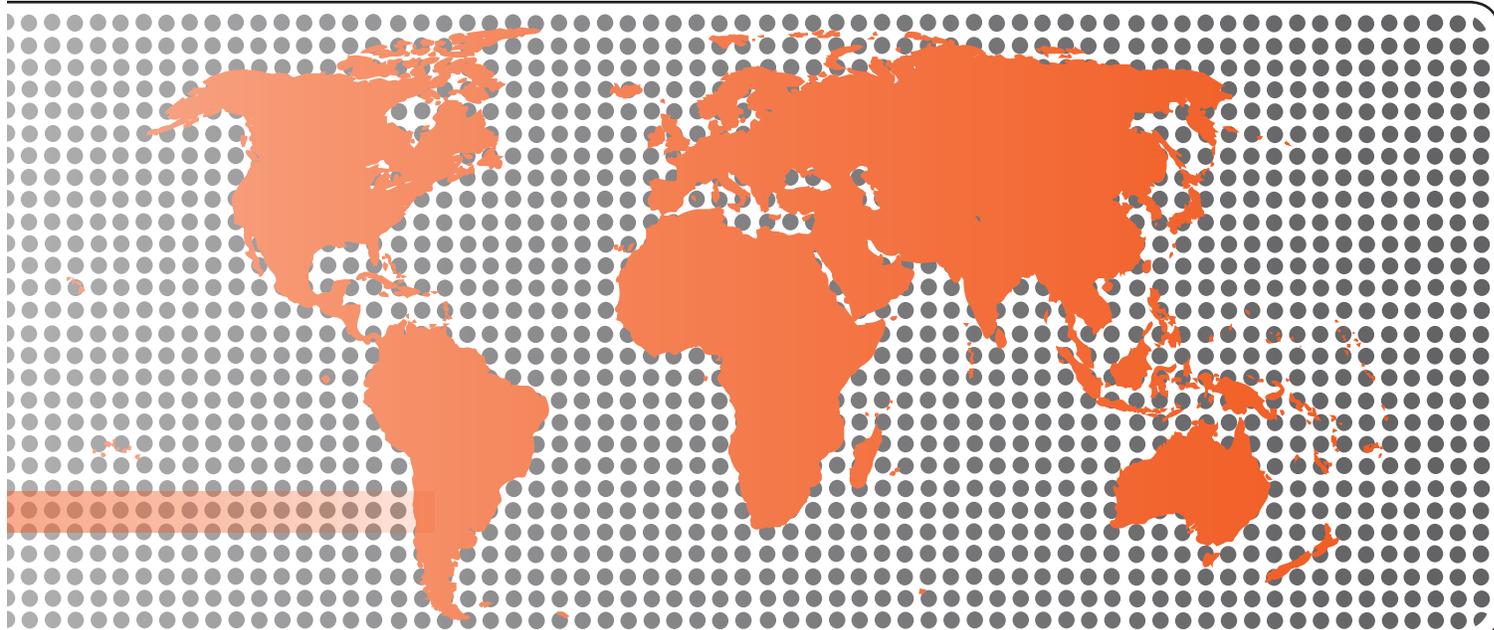


GTSS Global Adult Tobacco Survey (GATS)



Manuel de conception de l'échantillonnage





**Enquête mondiale sur le tabagisme chez  
l'adulte (GATS, Global Adult Tobacco  
Survey)  
Manuel de conception de  
l'échantillon**

Version 2.0  
Novembre 2010

## **Enquête mondiale sur le tabagisme chez l'adulte (GATS) Protocole standard complet**

.....

### **Questionnaire GATS**

Questionnaire principal contenant des questions facultatives  
Spécifications question par question

### **Modèle d'échantillonnage d'une enquête GATS**

Manuel de conception de l'échantillon  
Manuel de pondération de l'échantillon

### **Mise en application sur le terrain d'une enquête GATS**

Manuel de l'enquêteur de terrain  
Manuel du superviseur de terrain  
Manuel de cartographie et d'enregistrement

### **Gestion des données de l'enquête GATS**

Guide du programmeur du système général de l'enquête  
Spécifications de base de programmation du questionnaire  
Plan d'application de la gestion des données  
Guide de formation à la gestion des données

### **Assurance qualité GATS : Directives et documentation**

#### **Kit d'analyse et de rapport d'une enquête GATS**

Modèle d'aide-mémoire  
Rapport national : Plan et lignes directrices relatives aux tabulations  
Définitions des indicateurs

#### **Publication et diffusion des données d'une enquête GATS**

Politique de diffusion des données  
Diffusion des données : Directives pour la diffusion initiale des données

**Questions sur le tabagisme destinées aux enquêtes : un ensemble de questions clés tiré de l'enquête mondiale sur le tabagisme chez les adultes (GATS)**

## **Citation proposée**

Global Adult Tobacco Survey Collaborative Group. *Global Adult Tobacco Survey (GATS): Sample Design Manual, Version 2.0*. Atlanta, GA : Centers for Disease Control and Prevention, 2010.

## **Remerciements**

### ***Organismes participant à l'enquête GATS***

- Centers for Disease Control and Prevention
- Fondation CDC
- Johns Hopkins Bloomberg School of Public Health (École de santé publique Johns Hopkins Bloomberg)
- RTI International (Research Triangle Institute)
- École Gillings de santé publique, Université de Caroline du Nord
- Organisation mondiale de la santé

### ***Soutien financier***

La réalisation de ce manuel est soutenue financièrement par l'Initiative Bloomberg de lutte contre le tabagisme (Bloomberg Initiative to Reduce Tobacco Use), un programme de Bloomberg Philanthropies, par le biais de la Fondation CDC.

---

Clause de non-responsabilité : Les opinions exprimées dans ce manuel ne sont pas nécessairement celles des organisations participant à l'enquête GATS.



# Table des matières

---

Chapitre	Page
<b>1. Introduction</b> .....	<b>1-1</b>
1.1 Vue d'ensemble de l'enquête mondiale sur le tabagisme chez l'adulte .....	1-1
1.2 Utilisation de ce manuel .....	1-2
1.3 Décider d'un plan d'échantillon pour la GATS .....	1-3
<b>2. Objectifs de la conception de l'enquête</b> .....	<b>2-1</b>
<b>3. Population cible et base de sondage</b> .....	<b>3-1</b>
3.1 Définitions de la population cible de la GATS et admissibilité à l'étude .....	3-1
3.2 Zone de la base de sondage.....	3-5
<b>4. Spécifications de base de la conception de l'enquête</b> .....	<b>4-1</b>
4.1 Caractéristiques de la conception de l'échantillon .....	4-1
4.2 Tailles de l'échantillon et la précision attendue.....	4-3
4.3 Processus de recueil des données .....	4-7
<b>5. Aperçu de la méthode d'échantillonnage</b> .....	<b>5-1</b>
<b>6. Former des unités primaires d'échantillonnage (PSU)</b> .....	<b>6-1</b>
<b>7. Première étape de l'échantillonnage : Sélection des PSU</b> .....	<b>7-1</b>
7.1 Pays sélectionnant des zones géographiques (c.-à-d. les segments) en une seule étape de sélection .....	7-1
7.2 Pays sélectionnant des zones géographiques en plusieurs étapes de sélection.....	7-4
<b>8. Sélectionner les foyers</b> .....	<b>8-1</b>
8.1 Créer une liste des foyers pour un segment échantillonné en visitant le segment avant de sélectionner l'échantillon de foyers .....	8-1
8.2 Créer une liste des foyers pour un segment échantillonné à l'aide des registres administratifs .....	8-2
8.3 Sélection de l'échantillon de foyers.....	8-3
8.4 Affectation facultative des foyers à des groupes de foyers masculins et féminins.....	8-4
8.5 Préparation d'un fichier électronique d'échantillon pour l'administration sur l'ordinateur de poche.....	8-5
8.6 Traiter les problèmes de foyers manqués et autres erreurs apparentes dans la liste des foyers.....	8-5
<b>9. Sélectionner un résident admissible dans chaque foyer présélectionné</b> .....	<b>9-1</b>
<b>10. Déterminer les tailles d'échantillon à chaque étape de la sélection et rapporter les taux de réponse définitifs</b> .....	<b>10-1</b>
10.1 Tailles de l'échantillon à chaque étape de la sélection.....	10-1

10.2 Exemple de calcul de la taille de l'échantillon .....	10-4
10.3 Taux de réponses .....	10-5
<b>11. Calcul de la pondération de l'échantillon final pour analyse .....</b>	<b>11-1</b>
<b>12. Bibliographie .....</b>	<b>12-1</b>
12.1 Échantillonnage : simple et bref .....	12-1
12.2 Échantillonnage : préliminaire .....	12-1
12.3 Échantillonnage : traitement approfondi .....	12-1
12.4 Manuels pratiques .....	12-1
12.5 Autres publications .....	12-2
12.6 Manuels de la GATS .....	12-2
<b>Annexe A. Justification de la taille recommandée de l'échantillon de la GATS .....</b>	<b>A-1</b>
<b>Annexe B. Options de conception de l'échantillon pour les nouveaux pays de la GATS .....</b>	<b>B-1</b>
<b>Annexe C. Options de conception de l'échantillon pour les pays effectuant la GATS une nouvelle fois .....</b>	<b>C-1</b>

## Documents annexes

---

Numéro	Page
Document 3-1. Exemples de résidence de la GATS .....	3-4
Document 5-1. Aperçu de la conception de l'échantillon de la GATS.....	5-3
Document 8-1. Résumé du processus de résolution du problème des adresses de foyers manqués dans la liste d'enregistrement des foyers .....	8-7
Document 10-1. Codes des décisions finales du questionnaire des foyers .....	10-7
Document 10-2. Codes des décisions finales du questionnaire individuel.....	10-7



## 1. Introduction

---

Le tabagisme est une cause majeure de mortalité prématurée et de maladies évitables dans le monde entier. Environ 5,4 millions de personnes meurent chaque année en raison de maladies liées au tabac, un chiffre qui devrait dépasser 8 millions par an d'ici 2030. Si les tendances actuelles se confirment, le tabagisme sera responsable de la mort d'un milliard de personnes d'ici la fin de ce siècle. On estime que plus des trois quarts de ces décès auront lieu dans les pays à faible et moyen revenu<sup>1</sup>. Un mécanisme de surveillance efficace et systématique est indispensable pour suivre et prendre en charge l'épidémie.

L'*enquête mondiale sur le tabagisme chez l'adulte* (GATS), une composante du système de surveillance mondiale du tabagisme (GTSS), est la norme mondiale en matière de surveillance systématique de l'usage du tabac chez les adultes et du suivi des indicateurs clés de contrôle du tabac. La GATS est une enquête représentative à l'échelle nationale, menée auprès de foyers d'adultes de 15 ans ou plus, et effectuée en utilisant un questionnaire principal normalisé, une conception de l'échantillon, et des procédures de recueil et de gestion des données, qui ont été examinés et approuvés par des experts internationaux. Elle est destinée à renforcer la capacité des pays à concevoir, appliquer et évaluer les programmes de contrôle du tabac.

Afin de maximiser l'efficacité des données recueillies par le biais de la GATS, une série de manuels a été créée. Ces manuels sont conçus pour soumettre des exigences normalisées aux différents pays, ainsi que des recommandations sur la conception et la mise en œuvre de l'enquête, à chaque étape du déroulement de la GATS. Ils visent également à apporter des conseils sur la façon dont un pays donné peut adapter les caractéristiques du protocole de la GATS, en vue de maximiser l'utilité des données dans le pays. Il est fortement encouragé de suivre le protocole standard, afin de maintenir la cohérence et la comparabilité entre les pays.

---

**Les manuels de la GATS fournissent des recommandations systématiques sur la conception et la mise en œuvre de l'enquête.**

---

### 1.1 Vue d'ensemble de l'enquête mondiale sur le tabagisme chez l'adulte

La GATS est conçue pour générer des estimations nationales et sous-nationales parmi les adultes à travers le pays. La population cible comprend tous les hommes et femmes, n'habitant pas dans un établissement spécialisé, de 15 ans ou plus, qui considèrent le pays comme leur lieu habituel de résidence. Tous les membres de la population cible seront échantillonnés à partir du foyer (HH) qui est leur lieu habituel de résidence.

---

<sup>1</sup> Mathers, C.D., and Loncar D. Projections of Global Mortality and Burden of Disease from 2002 to 2030. *PLoS Medicine*, 2006, 3(11):e442.

La GATS utilise une méthodologie d'échantillonnage en plusieurs étapes regroupées géographiquement, pour identifier les foyers particuliers que les enquêteurs de terrain contacteront. Dans un premier temps, un pays est divisé en unités primaires d'échantillonnage, en segments au sein de ces unités primaires d'échantillonnage, et en foyers à l'intérieur des segments. Ensuite, un échantillon de foyers est sélectionné aléatoirement pour participer à la GATS.

---

**L'entretien de la GATS comporte deux parties : le questionnaire du foyer et le questionnaire individuel. Ces questionnaires sont administrés en utilisant un dispositif électronique de recueil des données.**

---

L'entretien de la GATS se compose de deux parties : le *questionnaire du foyer* (présélection des foyers) et le *questionnaire individuel* (entretien individuel). Les deux entretiens sont menés en utilisant un dispositif électronique de recueil de données.

