de préparer les formulaires de recueil des données, il faut attribuer un numéro d'identification à tous les enquêteurs et à toutes les grappes où des ménages et/ou des individus seront sélectionnés. Il faut également attribuer un numéro d'identification unique à tous les ménages et participants sélectionnés. Le tableau ci-dessous donne des instructions plus détaillées sur l'attribution de ces numéros d'identification.

Variable	Description	Fourchette de valeurs
N° d'identification de l'enquêteur	Il faut attribuer un numéro d'identification unique à chaque enquêteur.	1-999
Numéro de la grappe	Il faut attribuer un numéro unique à toutes les unités d'échantillonnage où des ménages et/ou des individus seront sélectionnés. Ces unités sont souvent des villages mais il peut s'agir de pâtés de maisons ou de quartiers, selon la base de sondage. Note : Si la sélection des ménages ou des individus est <u>le premier ou le seul</u> degré d'échantillonnage, il n'est pas nécessaire d'attribuer des numéros aux grappes.	1-999
Numéro du ménage	Il faut attribuer un numéro d'identification unique à chaque ménage qui sera contacté, en partant de 1 jusqu'à avoir numéroté tous les ménages concernés. Il est possible d'attribuer ce numéro avant même de commencer à recueillir les données car le nombre total de ménages à contacter doit être déterminé dès la conception du plan d'échantillonnage. Si on ne fait pas d'entrevue dans l'un des ménages sélectionnés, le numéro qui lui a été attribué n'est tout simplement pas utilisé.	1-99999

Suite à la page suivante

Préparation des formulaires de collecte des données, suite

Attribution de numéros d'identification uniques (suite)

Variable	Description	Fourchette de valeurs
Numéro d'identification du participant	Il faut attribuer un numéro d'identification unique à chaque participant, sans qu'il soit nécessaire que ces numéros se suivent. Il est parfois souhaitable d'opérer un regroupement par numéro de grappe (d'associer une séquence de numéros d'identification à un numéro de grappe). Ainsi, les participants 101 à 120 peuvent être rattachés à la grappe n° 1, les participants 201 à 220 à la grappe n° 2, etc...	1-999999

Il faudra également attribuer les trois numéros d'identification suivants, qui devront être communiqués à l'équipe chargée du recueil des données, si besoin.

Variable	Description	Fourchette de valeurs
Numéro d'identification de l'équipe chargée du recueil des données	Il faut attribuer un numéro d'identification unique à chaque équipe chargée du recueil des données et noter les numéros d'identification des enquêteurs rattachés à chaque équipe.	A-Z
Numéro d'identification du technicien	Si une autre personne que l'enquêteur (par exemple, un clinicien) se charge des Steps 2 et 3, il faut aussi lui attribuer un numéro d'identification.	1-999
Numéro d'identification du matériel	Si l'on décide de mettre en oeuvre les Steps 2 et 3, le matériel utilisé à cette occasion devra porter un numéro d'identification unique.	1-99

Note : Les fourchettes recommandées correspondent aux valeurs attendues dans les modèles standards de STEPS pour la saisie des données. Ces modèles **ne** permettent **pas** d'entrer des valeurs alphanumériques (p. ex. A21) pour un même identifiant.

Suite à la page suivante

Préparation des formulaires de collecte des données, suite

Formulaire de suivi des entretiens

Tous les sites devraient utiliser le Formulaire de suivi des entretiens, quel que soit leur plan d'échantillonnage. Cette information est utilisée à la fois pour calculer les coefficients de pondération et les taux de réponse pour le Step 1, le Step 2 et le Step 3 (le cas échéant).

Avant de commencer à recueillir les données, il faut préparer une série de formulaires de suivi des entretiens pour chaque numéro de grappe. On doit préparer au moins un formulaire pour chaque enquêteur chargé des entretiens dans cette grappe.

Avant le recueil des données, il faut indiquer sur chaque formulaire de suivi des entretiens :

- le numéro de la grappe
- le numéro de l'enquêteur
- les numéros des ménages
- les numéros d'identification des participants

Note : Si la sélection des ménages ou des individus correspond au premier degré de l'échantillonnage (c'est-à-dire si on ne numérote pas les grappes), préparer une série de formulaires de suivi des entretiens pour chaque enquêteur et indiquer au moins les trois dernières informations de la liste ci-dessus.

Feuille de couverture Kish des ménages

La feuille de couverture Kish des ménages devrait être utilisée lorsque l'équipe de collecte des données est amenée à sélectionner aléatoirement des participants au sein de chaque ménage. Il faut préparer une feuille de couverture Kish pour chaque ménage à contacter.

Avant de commencer à recueillir les données, il faut indiquer sur chaque feuille :

- le numéro du ménage
- le numéro de la grappe
- le numéro d'identification des participants

On peut, en outre, entourer la colonne à utiliser dans le tableau de sélection de Kish afin de s'assurer que la méthode est appliquée correctement.

Suite à la page suivante

Préparation des formulaires de collecte des données, suite

Formulaires de recueil des données pour le Step 3

Si le Step 3 est mis en oeuvre, il faut préparer trois formulaires pour chaque lieu (p. ex. centre de soins) où les analyses de sang seront effectuées :

- un formulaire d'entrée au centre de soins
- un formulaire pour les prélèvements sanguins
- un formulaire pour les analyses biochimiques

où on indiquera, avant le recueil des données :

- le numéro de la grappe
 - le numéro d'identification du technicien.
-