Pour chaque adresse de l'échantillon, les enquêteurs de terrain administreront le *questionnaire du foyer* à un adulte qui fait partie du foyer. Les objectifs du *questionnaire du foyer* sont de déterminer si le foyer sélectionné répond aux exigences d'admissibilité de la GATS, et de faire une liste, ou un tableau de tous les membres admissibles du foyer. Une fois que la liste des résidents admissibles du foyer est dressée, un individu sera choisi au hasard pour répondre au *questionnaire individuel*. Le *questionnaire individuel* demande des renseignements généraux, sur le tabac à fumer, le tabac sans fumée, l'arrêt, le tabagisme passif, les facteurs économiques, les médias, ainsi que les connaissances, les attitudes et les perceptions du tabac.

## 1.2 Utilisation de ce manuel

Ce manuel est conçu pour présenter des exigences et des recommandations, ainsi que les directives suggérées aux pays alors qu'ils développent une conception appropriée de l'échantillon qu'ils utiliseront pour la GATS. Ce chapitre fournit des informations générales sur la GATS. Les chapitres suivants sont résumés ci-dessous :

- Le **chapitre 2** résume les objectifs de l'enquête.
- Le **chapitre 3** présente une définition de la population cible de l'enquête et aborde la base de la zone de sondage qui sera utilisée. Une définition claire de la population cible et de la zone de la base de sondage est impérative. Une définition claire de la population cible permet de mieux apprécier l'étendue de la couverture de la population cible. Cette définition sera aussi utilisée pour déterminer l'admissibilité à l'enquête pour l'échantillonnage et le recueil des données. Ce chapitre fournit également une définition d'un « foyer »<sup>2</sup>.
- Le **chapitre 4** présente les caractéristiques de base de la conception de l'échantillon nécessaires pour que les données soient comparables entre pays et de haute qualité statistique. En outre, les chapitres suivants présentent d'autres exigences de conception et d'autres recommandations.

---

<sup>2</sup>Il convient de noter que tandis que nous utilisons les termes « foyer » et « unité de logement » (ou « unité d'habitation ») de façon un peu interchangeable dans ce manuel, ils ont des significations légèrement différentes. Un foyer est une personne ou un groupe de personnes qui vivent ensemble dans une unité de logement. Ainsi, l'unité de logement désigne la structure physique (par exemple, une maison unifamiliale, un appartement, un mobile home, un groupe de chambres, etc.) dans laquelle résident la ou les personnes composant le foyer.

- Le **chapitre 5** présente un bref aperçu de l'approche recommandée pour l'échantillonnage de la GATS.
- Les **chapitres 6 à 9** résument le processus de sélection de l'échantillon à chaque étape de la conception de l'échantillon.
- Le **chapitre 10** offre des suggestions sur la détermination d'une taille d'échantillon suffisante à la sélection, à chaque étape de la conception.
- Le **chapitre 11** donne un bref aperçu du processus de calcul de pondération de l'échantillon pour chaque répondant. Ces pondérations sont généralement corrigées pour les non-réponses et étalonnées en fonction du dénombrement de la population cible avant d'être utilisées pour l'analyse des données.
- Le **chapitre 12** fournit une liste de références qui pourraient être utiles aux statisticiens lorsqu'ils élaborent le plan d'échantillonnage de la GATS spécifique à leur pays.
- L'**annexe A** contient la justification de la taille de l'échantillon global recommandée pour les enquêtes de la GATS.
- L'**annexe B** propose un ensemble général d'« options N » pour la conception de l'échantillon dans les nouveaux pays de la GATS après la diffusion de la deuxième édition de ce manuel en 2010.
- L'**annexe C** présente un ensemble général d'« options R » pour les pays qui ont achevé un cycle de l'enquête de la GATS et sont sur le point d'effectuer un cycle de répétition.

### 1.3 Décider d'un plan d'échantillon pour la GATS

Ce manuel fournit un modèle statistique général pour les échantillons de probabilité qui doivent être utilisés pour toute enquête de la GATS. Le but de cette section est d'exposer plusieurs options de base pour les deux types de situations de conception pertinents à ce stade du développement du système des enquêtes de la GATS, dans lesquelles se trouvent plusieurs pays qui ont déjà effectué une GATS et d'autres qui cherchent à en réaliser une pour la première fois. Les deux types de situations sont les suivantes : (i) les nouveaux pays espérant mener leur premier cycle d'enquête GATS, et (ii) les pays qui souhaitent répéter une enquête GATS en menant un autre cycle d'enquête.

#### Options de conception des échantillons pour les pays qui effectuent leur premier cycle d'enquête de la GATS

Les options de conception proposées aux nouveaux pays sont définies en fonction de ce qui suit : (i) si l'échantillon national utilisé pour la GATS est statistiquement dérivé d'un échantillon national existant (c'est-à-dire, si la conception de l'échantillon de la GATS est « intégrée » à une conception d'échantillon existante), et (ii) si les tailles des échantillons doivent également être suffisantes pour faire des comparaisons entre les zones urbaines et rurales (en plus de mesurer les différences entre les sexes). Une conception de la GATS intégrée à un échantillon hautement reconnu et des infrastructures de recueil de données existantes peut signifier une enquête à des coûts réduits par rapport à ceux d'un plan d'échantillonnage autonome (c'est-à-dire, développé exclusivement pour la GATS). Par ailleurs, un échantillon de la GATS développé uniquement pour mesurer adéquatement les différences entre les deux catégories de sexe peut être de taille inférieure à un échantillon avec lequel on souhaite obtenir des

estimations suffisamment précises pour quatre sous-groupes de population formés par la classification croisée des sexes et de l'environnement urbain ou rural.

Les trois options de conception envisagées dans ce cadre sont les suivantes :

- **OPTION N1**, une conception autonome standard et une taille d'échantillon global de 8 000, suffisante pour produire des estimations pour les quatre sous-groupes de population sexe-degré d'urbanisation.
- **OPTION N2**, une conception autonome standard inférieure et une taille d'échantillon global de 4 000, suffisante pour produire des estimations pour les deux groupes de sexes et les deux sous-groupes de degré d'urbanisation séparément, mais pas pour les quatre sous-groupes de population sexe-degré d'urbanisation.
- **OPTION N3**, avec une conception de l'échantillon intégrée à celle d'un échantillon national hautement reconnu existant et dont la taille de l'échantillon des répondants peut être du niveau standard (8 000) ou inférieure.

Voir l'**annexe B** pour plus de détails sur ces options.

### **Options de conception de l'échantillon pour les pays répétant la GATS**

Dans le cas des pays répétant la GATS, l'accent est mis sur la façon dont l'échantillon pour le prochain (répétition) cycle se rapporte à l'échantillon du premier (de référence) cycle. Les options dans cette situation sont établies en fonction de ce qui suit : (i) si les deux échantillons ont des chevauchements prévus, et (ii) si les structures (c'est-à-dire définies par les unités d'échantillonnage à chaque stade et par les définitions de strate) des conceptions des échantillons de la répétition et de la référence sont les mêmes.

Les trois options de conception envisagées dans ce cadre sont les suivantes :

- **OPTION R1**, chevauchement partiel (prévu) des échantillons du cycle de référence et du cycle de répétition.
- **OPTION R2**, les structures des échantillons des cycles de référence et de répétition sont identiques, mais les deux échantillons sont choisis de façon indépendante.
- **OPTION R3**, les structures des échantillons des cycles de référence et de répétition ne sont pas identiques, et les deux échantillons sont choisis de façon indépendante.

Voir l'**annexe C** pour plus de détails sur ces options.

## 2. Objectifs de la conception de l'enquête

---

Des exigences et des recommandations pour la conception de l'enquête de la GATS ont été développées afin que des estimations de haute qualité soient générées pour chaque pays dans son ensemble, ainsi que pour deux groupes d'analyse définis par degré d'urbanisation et par sexe.

Comme il sera mentionné dans le **chapitre 4**, la taille ciblée de l'échantillon des répondants est généralement établie de sorte que la moitié environ de l'échantillon de répondants réside dans des zones urbaines, et l'autre moitié dans des zones rurales. Cette répartition de l'échantillon se fera en stratifiant explicitement l'échantillon par degré d'urbanisation (c'est-à-dire, zone urbaine ou rurale) au(x) stade(s) le(s) plus précoce(s) possible(s) de la conception. En option, le plan d'échantillonnage a également été conçu pour que le mélange des hommes et des femmes dans l'échantillon de répondants puisse être contrôlé par l'attribution aléatoire des foyers sélectionnés à des groupes de sexe masculin ou féminin par le biais d'un processus appelé *randomisation des sexes*. Par exemple, si un foyer est attribué au hasard au groupe des hommes, seuls les résidents admissibles masculins seront listés et sélectionnés à partir du foyer pour un entretien. De la même manière, si un foyer est attribué au hasard au groupe féminin, seules les résidentes admissibles seront listées et sélectionnées à partir du foyer pour un entretien. La randomisation des sexes a l'un des deux, ou les deux objectifs suivants : (i) prédéterminer le sexe des membres du foyer sélectionné pour apparier plus facilement le sexe de l'enquêteur de terrain avec celui du répondant, dans les cultures où une disparité entre les sexes doit être évitée, ou (ii) échantillonner de manière disproportionnée les hommes ou les femmes en fonction des besoins en taille de l'échantillon suivant le sexe.

La randomisation des sexes sera facultative pour les pays de la GATS. Par conséquent, un paramètre sera installé dans le code du programme pour permettre aux pays d'établir que des listes des foyers spécifiques au sexe soient créées ou non dans les foyers participants échantillonnés. L'option par défaut de ce paramètre de programmation conduira à l'identification des résidents admissibles des foyers en considérant que la randomisation des sexes n'a **pas** été effectuée.

Il est préférable pour la GATS de concevoir l'enquête de façon à générer des estimations transversales par sexe et par degré d'urbanisation, principalement pour permettre des comparaisons d'estimations suivant ces caractéristiques entre les différents pays participant à cet effort. Certains pays peuvent avoir d'autres caractéristiques démographiques hautement prioritaires pour rapporter les estimations de l'enquête. Par exemple, un pays peut souhaiter produire des estimations précises pour certaines régions géographiques, où les régions peuvent être individuelles ou désigner des groupes d'États, de provinces, ou d'autres types d'entités géographiques définis par des repères, des convictions politiques ou d'autres facteurs culturels. Ce type d'exigences de données supplémentaires imposé par un pays est acceptable, à condition que des estimations correspondantes puissent être produites avec une précision statistique acceptable et qu'il ne compromette pas la qualité statistique des estimations pour les analyses intra et inter-pays des caractéristiques d'intérêt majeur.

---

**Recommandation :**  
**la conception de la version standard de la GATS vise à obtenir des estimations précises par degré d'urbanisation ainsi que par sexe.**

---

Il est à noter que le succès de la génération d'estimations par régions géographiques, en particulier, implique généralement que la taille de l'échantillon global soit sensiblement supérieure à celle d'un échantillon lorsqu'aucune estimation régionale n'est nécessaire. Dans les sections suivantes de ce manuel, on remarquera que l'échantillon de la GATS doit être conçu pour obtenir un nombre minimum de 8 000 répondants par pays, dont environ la moitié se trouve dans des zones urbaines et l'autre moitié dans des zones rurales. Si un pays souhaite produire des estimations régionales et veut comparer les estimations entre zones urbaines et rurales au sein de chaque région, il est recommandé de concevoir l'enquête de façon à avoir 8 000 répondants *dans chaque région* (la moitié répartie dans des zones urbaines et l'autre moitié dans des zones rurales). Si un pays veut produire des estimations régionales et **ne souhaite pas** comparer les estimations entre zones urbaines et rurales de la région, il est recommandé de concevoir l'enquête de façon à avoir 4 000 répondants *dans chaque région*. Ces tailles minimales d'échantillon permettront de produire des estimations qui répondent aux exigences de précision attendue de la GATS.

### 3. Population cible et base de sondage

---

#### 3.1 Définitions de la population cible de la GATS et admissibilité à l'étude

En général, la population cible de la GATS doit inclure toutes les personnes admissibles résidant dans toutes les zones géographiques à l'intérieur d'un pays de la GATS. *Dans de rares cas*, il peut être nécessaire d'exclure certaines zones géographiques. L'exclusion de certaines zones sera envisagée à condition qu'une proportion importante de la population du pays ne réside pas dans ces zones. Par exemple, il est possible d'envisager l'exclusion des zones extrêmement reculées ou des zones où l'on ne peut se rendre en raison d'une guerre, de troubles politiques, etc., de la population cible de la GATS. Pour proposer une exclusion, un pays doit (i) définir explicitement les zones géographiques du pays incluses et exclues de la population cible, et (ii) fournir une estimation du pourcentage de la population qui réside dans les régions à exclure. Ce pourcentage fournira une indication de l'effet potentiel de biais de toute exclusion de zone sur des estimations qui sont par ailleurs destinées à être de portée nationale.

---

**Les pays ne peuvent exclure des portions géographiques du pays de la population cible de la GATS que dans de rares cas.**

---

La population cible des enquêtes de la GATS doit comprendre tous les civils, hommes et femmes, ne résidant pas dans des établissements spécialisés, de 15 ans ou plus, et vivant dans les zones incluses du pays, qui répondent aux exigences de résidence de la GATS au niveau du pays et du foyer. Cette définition est importante pour remplir le *questionnaire du foyer*, qui comprend une section où tous les membres admissibles du foyer sont répertoriés et l'un d'entre eux est choisi au hasard pour répondre au *questionnaire individuel*. Les résidents admissibles uniquement du groupe du sexe désigné pour le foyer doivent être répertoriés si la randomisation des sexes des foyers est effectuée.

---

**Population cible de la GATS :**  
**Celle-ci comprend tous les civils, hommes et femmes, ne résidant pas dans des établissements spécialisés, de 15 ans ou plus, et vivant dans les zones incluses du pays, qui répondent aux exigences de résidence de la GATS au niveau du pays et du foyer où a lieu l'échantillonnage.**

---

Les individus âgés de 15 ans ou plus, qui sont explicitement **exclus** de l'enquête, sont ceux qui au moment de l'entretien du *questionnaire du foyer* sont :

- des non-citoyens qui visitent le pays pour quelques semaines (par exemple des touristes, dans le pays pour voir des amis/des parents, etc.) ;
- des citoyens dans l'armée qui indiquent que leur lieu habituel de résidence est soit dans ou hors d'une base militaire, OU
- des citoyens qui habitent dans des établissements spécialisés, y compris les personnes résidant dans les hôpitaux, les prisons, les maisons de retraite, et autres établissements similaires, qui ne seront pas échantillonnés dans la GATS.

## Exigence de résidence de la GATS au niveau du pays :

Les individus sont considérés comme des résidents d'un pays de la GATS s'ils sont : (i) des citoyens du pays et y résident, ou (ii) des non-citoyens vivant dans le pays, considérant le pays comme leur pays de résidence **principal** (c'est-à-dire, ils ont vécu dans le pays de la GATS pendant au moins la moitié du temps des 12 mois précédant l'entretien du *questionnaire du foyer*).

## Exigence de résidence de la GATS au niveau du foyer :

Tous les civils, hommes et femmes, ne vivant pas dans un établissement spécialisé, âgés de 15 ans ou plus, qui sont inscrits dans une école et résident dans un dortoir pendant l'année scolaire, mais qui considèrent que le foyer sélectionné est normalement leur lieu de résidence, répondent automatiquement à l'exigence de résidence de la GATS au niveau des foyers.<sup>1</sup>

Tous les autres civils, hommes et femmes, ne résidant pas dans un établissement spécialisé, âgés de 15 ans ou plus dans un foyer de l'échantillon satisfont à l'exigence de résidence du foyer si le foyer échantillonné est considéré comme étant leur lieu habituel de résidence au moment où le *questionnaire du foyer* est rempli. Un foyer échantillonné est le *lieu habituel de résidence* d'une personne autrement admissible qui y vit, si cette personne n'a pas d'autre résidence, ou a de multiples résidences, mais a vécu dans le foyer échantillonné pendant au moins la moitié du temps au cours des 12 derniers mois.

En outre, une personne qui a récemment emménagé dans le foyer échantillonné pour en faire son unique résidence est considérée comme un membre de ce foyer si elle n'a pas l'intention de retourner dans son foyer précédent. Inversement, une personne qui a récemment déménagé hors du foyer de l'échantillon, sans intention connue d'y retourner n'est plus considérée comme un membre du foyer de l'échantillon.

Sauf pour les étudiants vivant dans les dortoirs, le lieu habituel de résidence des membres du foyer ayant plus d'une résidence est l'endroit où ils ont passé au moins la moitié du temps au cours des 12 derniers mois. En termes de procédure, il est important de noter que si les résidents choisis au hasard pour remplir un *questionnaire individuel* sont temporairement absents du foyer au moment de la sélection, tous les efforts possibles doivent être entrepris pour réaliser l'entretien avec eux à un moment ultérieur. Ils ne doivent pas être arbitrairement (ou même aléatoirement) remplacés par un autre membre admissible du foyer.

---

**Certains non-citoyens d'un pays de la GATS peuvent satisfaire à l'exigence de résidence au niveau du pays. Des non-citoyens sont considérés comme résidents « habituels » d'un pays de la GATS s'ils ont habité dans ce pays pendant au moins la moitié du temps au cours des 12 derniers mois.**

---

**Tous les étudiants/élèves vivant dans les dortoirs remplissent la condition de résidence de la GATS au niveau du foyer.**

---

---

**Un membre « habituel » d'un foyer échantillonné est tout résident autrement admissible qui : (i) n'a pas d'autre résidence, ou (ii) a de multiples résidences, mais a vécu dans le foyer sélectionné au moins la moitié du temps au cours des 12 derniers mois.**

---

<sup>1</sup> Puisque les dortoirs scolaires ne seront pas inclus sur les listes utilisées pour sélectionner les foyers, mais les élèves, qui vivent temporairement hors de chez eux, sont inclus dans la population cible de la GATS, nous devons considérer les personnes vivant dans ces dortoirs comme des résidents du foyer dont ils font partie quand ils ne sont pas à l'école.

Le **document 3-1** présente quelques exemples spécifiques de l'application des exigences de résidence au niveau du pays et du foyer lors de la constitution des listes des foyers des hommes et femmes, ne résidant pas dans des établissements spécialisés, âgés de 15 ans ou plus. Il indique s'ils peuvent être sélectionnés pour compléter un *questionnaire individuel*.

### Document 3-1. Exemples de résidence de la GATS

<b>Exemple de personne pour laquelle le pays et le foyer de résidence doivent être déterminés</b>	<b>L'enquêteur doit-il considérer cette personne comme admissible à la sélection pour le <i>questionnaire individuel</i> et inclure cette personne sur la liste du foyer ?<sup>2</sup></b>
Une citoyenne est étudiante et vit actuellement dans un dortoir scolaire, mais elle retourne au foyer échantillonné quand l'école est fermée.	Oui. Elle répond au critère d'admissibilité du pays, et puisque les dortoirs ne sont pas échantillonnés, la seule façon pour cette personne d'avoir une chance d'être sélectionnée est de considérer qu'elle est résidente du foyer échantillonné.
Un citoyen est étudiant et vit dans un dortoir scolaire durant l'année scolaire, mais il est chez lui en vacances.	Oui, pour la même raison que dans l'exemple précédent.
Un travailleur agricole temporaire est un citoyen qui considère le foyer échantillonné comme étant « sa maison », mais il a travaillé ailleurs dans le pays pendant 9 des 12 derniers mois.	Non. Cette personne répond au critère du pays puisqu'il est citoyen, mais il a de multiples résidences et n'a pas habité dans la résidence échantillonnée pendant plus de la moitié de l'année écoulée.
Un citoyen étranger dont le seul domicile est le foyer échantillonné depuis les 10 derniers mois, alors qu'il achève une mission de travail temporaire pour son employeur étranger.	Oui. Cet homme répond bien au critère de résidence du pays, car il a vécu dans le pays pendant au moins la moitié de l'année écoulée, il remplit aussi le critère de résidence du foyer puisque c'est sa seule résidence.
Une citoyenne vient d'emménager dans son unique résidence nouvellement achetée, qui est le foyer échantillonné.	Oui. Elle répond au critère de résidence du pays, et a l'intention d'habiter dans ce qu'elle considère être son unique résidence. Elle satisfait ainsi au critère de résidence du foyer.
Au moment où la liste des foyers est effectuée, une citoyenne habite dans sa maison de vacances, qui a été choisie comme le foyer échantillonné et où elle a passé deux jours par semaine pendant l'année écoulée.	Non. Cette personne répond au critère du pays puisqu'elle est citoyenne, mais elle a de multiples résidences et n'a pas habité dans la résidence échantillonnée pendant plus de la moitié de l'année écoulée.
Une citoyenne est propriétaire de deux maisons et celle qui est sélectionnée pour la GATS est sa résidence principale où elle a vécu toute l'année sauf un mois, mais elle habite dans l'autre maison (de vacances) au moment où la liste est dressée.	Oui. Cette personne est citoyenne et par conséquent répond au critère du pays. Elle a de multiples résidences et a habité dans le foyer échantillonné pendant plus de la moitié de l'année écoulée.

En résumé, lors de la visite initiale faite par l'enquêteur à chaque foyer sélectionné, il dressera une liste de tous les résidents admissibles qui considèrent le foyer sélectionné comme étant leur lieu habituel de résidence au moment où la liste est effectuée. Par ailleurs, tous les résidents admissibles des foyers doivent figurer sur la liste, et tous les efforts doivent être entrepris, dans la mesure du possible, pour les interroger s'ils sont sélectionnés afin de compléter le *questionnaire individuel*.

<sup>2</sup> Supposons que la personne de référence dans chaque exemple répond à tous les autres critères d'admissibilité de la GATS.

Le **Manuel de l'enquêteur de terrain de la GATS** fournira des informations spécifiques au pays sur les personnes à inclure et à exclure de la population cible, ainsi qu'une discussion sur quand effectuer un suivi des personnes sélectionnées qui n'habitent pas actuellement dans leur lieu habituel de résidence.

### 3.2 Zone de la base de sondage

En général, la base de sondage d'un échantillon d'une enquête est la liste de toutes les unités d'échantillonnage de la population de l'enquête à partir de laquelle l'échantillon doit être établi. Il existe deux types de bases utilisés dans la plupart des enquêtes en face à face sur des sujets humains : les bases constituées par des listes et les bases constituées par des grappes (par exemple, des zones). Une base de type liste est une simple liste des membres de la population (par exemple, une liste des élèves dans les écoles d'une ville, une liste de fichiers à sélectionner au hasard à partir d'une salle pleine de classeurs à tiroirs). Une base de type grappes est une liste d'entités qui fournissent une référence indirecte à des individus en les regroupant d'une certaine manière (par exemple, une liste des écoles de la ville, chacune étant un regroupement d'élèves, ou une liste des classeurs à tiroirs de la salle, chacun contenant plusieurs fichiers). Cette méthode est dite indirecte, car un échantillon de grappes doit d'abord être sélectionné, puis l'échantillon d'individus est identifié à partir des individus liés à des grappes sélectionnées.

Pour la GATS, comme pour la plupart des études de population, une liste complète de toutes les personnes admissibles à l'enquête dans un pays n'existe tout simplement pas, ce qui rend l'échantillonnage à partir d'une base de type liste pratiquement impossible. Par conséquent, une base de sondage indirecte de type grappes doit être utilisée pour sélectionner l'échantillon. En résumé, les zones géographiques de chaque pays seront sélectionnées aléatoirement, une liste complète de tous les foyers de chaque zone sélectionnée sera établie, un échantillon de foyers sera sélectionné au hasard à partir de chaque liste, et enfin une personne sera choisie au hasard dans chaque foyer sélectionné afin de répondre à l'entretien. Ce type de conception suppose que chaque individu sélectionné admissible à l'enquête ne peut être relié qu'à un seul foyer dans le pays.

#### Définition d'un « foyer »

La section précédente a donné une définition reconnue d'un individu pour l'enquête de la GATS. Il est tout aussi important de fournir une définition reconnue d'un foyer. Bien que cette définition puisse être un peu modifiée suivant le pays, en général nous suggérons aux pays d'utiliser la définition d'un foyer telle qu'établi par les Nations Unies dans leur rapport **Principes et recommandations pour le recensement de la population et des logements**. Le rapport définit un foyer<sup>3</sup> comme suit :

*« Un foyer d'une personne, définie comme un arrangement dans lequel une personne se procure sa propre nourriture ou autres besoins vitaux sans s'associer avec d'autres personnes pour former un foyer de plusieurs personnes, ou un foyer de plusieurs personnes, défini comme un groupe de deux ou plusieurs personnes vivant ensemble qui pourvoient en commun à la nourriture ou autres besoins vitaux. Les personnes du groupe peuvent mettre en commun leurs revenus et inclure des personnes apparentées ou non, ou une combinaison de personnes*

---

<sup>3</sup>Cette définition d'un foyer est tirée du site [http://unstats.un.org/unsd/cdb/cdb\\_dict\\_xrxx.asp?def\\_code=327](http://unstats.un.org/unsd/cdb/cdb_dict_xrxx.asp?def_code=327).

*apparentées et non apparentées. Cet arrangement illustre le concept de ménage. Dans une autre définition utilisée dans de nombreux pays illustrant le concept de ménage-logement, un foyer se compose de toutes les personnes qui vivent ensemble dans une unité de logement. »*

La définition complète et définitive d'un foyer qu'un pays envisage d'utiliser doit être incluse dans la proposition du pays de la conception de l'échantillon de la GATS et plus tard, dans son rapport des conclusions de la GATS.

## 4. Spécifications de base de la conception de l'enquête

---

L'unité d'échantillonnage à chaque étape de sélection désigne les entités qui sont sélectionnées pour l'enquête. Dans cette enquête, les unités finales d'échantillonnage sont les foyers et une personne résidant dans le foyer sélectionné. En général, l'échantillon de la GATS a été conçu pour être sélectionné en utilisant une conception en plusieurs étapes, par grappes géographiques. Une conception en plusieurs étapes désigne une conception de l'échantillon qui consiste à sélectionner progressivement des sous-échantillons d'un échantillon préalablement sélectionné jusqu'à ce que les dernières unités d'échantillonnage soient sélectionnées. Dans le cas de la GATS, des zones géopolitiques existantes seront sélectionnées dans les premiers stades de la conception, les foyers, et pour finir, les individus au sein des foyers, seront sélectionnés lors des dernières étapes de la conception. Les sections suivantes décrivent des détails supplémentaires de la conception.

Comme indiqué dans le **chapitre 1**, certaines exigences et recommandations doivent être suivies afin de maximiser la comparabilité des résultats entre les pays qui participent à la GATS. Cependant, chaque pays a la possibilité d'apporter des améliorations à la conception susceptibles d'augmenter la facilité d'utilisation des résultats de cette enquête (par exemple, sélection de l'échantillon de façon à assurer des estimations précises par région). Dans ce chapitre, nous présentons quelques-unes des exigences de base de la conception de l'enquête. Les améliorations de conception qu'un pays peut souhaiter mettre en place seront généralement acceptables pourvu qu'elles n'interfèrent pas avec ces exigences fondamentales. D'autres exigences spécifiques de conception de l'échantillon sont présentées dans les sections restantes de ce manuel.

En résumé, certaines des caractéristiques de base requises pour la conception de l'enquête de la GATS sont présentées dans les **sections 4.1 à 4.3**.

### 4.1 Caractéristiques de la conception de l'échantillon

Les exigences relatives à la conception de l'échantillon comprennent :

1. Une sélection aléatoire doit être utilisée à chaque étape d'échantillonnage de façon à ce que chaque membre de la population cible ait une chance non nulle d'être inclus dans l'échantillon.
2. La probabilité de sélection pour chaque unité (foyer et individuel) sélectionnée à chaque étape de la conception doit être connue et stockée dans les fichiers d'analyse finale de l'enquête. Par exemple, si l'échantillon est sélectionné en quatre étapes (sélection de l'unité primaire d'échantillonnage, ou PSU, sélection du segment<sup>1</sup>, sélection des foyers, et sélection individuelle), les fichiers d'analyse finale doivent contenir les quatre probabilités de sélection correspondantes. Au minimum, chaque pays doit créer et maintenir deux fichiers :

---

**EXIGENCE :**

**À chaque étape de sélection, toutes les unités doivent être choisies au hasard et les probabilités de sélection à chaque étape doivent être enregistrées.**

---

<sup>1</sup> Un segment est défini comme une zone géographique réduite contenue dans une unité primaire d'échantillonnage donnée. Une ou plusieurs étapes d'échantillonnage intermédiaires peuvent être nécessaires pour sélectionner les segments à l'intérieur des PSU.

**a. Fichier au niveau des foyers**

Ce fichier contiendra un dossier pour chaque foyer *sélectionné* pour la GATS. Chaque dossier doit comprendre un numéro d'identification unique du foyer, les probabilités géographiques de sélection (par exemple, les probabilités de sélection de la PSU et du segment), et la probabilité de sélection du foyer dans chaque segment sélectionné, toutes les variables des strates, et les identificateurs de grappe. Un indicateur des résultats de l'attribution des sexes (pour le groupe masculin ou féminin), avec la probabilité correspondant à l'attribution au groupe du sexe affecté, doit également être inclus dans ce fichier, si l'assignation des sexes est utilisée par le pays. Enfin, les pays peuvent envisager de créer une variable indicatrice (dérivée du code de décision finale du *questionnaire du foyer*) signalant l'admissibilité à la GATS, et une deuxième variable indicatrice montrant que la liste du foyer est terminée.

**b. Fichier au niveau individuel**

Ce fichier contiendra un dossier pour chaque individu sélectionné pour un entretien après une présélection réussie du foyer. Chaque dossier comprendra toutes les variables du fichier du foyer, plus un identifiant unique de l'individu et la probabilité de sélection de l'individu. En outre, les pays peuvent envisager de créer une variable indicatrice (dérivée du code de décision finale du *questionnaire individuel*) signalant l'admissibilité à la GATS de la personne sélectionnée, et une seconde variable indicatrice, basée sur le codage de l'arrangement final pour l'individu, montrant que la personne sélectionnée a complété le *questionnaire individuel*.

Des informations supplémentaires sur ces variables sont présentées dans les sections suivantes de ce manuel.

3. Les pays doivent créer des identifiants uniques pour chaque unité d'échantillonnage sélectionnée à chaque étape de la conception, et ces identifiants doivent être conservés dans le fichier d'analyse finale. Par exemple, si un échantillon d'individus est sélectionné en quatre étapes, le fichier d'analyse doit contenir les identifiants uniques au niveau de la PSU, du segment, du foyer, et de l'individu. Ces identifiants sont nécessaires pour contrôler la probabilité des sélections à chaque étape de la conception et peuvent également être nécessaires pour estimer les variances des estimations produites à partir de l'enquête. À l'exception de l'unité d'échantillonnage de la première étape et les identifiants de stratification, les pays ne doivent pas mettre ces identifiants dans un fichier accessible au public afin de préserver la confidentialité de chaque répondant à l'enquête.
4. Comme indiqué précédemment, l'échantillon doit être sélectionné en utilisant une conception de l'échantillon en plusieurs étapes, regroupées géographiquement. Les détails de ce type de conception de l'échantillon seront décrits dans les sections suivantes de ce manuel. On remarquera dans ces sections que pour la première étape de sélection, les pays doivent diviser leur pays en zones géographiques, compactes, qui ne se chevauchent pas. Pour la GATS, ces zones seront désignées sous les termes d'unités primaires d'échantillonnage, ou PSU. Ces zones peuvent être définies en utilisant des frontières naturelles, politiques, économiques ou géographiques telles que celles des États, comtés, provinces, villages, ou des secteurs de recensement. Notez que l'une des principales raisons de regroupement dans un plan d'échantillonnage est de minimiser les besoins de

---

**EXIGENCE :**

**Créer des identifiants uniques pour chaque unité d'échantillonnage à chaque étape de sélection.**

---

déplacement du personnel chargé du recueil des données en concentrant les travaux dans des poches géographiques, ou grappes.

5. En général, chaque PSU doit comprendre au moins 250 à 500 foyers (certains pays peuvent en avoir beaucoup plus), et au moins 100 PSU doivent être sélectionnées à la première étape de la conception. Cette taille minimale de l'échantillon des PSU devrait être suffisante pour fournir une représentation géographique adaptée au pays, tout en offrant le coût réduit souhaité du recueil de données, recherché dans une conception en grappes. Il est important de maintenir le nombre de PSU échantillonnées aussi grand que possible, car plus la taille de l'échantillon des PSU est faible, plus la taille moyenne de l'échantillon des répondants par PSU est importante, et les estimations de l'enquête relativement moins précises seront comparées à un échantillon de répondants non regroupé de la même taille.
6. Le nombre de PSU qui divisent un pays au premier stade de la conception doit être suffisamment grand (> 1 000), pour que le taux d'échantillonnage pour la sélection des PSU ne soit pas supérieur à environ 10 %. De grandes régions géographiques peuvent être utilisées pour former des strates, mais elles ne doivent pas être utilisées comme PSU. Des informations complémentaires sur la formation des PSU figurent dans le **chapitre 6** de ce manuel.

---

**EXIGENCE :**

**Sélectionnez au moins 100 PSU, quelle que soit la taille de l'échantillon global des répondants.**

---

## 4.2 Tailles de l'échantillon et la précision attendue

Les exigences et les recommandations relatives à d'autres mesures de la taille de l'échantillon sont basées sur les indicateurs de la qualité statistique ci-dessous, établis pour les résultats de la GATS :

1. La GATS doit être conçue pour produire des estimations qui répondent aux exigences de précision suivantes :
  - Les estimations calculées à l'échelon national, par degré d'urbanisation, par sexe et par croisement du sexe et du degré d'urbanisation doivent avoir un intervalle de confiance à 95 % avec une marge d'erreur de 3 points de pourcentage ou moins pour les taux de tabagisme de 40 %.
  - Les tailles d'échantillon pour le premier cycle de la GATS, ainsi que pour les éventuels cycles ultérieurs,<sup>2</sup> doivent être suffisamment grandes pour se conformer aux exigences suivantes des tests effectués pour détecter les différences entre les cycles d'enquêtes reposant sur des échantillons choisis indépendamment à chaque cycle :
    - *réduction du taux de prévalence du tabagisme de 40 % à 34 % avec une puissance de 80 % et à 33 % avec une puissance de 90 %, erreur de type I de 0,05, et une hypothèse alternative bilatérale.*
    - *réduction du taux d'exposition passive de 80 % à 70 % avec une puissance > 90 %, erreur de type I de 0,05, et une hypothèse alternative bilatérale.*

---

<sup>2</sup> Notez que le financement de la GATS est prévu un cycle à la fois.

- *Augmentation du taux de tentative de sevrage de 10 % à 20 % avec une puissance > 90 %, erreur de type I de 0,05, et une hypothèse alternative bilatérale.*

2. **L'effet de la conception** associée à une estimation particulière provenant d'une enquête est défini comme l'inflation de la variance due à des caractéristiques de conception complexe de l'enquête comme une pondération inégale et l'échantillonnage par grappes. Mathématiquement, il correspond au rapport de la variance d'une estimation basée sur la conception complexe d'une enquête sur la variance correspondante de la même taille de l'échantillon en utilisant un échantillonnage aléatoire simple. Bien qu'il soit théoriquement possible d'atteindre un effet de conception inférieur à un, dans la pratique, les caractéristiques d'une conception complexe d'une enquête nuisent presque toujours à la précision des estimations. Par conséquent, pour la plupart des études, les effets de conception seront supérieurs à un.

En supposant un effet de conception de 2,00 pour les estimations calculées au niveau national, par degré d'urbanisation, par sexe, et par classification croisée par sexe et degré d'urbanisation, la taille minimale de l'échantillon nécessaire pour se conformer aux normes de la qualité statistique de la GATS que nous venons de décrire est de 2 000 répondants. Si cette hypothèse est appliquée à chacun des quatre groupes définis par la classification croisée par degré d'urbanisation et par sexe, il en résulte un échantillon minimum recommandé de 8 000 répondants, établi comme la taille de l'échantillon global recommandée lorsque la classification croisée par sexe et par degré d'urbanisation à l'échelon national fixe les caractéristiques clés de rapport des résultats de la GATS. Des détails supplémentaires sur la façon dont cette taille d'échantillon de répondants a été calculée sont présentés dans l'**annexe A**.

3. Si un pays souhaite produire des estimations régionales et veut générer des estimations à la fois par sexe et par degré d'urbanisation au sein de chaque région, il est recommandé de concevoir l'enquête de façon à avoir 8 000 répondants dans chaque région (la moitié répartie dans des zones urbaines et l'autre moitié dans des zones rurales). Si un pays souhaite obtenir des estimations régionales, mais veut seulement rapporter les estimations par sexe et par degré d'urbanisation séparément, l'enquête doit être conçue pour comprendre 4 000 répondants dans chaque région.
4. Si un pays introduit des améliorations qui se traduiraient par un effet de conception attendu dépassant généralement 2,00, ou présente des données empiriques

---

**Recommandation :**  
une taille d'échantillon de 2 000 répondants est recommandée pour chaque caractéristique de rapport clé afin de répondre aux normes de qualité statistique de la GATS fixées pour l'estimation de la caractéristique. Cela correspond à une taille d'échantillon de répondants global recommandée d'au moins 8 000 pour les estimations à l'échelon national qui doivent être rapportées conjointement par sexe et par degré d'urbanisation. Interrogez au moins 8 000 répondants pour les estimations à l'échelon national, par sexe et par degré d'urbanisation.

---

---

**Recommandation :**  
une taille d'échantillon des répondants d'au moins 8 000 est recommandée pour chaque région lorsque les estimations doivent être rapportées conjointement par sexe et par degré d'urbanisation dans chaque région. Lorsque les estimations sont présentées séparément par sexe et par degré d'urbanisation dans chaque région, 4 000 répondants par région sont suffisants.

---

qui indiquent que la plupart des effets de conception sont susceptibles d'être supérieurs à 2,00, alors la taille de l'échantillon au sein du groupe doit être ajustée à la hausse en conséquence. Par exemple, si un pays assume un effet de conception de 2,5, leur conception doit être corrigée pour obtenir :

$$8,000 \cdot \frac{2.50}{2.00} = 10,000 \text{ répondants.}$$

5. La conception de l'enquête doit tenir compte correctement des niveaux prévus de non-réponse et de non admissibilité en déterminant le nombre de foyers à *sélectionner* afin d'obtenir le nombre recommandé de répondants indiqué ci-dessus. Des non-réponses et non admissibilités peuvent être observées tant au niveau des foyers qu'au niveau de l'individu sélectionné. Par exemple, un résident adulte du foyer invité à remplir une liste de résidents admissibles, ou une personne sélectionnée pour l'entretien peut refuser de participer (non-réponse). De même, un foyer sélectionné peut se révéler vacant ou une personne choisie peut indiquer qu'elle est âgée de moins de 15 ans, et par conséquent non-admissible. Une définition plus complète de la non-réponse et de la non-admissibilité figure dans le ***Manuel de l'enquêteur de terrain de la GATS***.

---

**EXIGENCE :**

**la conception de l'échantillon de la GATS doit répondre aux éventuelles non-réponses et non admissibilités à chaque étape.**

---

Si, par exemple, l'enquête d'un pays est conçue pour atteindre 8 000 répondants, on s'attend à observer ce qui suit :

Taux	Commentaire	Hypothèse
<b>Taux d'admissibilité des foyers</b>	Les foyers non admissibles comprennent les foyers vacants.	90 %
<b>Taux de présélection des foyers</b>	Comptes des foyers sans adulte, mineur âgé de 15 ans ou plus. Si la randomisation des sexes est utilisée, cela tient compte également des foyers sélectionnés pour le groupe des foyers masculins, lorsqu'aucun homme ne réside dans le foyer. De même cela comprend les foyers sélectionnés pour le groupe des foyers féminins lorsqu'aucune femme ne réside dans le foyer.	95 %
<b>Taux de réponse des foyers</b>	Tient compte des foyers admissibles présélectionnés où un <i>questionnaire du foyer</i> (la liste du foyer) n'est pas complet.	98 %
<b>Taux d'admissibilité individuelle</b>	Dans certains cas, un individu peut être interrogé pour la GATS et, plus tard jugé non-admissible, par exemple, s'il est en fait âgé de moins de 15 ans. Ce taux reflète ce type de cas.	98 %
<b>Taux de réponse individuel</b>	Comprend les résidents du foyer admissibles qui sont sélectionnés, mais ne remplissent pas le <i>questionnaire individuel</i> .	85 %
<b>Taux de réponse individuel des femmes</b>	Un taux de réponse individuel lorsque le foyer est sélectionné de manière aléatoire pour être un « foyer de sexe féminin ».	90 %
<b>Taux de réponse individuel des hommes</b>	Un taux de réponse individuel lorsque le foyer est sélectionné de manière aléatoire pour être un « foyer de sexe masculin » ; des études antérieures suggèrent que les hommes répondent à des taux légèrement inférieurs à celui des femmes.	80 %

Dans cet exemple, l'enquête doit être conçue pour obtenir un échantillon sélectionné de

$$\frac{4,000}{.90 \cdot .98} = 4,535 \text{ femmes sélectionnées,}$$

$$\frac{4,000}{.80 \cdot .98} = 5,102 \text{ hommes sélectionnés, et}$$

$$\frac{4,535 + 5,102}{.90 \cdot .98 \cdot .95} = 11,501 \text{ foyers sélectionnés.}$$

Des directives supplémentaires pour déterminer une taille d'échantillon appropriée à chaque étape de la conception de l'échantillon de la GATS figurent dans le **chapitre 10**.

### 4.3 Processus de recueil des données

Les exigences relatives aux procédures de recueil des données sont les suivantes :

1. Chaque pays doit recueillir les données en utilisant une méthode de collecte de données en face à face. En outre, chaque pays doit utiliser la partie fondamentale du questionnaire de la GATS présentée dans le manuel du **questionnaire principal contenant des questions facultatives de la GATS**.
2. Chaque pays doit élaborer des procédures de recueil de données pour parvenir à un *taux de réponse combiné* de 80 % ou plus. À chaque étape de la conception, le taux de réponse est défini comme le nombre total de répondants divisé par le nombre total de personnes admissibles. Le taux de réponse combiné est défini comme le produit des taux de réponse spécifiques au stade. Dans la GATS, le taux de réponse combiné est défini comme le produit des taux de réponse au niveau des foyers et au niveau individuel. Des informations spécifiques sur la définition des taux de réponse dans la GATS sont présentées dans le **chapitre 10** de ce manuel ainsi que dans **Assurance qualité de la GATS : directives et documentation**.
3. Dans certaines études, les substitutions d'échantillon sont autorisées. Cela désigne la pratique consistant à remplacer les non-répondants admissibles ou les répondants non-admissibles avec une autre unité choisie au hasard dans la base de sondage. Les substitutions d'échantillon peuvent conduire à des estimations biaisées et ne sont donc **pas autorisées** dans le contexte de la GATS.
4. La supplémentation de l'échantillon correspond à l'introduction d'un échantillon supplémentaire dans l'enquête. Ceci a pour objectif d'augmenter la taille de l'échantillon final des répondants lorsque les niveaux de non admissibilité et/ou de non-réponse sont plus élevés que prévu dans une strate ou une zone particulière. La supplémentation de l'échantillon est autorisée à condition qu'il reste suffisamment de temps suivant le calendrier du recueil des données pour permettre aux nouveaux cas de passer par toutes les phases de la collecte des données d'une manière normale.
5. Un maximum d'un répondant par foyer doit être sélectionné pour un entretien. Comme les ordinateurs de poche sont utilisés pour répertorier les foyers et le recueil des données dans toutes les opérations de collecte de données de la GATS, l'appareil de poche utilisera un générateur de nombres aléatoires pour sélectionner les répondants et affichera à l'écran le nom de la personne choisie. Le **chapitre 9** présente des détails sur l'échantillonnage dans les foyers.

---

**EXIGENCE :**

**obtenir un taux de réponse combinée d'au moins 80 %.**

---



---

**EXIGENCE :**

**aucune substitution de non-répondants admissibles ou de répondants non-admissibles.**

---

Si un des pays mettant en œuvre la GATS a des questions concernant les exigences de conception et les recommandations présentées dans cette section, ou si un pays souhaite demander la permission de s'en écarter, il doit prendre contact avec les CDC ou l'OMS.

## 5. Aperçu de la méthode d'échantillonnage

---

L'échantillon de la GATS doit être sélectionné en utilisant une conception en plusieurs étapes, en grappes géographiques pour assurer une couverture adéquate de l'ensemble de la population cible, tout en minimisant les coûts de la collecte des données. Les coûts sont réduits parce que la conception de l'échantillon par grappes géographiques minimisera les déplacements nécessaires aux enquêteurs pour visiter chaque foyer sélectionné. Cette conception est décrite dans le **document 5-1**. En bref, ce processus de sélection de l'échantillon doit procéder comme suit :

La ou les premières étapes de la conception impliquent la sélection d'unités régionales géopolitiques établies. Ce processus de sélection dépendra fortement du pays qui applique la GATS sur le terrain. Certains pays peuvent avoir des données suffisantes pour qu'une seule étape de sélection géographique soit nécessaire. D'autres pays peuvent avoir besoin de multiples étapes de sélection géographique, où chaque étape est conçue pour sous-sélectionner de façon aléatoire des zones géographiques à partir de zones plus vastes sélectionnées à un stade antérieur.

Le but ultime est de sélectionner au moins 100 unités d'échantillonnage primaires et au moins 400 unités géographiques définitives pour l'enquête. Les unités géographiques définitives, aussi appelées **segments de l'échantillon**, sont les plus petites zones géographiques sélectionnées pour la GATS. Notez que si un pays sélectionne des zones géographiques en une seule étape de sélection, il devrait échantillonner environ 400 segments directement à partir de la base de sondage de la zone initiale, car le segment est en fait l'unité primaire d'échantillonnage dans cette conception de l'enquête. En outre, des conceptions d'échantillon de certains pays peuvent faire appel à des étapes intermédiaires d'échantillonnage pour sélectionner les segments dans les PSU de l'échantillon.

Les zones géographiques doivent être sélectionnées à chaque étape avec une probabilité proportionnelle à la taille (PPS). La mesure de taille peut être calculée à partir d'un recensement récent et/ou de registres administratifs et, en général, doit être égale à la population totale admissible, âgée de 15 ans ou plus, ou au nombre de foyers dans la région. Si ces dénombrements ne sont pas disponibles, le dénombrement de la population totale correspondante ou des foyers peut être utilisé comme mesure de la taille.

Si des estimations régionales sont souhaitées et un pays veut comparer les zones rurales de la région avec les zones urbaines, il devra sélectionner au moins 100 PSU et au moins 400 segments à partir de chaque région, avec une probabilité proportionnelle à la taille. Si des estimations régionales sont souhaitées, mais un pays n'a pas besoin de comparer les zones rurales de la région avec les zones urbaines, il devra sélectionner au moins 50 PSU et au moins 200 segments à partir de chaque région, avec une probabilité proportionnelle à la taille.

---

### EXIGENCES :

**sélectionnez au moins 100 PSU géographiques et 400 segments.**

**Sélectionnez les zones géographiques en utilisant une technique aléatoire (par ex., systématique) PPS. La taille est le dénombrement de la population admissible (préférable) ou le total des foyers.**

**Sélectionnez au moins 100 PSU et 400 segments dans chaque région si le pays veut comparer les estimations par degré d'urbanisation dans la région. Sinon, sélectionnez 50 PSU et 200 segments par région.**

**Les foyers des segments doivent être sélectionnés au hasard avec une probabilité égale et sans remplacement.**

---

Une fois la phase finale de l'échantillonnage des zones géographiques terminée (c'est-à-dire l'étape d'échantillonnage des segments), des listes séparées de tous les foyers de chaque segment de l'échantillon doivent être dressées. Ces listes peuvent être établies à l'aide des registres administratifs, par le superviseur de terrain chargé de visiter les foyers sélectionnés dans la zone, ou par d'autres membres du personnel spécialement formés au dénombrement et à l'établissement de liste des foyers dans une zone sélectionnée. Le **Manuel de cartographie et d'enregistrement de la GATS** contient un exemple du type de liste des foyers qui doit être établie pour chaque zone sélectionnée, ainsi que des recommandations sur la façon de créer cette liste au cours d'une visite de la zone.

La liste des foyers dans chaque segment de l'échantillon servira de base de sondage pour la sélection des foyers dans chaque segment sélectionné. Les foyers doivent être sélectionnés au hasard à partir de la liste avec une probabilité égale et sans remplacement. L'échantillonnage aléatoire simple ou un échantillonnage systématique après un départ aléatoire peuvent être utilisés.

Après la sélection de l'échantillon de foyers dans chaque segment, et si le pays a décidé de procéder à une randomisation des sexes, chaque foyer échantillonné doit être assigné au hasard au groupe des foyers masculins ou au groupe des foyers féminins. Seuls les hommes admissibles à l'enquête des foyers du groupe masculin seront répertoriés dans la liste lors de la visite de présélection, et par la suite seuls les hommes des foyers du groupe masculin seront sélectionnés pour l'entretien de la GATS. De même, seules les femmes admissibles à l'enquête des foyers du groupe féminin seront répertoriées dans la liste lors de la visite de présélection, et par la suite seules les femmes des foyers du groupe féminin seront sélectionnées pour l'entretien de la GATS. (Pour plus d'informations sur la randomisation des sexes, consultez le **chapitre 2**.)

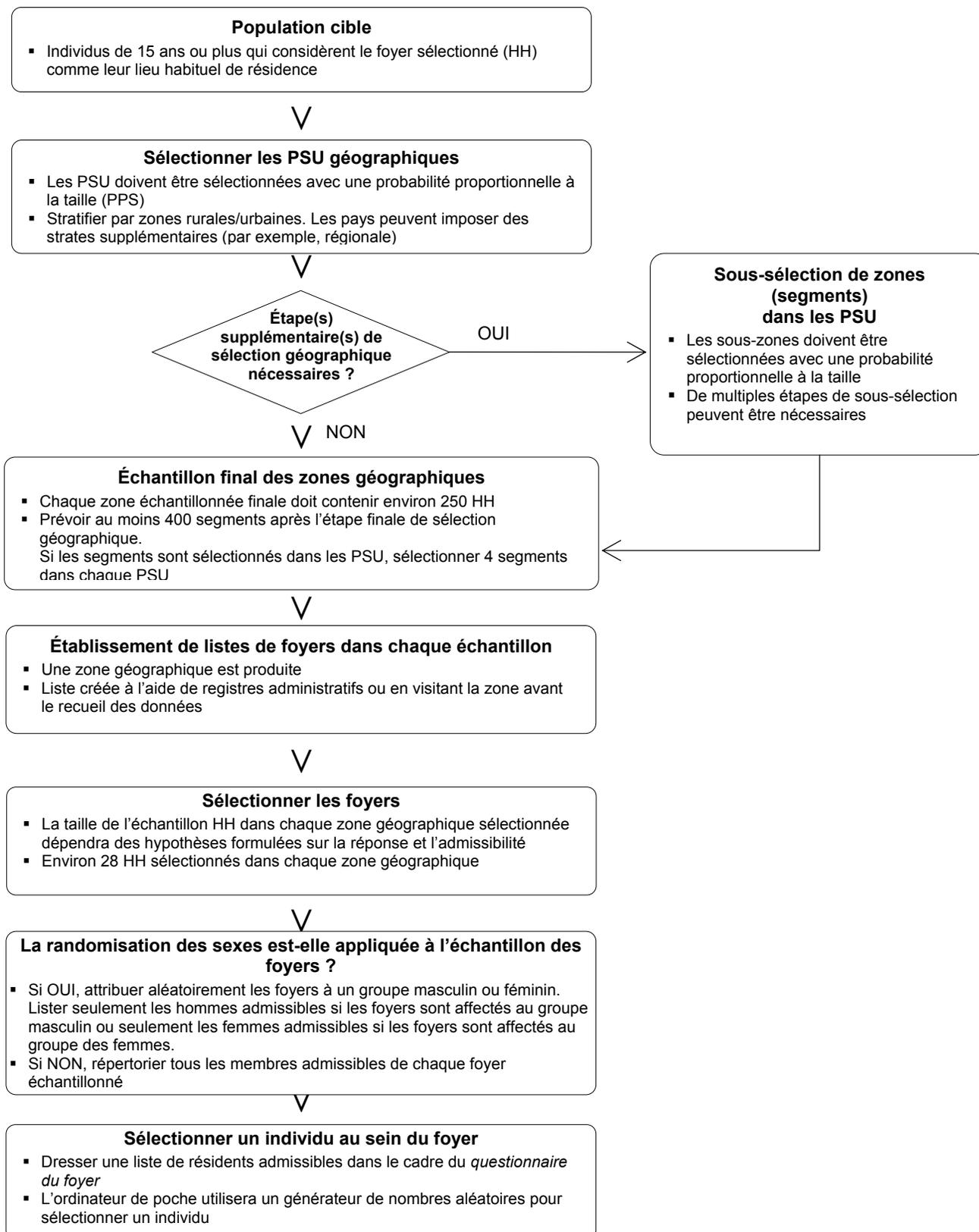
Les enquêteurs se rendront dans chaque foyer sélectionné. Si cette visite révèle que des foyers supplémentaires sont apparus dans une zone ou si l'enquêteur voit que le foyer sélectionné est en fait une structure contenant de multiples foyers (comme un immeuble), une opération de sous-échantillonnage peut être exigée par l'enquêteur, en consultation avec le superviseur de terrain. Ce type de sous-échantillonnage n'est généralement pas très souvent nécessaire.

Après la sélection d'une adresse de la liste et la confirmation qu'il s'agit d'un foyer unique, l'enquêteur se rendra au foyer et dressera une liste de tous les hommes et/ou femmes admissibles à l'enquête (suivant l'utilisation ou non de la randomisation des sexes) qui considèrent le foyer échantillonné comme étant leur lieu habituel de résidence. (Pour plus d'informations sur les critères de résidence et l'appartenance à la population cible de la GATS, voir la **section 3.1**.)

Pour les pays administrant l'enquête avec les ordinateurs de poche, l'appareil de poche utilisera un générateur de nombres aléatoires pour sélectionner un individu au sein du foyer. Seules ces personnes sélectionnées répondront au *questionnaire individuel* de la GATS car aucune substitution en cas de non-réponse n'est autorisée.

Chaque étape du processus d'échantillonnage est décrite plus en détail dans les prochaines sections.

## Document 5-1. Aperçu de la conception de l'échantillon de la GATS





## 6. Former des unités primaires d'échantillonnage (PSU)

---

La première tâche dans la conception de la GATS est de déterminer la définition la plus appropriée d'une unité d'échantillonnage primaire (PSU) et la base de sondage associée. Celles-ci seront fortement tributaires de la mise en œuvre de la GATS par le pays ; certains pays peuvent être plus petits et disposer de données de recensement récentes, leur permettant ainsi de former des PSU de taille suffisante pour que les enquêteurs les utilisent directement. En d'autres termes, l'échantillon de 400 segments peut être sélectionné en une étape de sélection.<sup>1</sup> D'autres pays peuvent avoir besoin de définir de plus grandes PSU puis utiliser les étapes ultérieures de sélection pour arriver à un ensemble de grappes géographiques qui sont de tailles approximativement égales et appropriées au travail des enquêteurs.

En général, l'objectif principal de la définition et de la sélection des PSU lors de la première étape est de réduire les coûts associés aux déplacements des enquêteurs de terrain. Les directives suivantes s'appliquent lors de la formation PSU :

Les PSU doivent être des zones clairement définies, compactes et mutuellement exclusives définies par des limites claires. Certains pays peuvent utiliser la surface des comtés ou les secteurs de recensement comme PSU. Ces entités comprennent, par exemple, les comtés/districts municipaux ou les secteurs de recensement.

Le nombre de PSU qui divisent un pays doit être suffisamment grand (> 1 000), pour que le taux d'échantillonnage pour la sélection des PSU ne soit pas supérieur à environ 10 %. De grandes régions géographiques peuvent être utilisées pour former des strates, mais elles **ne doivent pas** être utilisées comme PSU.

---

**Recommandation :**  
**attendez-vous à moins 1 000 PSU avec la première base de sondage.**

---

Il est conseillé aux pays qui ne sont pas subdivisés en secteurs de recensement d'envisager de sélectionner les zones géographiques pour la GATS en plusieurs étapes. La sélection de l'échantillon en plusieurs étapes entraînera la concentration de l'échantillon dans des sous-zones de petite taille à l'intérieur du pays, définies par les limites des PSU sélectionnées à la première étape avant la segmentation. Cela devrait permettre d'obtenir des coûts de recueil des données beaucoup plus bas comparés à ceux d'un échantillon de segments largement dispersés.

L'échantillon des PSU doit être sélectionné parmi une liste de toutes les PSU dans le pays. Par exemple, pour une enquête de la Fédération de Russie, la base de sondage des PSU doit inclure l'ensemble des îles de Russie au nord et à l'est de la Russie continentale (par exemple, îles de Nouvelle-Sibérie, île de Sakhaline) et pas seulement la Russie continentale.

---

<sup>1</sup>Si les segments peuvent être sélectionnés dans un pays en une étape de sélection, les segments sont en fait les unités primaires d'échantillonnage ou PSU de l'enquête.

Il peut y avoir des exceptions à cela. Des régions extrêmement éloignées ou des zones qui ne peuvent pas être visitées en raison de la guerre, de troubles politiques, etc., peuvent être exclues de la base des PSU, à condition qu'une proportion importante de la population du pays ne réside pas dans les zones exclues. Chaque pays doit définir explicitement les zones géographiques du pays incluses et exclues de leur base géographique initiale, et fournir une estimation du pourcentage de la population qui réside dans les régions à exclure. Ce pourcentage fournira une évaluation du potentiel de biais de couverture dans les estimations finales de la GATS qui seront générées à partir des données du pays.

En général, les zones géographiques finales sélectionnées pour la GATS doivent contenir environ 250 foyers. Si les PSU ne peuvent pas être formées pour contenir seulement environ 250 foyers, le pays doit envisager de sélectionner les zones géographiques en plusieurs étapes.

Une estimation du nombre de membres de la population cible (résidents, âgés de 15 ans ou plus) est nécessaire pour la sélection PPS des PSU. Cette estimation peut être obtenue à partir de sources telles que les registres administratifs ou le dernier recensement effectué dans le pays. Si une estimation de la population admissible n'est pas disponible, un pays peut utiliser une mesure fortement corrélée avec la population admissible à l'enquête telle que le nombre total de foyers dans la région. L'identification des données appropriées pour chaque PSU de la base est importante car elles seront la mesure de taille utilisée pour sélectionner les zones géographiques de la GATS.

La liste complète des PSU doit être facilement obtenue et manipulable, c'est-à-dire triable, pour la sélection de l'échantillon. Un formulaire de cette liste lisible par une machine est hautement préférable pour que les procédures de sélection des PSU puissent être mises en œuvre facilement et avec précision.

Certains pays peuvent souhaiter définir les PSU d'une manière qui s'écarte sensiblement des lignes directrices suggérées ci-dessus. Cet écart doit être discuté avec le point de contact du pays aux CDC ou à l'OMS avant que l'échantillon ne soit sélectionné. En général, pour garantir l'application de procédures d'échantillonnage normalisées, il est important que chaque pays qui participe à la GATS utilise les directives présentées ci-dessus et dans ce manuel.

---

**EXIGENCE :**

**la documentation de la conception doit inclure une estimation du pourcentage de la population du pays qui n'est pas inclus dans la zone cible, si des portions du pays doivent être exclues de la GATS.**

---

---

**Recommandation :**

**les zones géographiques finales sélectionnées (segments) doivent contenir environ 250 HH.**

---

---

**EXIGENCE :**

**toute déviation de ce plan de définition des PSU doit être discutée avec les CDC ou l'OMS.**

---

## 7. Première étape de l'échantillonnage : Sélection des PSU

---

Le processus d'échantillonnage recommandé pour les pays qui choisiront leur échantillon final de zones géographiques en une seule étape est légèrement différent de celui des pays qui sélectionneront les zones en plusieurs étapes. Des lignes directrices pour ces deux scénarios sont présentées ci-dessous.

### 7.1 Pays sélectionnant des zones géographiques (c.-à-d. les segments) en une seule étape de sélection

Les pays qui prévoient de sélectionner les zones géographiques finales (par exemple, les zones de segments) en une seule étape de sélection doivent suivre les directives suivantes :

1. Les segments doivent être formés pour qu'un petit nombre d'enquêteurs puissent y travailler. En général, un segment doit être formé de façon à être utilisé par deux à quatre enquêteurs. Si les segments sont trop grands, les économies de coûts du processus de recueil des données seront perdues puisque le temps de déplacement de l'enquêteur pourra devenir considérable.
2. Chaque segment doit être clairement désigné comme rural ou urbain. La désignation du degré d'urbanisation doit être utilisée pour définir deux strates explicites lors de la première étape de la conception. La stratification par région doit également être envisagée, surtout si l'on s'attend à des différences régionales dans l'usage du tabac. Des stratifications supplémentaires peuvent être imposées à ce premier stade de la conception en fonction des améliorations qu'un pays veut apporter à la conception (par exemple, la stratification par la composition ethnique).

**EXIGENCE :**  
stratifiez la base de segments par degré d'urbanisation.
3. Si un pays cherche à sélectionner ses zones géographiques en une seule étape de sélection, le nombre total d'unités à sélectionner doit être d'environ 400. En général, nous prévoyons qu'un pays sélectionnera 28 à 30 foyers par segment, de sorte que cela équivaudrait à une taille d'échantillon de segments d'environ 400, selon les hypothèses faites sur la non-réponse et la non-admissibilité. Des informations supplémentaires sur la détermination d'une taille d'échantillon de segments appropriée figurent dans le **chapitre 10**.
4. Comme la taille des segments utilisée pour la conception de l'échantillon est susceptible de varier considérablement, en particulier puisque les segments sont les PSU dans ce scénario de conception, les segments doivent être sélectionnés avec une probabilité proportionnelle à la mesure de la taille (PPS). En général, cette mesure de la taille doit être définie comme le nombre total prévu d'individus admissibles à l'enquête dans le segment. Comme indiqué dans le **chapitre 6**, si ces données ne sont pas disponibles, le pays peut utiliser une mesure de taille étroitement corrélée avec le nombre d'individus admissibles à l'enquête, telle que le nombre de foyers dans la zone.

5. L'échantillon de segments doit être sélectionné suivant une PPS. Il existe plusieurs techniques standard<sup>1</sup> qui peuvent être utilisées pour sélectionner les segments suivant une PPS, y compris l'algorithme Hanuray-Vijayan qui sélectionne des unités sans remplacement [voir Vijayan (1968)] et la méthode de Sampford qui sélectionne des unités sans remplacement [ceci est une extension de la méthode de Brewer, discutée dans Cochran (1977) et Sampford (1967)]. Plusieurs techniques de sélection séquentielle PPS de l'échantillon peuvent aussi être utilisées et sélectionnent des unités avec remplacement ou avec un remplacement minimal (voir par exemple, Chromy, 1979). La technique de sélection séquentielle, avec remplacement minimal PPS est similaire à une technique séquentielle standard, avec remplacement PPS sauf qu'elle garantit que les unités de la première étape (dans ce cas, les segments) ne seront pas sélectionnées une fois de trop. Plus précisément, si un segment contient des individus admissibles à l'enquête et que l'on souhaite sélectionner des segments, cette technique sélectionnera un segment avec une probabilité :

$$p_i^{(1)} = \frac{I \cdot N_i}{\sum_i N_i}$$

où  $N_i$  est la mesure de la taille,  $I$  est le nombre de segments choisis, et  $\sum_i N_i$  est la somme des mesures de taille pour tous les segments. Nous utilisons l'exposant (1) sur  $p_i^{(1)}$  pour indiquer qu'il s'agit de la première étape de sélection de l'échantillon.

De plus, avec cette méthode de sélection avec remplacement minimal, chaque segment  $i$  est assuré d'être sélectionné soit  $\text{int}(p_i^{(1)})$  ou  $\text{int}(p_i^{(1)}) + 1$  fois, où  $\text{int}(p_i^{(1)})$  désigne la partie entière de  $p_i^{(1)}$ .

Quelle que soit la méthode utilisée, il est important que chaque unité géographique de la base ait une probabilité non nulle de sélection et que les probabilités de sélection soient conservées dans le fichier de l'analyse finale. La pondération d'échantillonnage pour les répondants définitifs de la GATS sera créée à partir de ces probabilités de sélection<sup>2</sup>.

---

**EXIGENCE :**

**sélectionnez les segments avec une probabilité proportionnelle à la taille.**

**La taille est égale à la population ou aux HH.**

---



---

**EXIGENCE :**

**chaque segment doit avoir une probabilité non nulle de sélection.**

**Conservez les probabilités sur le fichier de données finales.**

---

6. Si un algorithme séquentiel PPS est utilisé, la base doit être triée avant la sélection de l'échantillon suivant un ensemble de variables, afin d'introduire un niveau de stratification implicite dans la

---

<sup>1</sup> Beaucoup de ces techniques de sélection de l'échantillon sont maintenant disponibles dans des logiciels statistiques commerciaux très utilisés. Par exemple, SAS a introduit la procédure SurveySelect dans SAS version 8 qui effectue plusieurs de ces algorithmes de sélection suivant une PPS.

<sup>2</sup> Si le taux d'échantillonnage est élevé (par ex., la sélection de 10 % ou plus d'une base) alors les probabilités conjointes de sélection doivent également être consignées pour estimer les variances de manière plus précise.

conception. En général, les variables utilisées pour trier la base doivent être étroitement corrélées avec les mesures principales des résultats concernés de l'enquête, en vue de bénéficier de la stratification implicite. Il est recommandé de trier l'échantillon suivant la géographie, si aucune autre variable pertinente n'existe dans la base de sondage.

7. Si une technique de sélection PPS avec remplacement (ou avec remplacement minime) est choisie pour sélectionner l'échantillon d'unités de la première étape (c'est-à-dire les segments), il y a des chances qu'un segment puisse être sélectionné plus d'une fois. En fait, si un algorithme séquentiel PPS avec remplacement minime est utilisé (voir par exemple, Chromy, 1979), un segment est assuré d'être sélectionné au moins une fois si la taille du segment est suffisamment grande. Si un segment est sélectionné à plusieurs reprises (par exemple  $\alpha_i$  fois), alors  $\alpha_i$

échantillons de foyers doivent être sélectionnés dans chaque segment sélectionné et le nombre  $1, \dots, \alpha_i$  de groupes de chaque foyer sélectionné doit être conservé dans le fichier d'analyse finale. Le processus de traitement de ces situations dans les dernières étapes de la conception est discuté dans le **chapitre 8**.

8. Comme indiqué précédemment, la GATS exige notamment que la première étape de sélection comprenne au moins 100 PSU. Il a également été mentionné précédemment que les pays qui souhaitent sélectionner des zones géographiques en une seule étape doivent sélectionner un nombre considérablement plus élevé d'unités de la première étape (au moins 400). Puisqu'une enquête dont les résultats seront présentés au niveau national simultanément par sexe et par degré d'urbanisation est conçue pour inclure 4 000 répondants dans les zones rurales et 4 000 répondants dans les zones urbaines, le pays concerné doit sélectionner la moitié de l'échantillon de segments (environ 200) à partir de la strate urbaine et l'autre moitié (environ 200) à partir de la strate rurale. Certains écarts à cette répartition de l'échantillon des segments dans les strates rurales et urbaines sont acceptables, surtout si une répartition égale augmente considérablement l'effet de conception attendu. Toutefois, quel que soit le cas, l'échantillon doit être conçu pour comprendre 4 000 répondants dans les zones rurales et 4 000 dans les zones urbaines. Dans les pays visant à présenter les résultats nationaux séparément (mais pas conjointement) par sexe et par degré d'urbanisation, seulement 200 segments seront nécessaires pour obtenir la taille de l'échantillon global recommandée de 4 000 répondants.

---

**Recommandation :**

**lorsque la taille de l'échantillon cible des répondants est de 8 000, sélectionnez environ 200 unités lors de la première étape (segments) dans les zones urbaines et environ 200 dans les zones rurales.**

---

---

**Recommandation**

**lorsque la taille de l'échantillon cible des répondants est de 4 000, sélectionnez 200 segments. L'enquête doit être conçue pour obtenir 4 000 répondants dans les zones urbaines et 4 000 dans les zones rurales.**

---

## 7.2 Pays sélectionnant des zones géographiques en plusieurs étapes de sélection

Les pays qui prévoient de sélectionner les zones géographiques en plusieurs étapes de sélection doivent suivre la plupart des directives mentionnées ci-dessus. En résumé :

1. Les PSU seront généralement définies par de plus grandes zones géographiques. Pour la GATS, la formation d'au moins 1 000 PSU dans un pays est recommandée, et il est nécessaire qu'un échantillon aléatoire d'au moins 100 PSU soit sélectionné afin d'assurer la couverture géographique adéquate du pays.
2. La ou les variables de stratification utilisées à la première étape de la collecte dépendront du pays. Dans certains cas, il sera possible de désigner clairement une PSU comme étant rurale ou urbaine. Si cela est le cas, alors environ le même nombre de PSU dans les strates définies comme étant rurales et dans les strates urbaines doit être sélectionné de manière à obtenir au mieux les 4 000 répondants souhaités dans les zones rurales et 4 000 répondants dans les zones urbaines. Si les PSU ne peuvent pas être facilement classées en zones rurales et urbaines, la stratification en fonction du degré d'urbanisation peut être imposée à un stade ultérieur de la sélection géographique. Dans tous les cas, l'enquête doit être conçue pour comprendre 4 000 répondants dans les zones rurales et 4 000 répondants dans les zones urbaines.
3. Comme décrit dans le **paragraphe 5** de la **section 7.1**, les PSU doivent être sélectionnées avec une probabilité proportionnelle à la taille (PPS). La mesure de taille utilisée pour sélectionner les PSU doit être une estimation du nombre total des personnes admissibles à l'enquête qui résident dans le PSU. Si ces données ne sont pas disponibles, d'autres mesures de taille peuvent être utilisées, étroitement corrélées avec le dénombrement de la population, par exemple le nombre total de foyers dans la PSU.
4. Si un algorithme séquentiel PPS est utilisé pour sélectionner l'échantillon de la première étape, la base doit être triée avant de sélectionner l'échantillon en fonction d'un ensemble de variables afin d'introduire un niveau de stratification implicite dans la conception. En général, les variables utilisées pour trier la base doivent être étroitement corrélées avec les mesures principales des résultats concernés de l'enquête, en vue de bénéficier de la stratification implicite. Il est recommandé de trier l'échantillon suivant la géographie (par ex., par provinces), si aucune autre variable pertinente n'existe dans la base de sondage.
5. Si une technique de sélection PPS avec remplacement ou avec remplacement minime est utilisée pour sélectionner l'échantillon, il y a des chances qu'une PSU puisse être sélectionnée plus d'une fois. En fait, si l'algorithme PPS séquentiel avec remplacement minimal est utilisé, une PSU est assurée d'être sélectionnée au moins une fois si  $p_i^{(1)} > 1$ . Si une PSU est sélectionnée à plusieurs reprises (par exemple  $\alpha_i$  fois), alors le nombre de sous-zones géographiques sélectionnées à

---

### Recommandation :

**autant que possible, choisissez la moitié des PSU dans des zones urbaines et l'autre moitié dans des zones rurales.**

---

---

### EXIGENCE

**sélectionnez les PSU et les segments avec une probabilité proportionnelle à la taille. La mesure préférable de la taille est égale au nombre d'individus âgés de 15 ans ou plus, sinon la taille est égale au nombre de HH.**

---

l'étape suivante de sélection de la PSU doit être  $\alpha_i$  fois le nombre sélectionné dans une PSU qui a été choisie une seule fois.

6. Lorsque les PSU sont des zones géographiques relativement vastes, des sous-échantillonnages géographiques devront être effectués, afin d'obtenir un échantillon de zones géographiques qui sont de tailles convenables pour qu'un petit nombre d'enquêteurs puissent y travailler en se déplaçant un minimum. En général, l'unité d'échantillonnage géographique définitive (à savoir, le segment) doit contenir environ 250 foyers.

---

**Recommandation :**  
l'unité d'échantillonnage géographique définitive, à savoir, le segment, doit contenir environ 250 foyers.

---

7. Le nombre d'étapes de sélection dans une PSU sélectionnée peut varier, selon la taille de la PSU. À des fins de discussion, nous supposons qu'à l'étape finale de sélection géographique, une zone sera sélectionnée et celle-ci est généralement désignée sous le terme de **segment**.

8. Les segments d'une PSU doivent être choisis en utilisant une approche PPS qui peut être similaire à l'approche PPS utilisée pour sélectionner les PSU. La définition de la mesure de taille utilisée pour sélectionner les segments doit être équivalente à la définition de la mesure de taille utilisée pour former les PSU.

---

**EXIGENCE :**  
sélectionnez des sous-zones dans les PSU (à savoir les segments) avec une probabilité proportionnelle à la taille. La mesure préférable de la taille est égale au nombre d'individus âgés de 15 ans ou plus, sinon la taille est égale au nombre de HH.

---

Pour voir l'avantage d'utiliser la même définition de mesure de taille à l'étape de sélection des PSU et des segments, supposons qu'une PSU  $i$  a  $k = 1, \dots, \Omega_i$  segments. Par

ailleurs, supposons que la base de sondage indique que la PSU  $i$  a  $N_i = \sum_k N_{ik}$  personnes âgées

de 15 ans ou plus. Avec la plupart des approches standard de sélection PPS, la PSU sera choisie avec une probabilité  $\frac{I \cdot N_i}{\sum_i N_i}$  où  $I$  est le nombre de PSU sélectionnées. Supposons que nous

souhaitons sélectionner  $K_i$  segments dans la PSU  $i$  et que les segments sont sélectionnés avec une probabilité proportionnelle à la  $N_{ik}$ 's. Pour la plupart des conceptions nous nous attendrions à

ce que  $K_i$  soit égal à 4. Alors la probabilité de sélection du segment est  $\frac{K_i \cdot N_{ik}}{\sum_k N_{ik}} = \frac{K_i \cdot N_{ik}}{N_i}$  et la

probabilité inconditionnelle de sélection à partir des deux étapes se réduit à :

$$\frac{I \cdot N_i}{\sum_i N_i} \cdot \frac{K_i \cdot N_{ik}}{\sum_k N_{ik}} = \frac{I \cdot N_i}{\sum_i N_i} \cdot \frac{K_i \cdot N_{ik}}{N_i} = \frac{I \cdot K_i \cdot N_{ik}}{\sum_i N_i}$$

Remarquez comment le terme  $N_i$  du numérateur de la probabilité de sélection de la PSU s'annule avec le terme  $N_i$  du dénominateur de la probabilité de sélection du segment. En raison de cette

annulation, la probabilité finale inconditionnelle de sélection du segment est équivalente à ce qui aurait été obtenu si le segment avait été sélectionné avec une conception en une seule étape (au lieu d'une conception en deux étapes).

- 
9. Le nombre de sous-zones à sélectionner dans une PSU dépendra de la taille de la PSU. Par exemple, si une PSU est sélectionnée au hasard à quatre reprises durant le processus de sélection, alors le nombre de sous-zones sélectionnées de la PSU doit être quatre fois le nombre de sous-zones sélectionnées à partir d'une PSU choisie une fois. En général, il faudra toujours choisir au moins deux sous-zones à partir d'une zone précédemment sélectionnée. Si les zones géographiques sont sélectionnées en deux étapes et 100 PSU sont sélectionnées à la première étape, en moyenne, il devrait y avoir quatre segments sélectionnés à partir de chaque PSU.
- 
- Recommandation :**  
**sélectionnez en moyenne, quatre segments à partir de chaque PSU, avec un minimum de deux.**
- 
10. Après toutes les étapes de sélection géographique, le nombre total de segments sélectionnés à partir des zones urbaines doit être égal au nombre total de segments sélectionnés à partir des zones rurales. En général, environ 200 segments devraient être sélectionnés dans les zones urbaines et 200 segments dans les zones rurales.
- 
- Recommandation :**  
**après toutes les étapes de sélection géographique, dans un pays où les estimations nationales doivent être présentées conjointement par sexe et degré d'urbanisation, il doit y avoir environ 200 segments sélectionnés à partir des zones urbaines et environ 200 segments sélectionnés à partir des zones rurales.**
- 
11. En général, il est prévu qu'un pays sélectionne environ 28 à 30 foyers dans chaque segment de sorte que cela équivaldrait à une taille totale d'échantillon de segments d'environ 400. Des informations supplémentaires sur la détermination d'une taille d'échantillon de segments appropriée figurent dans le **chapitre 10**.
-

## 8. Sélectionner les foyers

---

Une fois que les zones géographiques sont sélectionnées pour l'enquête de la GATS, la prochaine étape de la conception de la GATS consiste à sélectionner des foyers à partir de l'échantillon final de segments préalablement choisi.

Il y a plusieurs façons de développer une base de sondage de foyers appropriée et de sélectionner l'échantillon de foyers pour ce stade de la conception. Certains pays peuvent trouver qu'une méthode fonctionne le mieux pour tous leurs segments échantillonnés, tandis que d'autres pays peuvent avoir besoin d'utiliser un mélange de méthodes, peut-être une méthode pour les segments situés dans les zones urbaines et une seconde méthode pour les segments situés dans les zones rurales. Certains pays peuvent aussi souhaiter utiliser une base de sondage maîtresse mise à jour qui peut avoir été développée pour le dernier recensement, une enquête récente sur la main d'œuvre, ou d'autres enquêtes nationales similaires. Les méthodes utilisées par un pays dépendront des données disponibles sur les foyers pour un segment, la qualité de ces données, et la facilité avec laquelle les enquêteurs et les superviseurs peuvent corriger des erreurs dans ces données lors de leur première visite au segment. En général, l'objectif principal de cette étape du processus d'échantillonnage est d'établir la liste la plus complète et précise de tous les foyers actuellement situés dans chaque segment de l'échantillon.

---

### EXIGENCE :

**la conception de l'échantillon final d'un pays doit comprendre une description du processus de développement de la base des foyers, et des indications de mesures qui seront prises pour garantir que les listes de foyers seront aussi complètes et précises que possible.**

---

Voici les descriptions des méthodes qu'un pays pourrait utiliser pour développer une nouvelle base de sondage des foyers pour chaque segment, et pour sélectionner l'échantillon des foyers. La conception de l'échantillon final d'un pays doit inclure une description détaillée de la façon dont un pays compte former une base de sondage adaptée à la sélection des foyers, dont il remédiera aux imperfections de la base qui ne sont pas représentées sur les listes existantes (telles que les foyers nouvellement construits dans la zone), et de la manière dont l'échantillon sera choisi dans chaque région. Ce chapitre présente également des recommandations sur la façon de traiter les foyers qui sont découverts par l'enquêteur ou le superviseur et sont absents de la liste des foyers (voir la **section 8.5**).

### 8.1 Créer une liste des foyers pour un segment échantillonné en visitant le segment avant de sélectionner l'échantillon de foyers

La méthode préférée de sélection d'un échantillon de foyers dans un segment consiste à créer d'abord une liste des foyers du segment. Cette liste des foyers, qui devient la base de sondage pour la phase de sélection des foyers de la conception, est développée en visitant la zone géographique et en énumérant tous les foyers qui apparaissent dans les limites géographiques du segment avant le début du recueil de données. Cette méthode génère une base de sondage des foyers qui est à jour et

---

### Recommandation :

**la méthode préférée de création d'une liste des foyers pour chaque segment consiste à avoir une liste énumérative de tous les foyers de la zone avant la collecte de données.**

---

particulièrement bénéfique pour les zones qui connaissent une croissance importante. S'il est déterminé qu'un segment a un bien plus grand nombre de foyers que prévu initialement, il est recommandé que l'équipe de conception de l'échantillon de la GATS du pays commence une autre étape de sélection géographique et sous-sélectionne une zone au sein du segment. Encore une fois, le but ultime est d'avoir des segments qui ne sont ni trop petits ni trop grands ; ils doivent contenir environ 250 foyers.

Lors de la visite de dénombrement où un agent de terrain créera la liste des foyers dans le segment échantillonné, les foyers de la liste doivent être classés dans un ordre *de trajet de déplacement continu* comme indiqué dans le **Manuel de cartographie et d'enregistrement de la GATS**. Présenter les foyers suivant l'ordre du trajet de déplacement continu est très important car l'ordre peut être utilisé par les enquêteurs et les superviseurs pour identifier les éventuels foyers manquant sur la liste des foyers. Si une zone est située dans une région extrêmement éloignée ou inaccessible du pays, ou si un trajet de déplacement continu logique n'existe pas pour le segment, le pays peut demander au recenseur d'enregistrer la localisation des foyers sur une carte du segment.

---

**Recommandation :**  
chaque fois qu'une nouvelle liste des foyers est créée au sein de chaque segment de l'échantillon de la GATS, il est préférable que les foyers soient listés suivant un ordre de trajet de déplacement continu.

---

Il faut remarquer qu'afin de traiter adéquatement les foyers qui n'ont pas été répertoriés dans un segment, ce qui peut se produire indépendamment de la méthode utilisée pour créer la liste des foyers, la liste des foyers doit être :

- une liste des foyers classés dans un ordre logique, suivant un trajet de déplacement continu à travers le segment, ou
- les foyers du segment doivent être notés sur une carte.

---

**EXIGENCE :**  
les foyers doivent être listés sur une feuille suivant un ordre de trajet de déplacement continu ou notés sur une carte du segment.

---

Les méthodes proposées pour le traitement des foyers manquants sur la liste des foyers sont présentées dans la **section 8.5**.

## **8.2 Créer une liste des foyers pour un segment échantillonné à l'aide des registres administratifs**

Une deuxième méthode de création de liste des foyers pour un segment de l'échantillon consiste à utiliser des registres administratifs, comme ceux d'un recensement antérieur. Cette méthode de création de liste des foyers peut être moins coûteuse que le développement d'une liste par l'intermédiaire d'une visite personnelle dans la zone effectuée par un agent recenseur. Cependant, la qualité de la liste peut être très inférieure, surtout si les registres administratifs utilisés pour créer la liste sont anciens. Le pays doit contacter son représentant de l'OMS/des CDC s'il a l'intention d'utiliser un répertoire au niveau individuel pour la création de sa liste des foyers. Certains des aspects à considérer lors de la création d'une base de sondage à l'aide des registres administratifs comprennent :

1. Comme noté précédemment, l'âge des données administratives est à prendre en compte. Des données anciennes peuvent produire des listes des foyers contenant de nombreuses inexactitudes,

en particulier pour les zones qui sont sujettes à des changements, telles que les zones urbaines en expansion.

2. Tenez compte de la facilité avec laquelle on peut faire une liste des foyers qui représentent un trajet logique de déplacement à travers la zone et/ou la facilité avec laquelle on peut noter les foyers sur une carte du segment. Disposer d'un trajet logique de déplacement ou d'une carte des foyers sera important pour le personnel de terrain, en particulier parce qu'ils sont susceptibles de trouver de nombreuses unités de logements manquées dans la zone.
3. Si plusieurs sources administratives sont utilisées pour faire une liste des foyers, on doit considérer la facilité avec laquelle il est possible de soustraire les doublons des sources. Une inspection visuelle peut être nécessaire pour ce faire. Par exemple, une source peut lister une maison à l'adresse « 110, boulevard central, app. D » et une seconde source peut lister une maison à l'adresse « 110 bd central, D. »

### 8.3 Sélection de l'échantillon de foyers

Dans les deux sections précédentes, il a été mentionné que la liste des foyers pouvait être créée à partir d'une liste distincte des foyers dressée sur le terrain ou de registres administratifs existants, tels que ceux du recensement le plus récent. Dans les deux cas, une liste des foyers (ou des adresses des foyers) doit être le résultat final. Un échantillon aléatoire simple ou un échantillon systématique des foyers du segment est ensuite choisi séparément dans chaque segment de l'échantillon. Le nombre recommandé de foyers sélectionnés dans chaque segment pour les deux méthodes est d'environ 28. Des détails supplémentaires sur la détermination de la taille de l'échantillon figurent dans le **chapitre 10**.

---

**EXIGENCE :**

**l'échantillonnage à partir d'une liste des foyers doit se faire en utilisant une méthode aléatoire simple ou systématique, sans remplacement.**

---

L'échantillonnage aléatoire simple implique une sélection aléatoire sans remplacement à partir de la liste, tandis que l'échantillonnage systématique consiste en l'échantillonnage de tous les *K*-ème foyers sur la liste après un point de départ aléatoire, où l'intervalle d'échantillonnage (*K*) pour chaque segment est basé sur le rapport du nombre total de foyers dans le segment sur le nombre désigné de foyers sélectionnés pour le segment (par exemple, 28). La plupart des manuels d'échantillonnage décrivent la sélection de l'échantillon par ces deux méthodes. Atteindre le nombre désigné de foyers sélectionnés est plus difficile avec un échantillonnage systématique lorsque le nombre total de foyers du segment n'est pas un multiple entier de la taille désignée de l'échantillon des foyers du segment. Kish (1965, section 7.5) décrit une façon de traiter ce problème.

La probabilité de sélection conditionnelle pour tout foyer sélectionné dans chaque segment dépend de la méthode de sélection utilisée. *Si un échantillonnage aléatoire simple est utilisé*, cette probabilité est simplement le taux d'échantillonnage des foyers (c'est-à-dire, le rapport de la taille de l'échantillon des foyers sur le nombre total des foyers de la base de liste). *Lorsque l'échantillonnage systématique est utilisé* pour sélectionner les foyers, cette probabilité de sélection est un divisé par l'intervalle de sélection (*K*). Consultez le **Manuel de pondération de l'échantillon de la GATS** pour plus de détails sur ces calculs et d'autres probabilités de sélection de la GATS.

## 8.4 Affectation facultative des foyers à des groupes de foyers masculins et féminins

Certains pays pourront envisager une modification optionnelle de la conception consistant à répartir aléatoirement les foyers de l'échantillon en deux groupes. Un ensemble sera affecté au groupe de foyers masculins et l'échantillon restant sera affecté au groupe des foyers féminins. Les foyers affectés au groupe masculin n'auront que des hommes admissibles, âgés de 15 ans ou plus, répertoriés pendant la partie de présélection de l'entretien, et par conséquent seul un homme pourra être sélectionné au sein du foyer pour l'entretien individuel de la GATS. De même, les foyers affectés au groupe féminin n'auront que des femmes admissibles, âgées de 15 ans ou plus, répertoriées pendant la partie de présélection de l'entretien, par conséquent seule une femme pourra être sélectionnée au sein du foyer pour l'entretien individuel de la GATS. Cette répartition de l'échantillon dans des groupes de foyers masculins et féminins peut être effectuée pour les raisons suivantes :

1. Dans certains pays, le recrutement des répondants sera moins compliqué et donc les taux de participation seront plus élevés si le sexe de l'enquêteur de terrain correspond à celui du répondant. Cette randomisation des sexes de l'échantillon des foyers au sein de chaque segment dans des groupes masculins et féminins permet au pays d'obtenir cet appariement des sexes.
2. Dans certains pays où les taux de prévalence du tabagisme féminin sont considérablement inférieurs aux taux correspondants masculins, il peut être nécessaire d'échantillonner les femmes de manière disproportionnée, à un taux plus élevé que les hommes. Ce genre de sur-échantillonnage en fonction du sexe s'obtient le plus facilement par randomisation des sexes des foyers.
3. Sur la base de l'expérience et des études antérieures, les hommes ont tendance à présenter des taux de réponse inférieurs à ceux des femmes ; cette répartition de l'échantillon permet à un pays de tenir compte de façon efficace de cette différence de propension à répondre en assignant aléatoirement un nombre légèrement plus élevé de foyers au groupe masculin par rapport au groupe féminin.

Nous recommandons d'effectuer l'affectation aléatoire des foyers échantillonnés à des groupes masculins et féminins en utilisant un processus d'échantillonnage aléatoire simple. Par exemple, comme indiqué dans le paragraphe 3 ci-dessus, dans la plupart des pays on peut s'attendre à ce que les hommes répondent à des taux inférieurs à ceux des femmes, par conséquent, un plus grand nombre de foyers échantillonnés doit être affecté au groupe masculin par rapport au groupe féminin.

Supposons que  $\bar{H}_{ik}$  foyers sont sélectionnés dans un segment  $k$  situé dans la PSU  $i$  et que nous souhaitons attribuer  $\bar{H}_{male,ik}$  de ces foyers au groupe masculin et  $\bar{H}_{female,ik}$  au groupe féminin. Donc  $\bar{H}_{male,ik} + \bar{H}_{female,ik} = \bar{H}_{ik}$ . Ensuite, un échantillon

---

**Option de conception :**  
attribuer aléatoirement les foyers sélectionnés au groupe des foyers masculins ou au groupe des foyers féminins.

---

---

**Recommandation :**  
la distribution aléatoire de foyers échantillonnés dans le groupe masculin/féminin doit se faire en utilisant une méthode d'échantillonnage aléatoire simple.

---

---

**EXIGENCE :**  
la probabilité d'attribution d'un foyer au groupe masculin ou féminin doit être consignée dans le fichier d'analyse utilisé pour créer la pondération de

---

aléatoire simple (sans remplacement) de  $\bar{H}_{male,ik}$  doit être sélectionné à partir des  $\bar{H}_{ik}$  foyers et affecté au groupe masculin. Les  $\bar{H}_{ik} - \bar{H}_{male,ik} = \bar{H}_{female,ik}$  restants doivent alors être affectés au groupe féminin. Par exemple, supposons que 28 foyers sont sélectionnés dans un segment et 15 de ces foyers doivent être affectés au groupe masculin. Alors 15 foyers doivent être choisis au hasard parmi les 28 et doivent recevoir une probabilité de sélection dans le groupe masculin de 15/28. Les 13 foyers restants seront affectés au groupe féminin et recevront une probabilité de sélection égale à 13/28.

Comme avec toutes les autres étapes de la sélection, la probabilité d'assigner un foyer au groupe masculin ou féminin doit être consignée dans le fichier d'analyse finale de la GATS.

### 8.5 Préparation d'un fichier électronique d'échantillon pour l'administration sur l'ordinateur de poche

Quand les ordinateurs de poche sont utilisés pour le recueil des données, une version électronique de l'échantillon doit être préparée. Ce fichier doit contenir les informations nécessaires pour charger les cas dans l'ordinateur de poche. L'équipe d'échantillonnage ou le personnel de soutien informatique doit être prêt à créer un fichier dérivé de l'échantillon maître qui contiendra les informations suivantes pour chaque unité d'habitation de l'enquête :

- Identifiants (numéro de cas, PSU, région)
- Informations sur le localisateur
- Désignation du groupe de randomisation de sexe masculin ou féminin (si l'option de conception est utilisée)
- Identifiant de l'enquêteur de terrain (FI) qui recevra les cas (facultatif)

Consultez le **Guide du programmeur du système général de l'enquête de la GATS** pour plus d'informations sur l'organisation et le contenu exact nécessaire pour créer ce fichier.

### 8.6 Traiter les problèmes de foyers manqués et autres erreurs apparentes dans la liste des foyers

Tous les éléments suivants peuvent entraîner l'omission d'adresses de foyers existants, sur la liste utilisée pour choisir l'échantillon de foyers dans le segment de la GATS : (i) la base peut être datée de plusieurs années de sorte que les habitations résidentielles nouvellement construites sont manquantes, (ii) les registres administratifs d'adresses peuvent exclure des segments entiers de la population résidentielle (par exemple, les personnes qui ne sont pas inscrites sur les listes électorales), et (iii) une liste nouvellement créée peut être d'une qualité inconnue si la liste des foyers a été établie par du personnel manquant d'expérience préalable ou de formation dans le dénombrement des foyers.

Dans tous les cas, l'utilisation de bases de liste avec un taux considérablement élevé de couverture incomplète peut conduire à des estimations biaisées à partir des échantillons de la GATS, en particulier lorsque le comportement tabagique des individus vivant dans les foyers inclus dans ces bases diffère

---

**Recommandation :**  
chaque pays de la GATS doit élaborer une stratégie pour faire face à la probabilité que les bases de sondage utilisées pour l'échantillonnage des foyers soient au moins partiellement incomplètes.

---

collectivement du comportement des personnes vivant dans les foyers absents de la base de liste utilisée pour la sélection des foyers. Des stratégies de prévention sont disponibles pour minimiser l'impact des biais de couverture de la base. Cette section présente un résumé du processus qui doit être utilisé pour traiter les ménages manqués ou d'autres erreurs trouvées dans la liste des foyers au moment où l'enquêteur de terrain et le superviseur de terrain visitent le segment juste avant de contacter chaque foyer échantillonné. Un résumé de ce processus est présenté dans le **document 8-1**.

Après la sélection de l'échantillon initial des foyers, un enquêteur ou un superviseur doit visiter la zone avant de prendre tout contact avec un foyer. L'enquêteur et/ou le superviseur doivent avoir avec eux une copie de la liste des foyers. Lors de cette visite initiale, l'enquêteur et/ou le superviseur doivent effectuer les opérations suivantes :

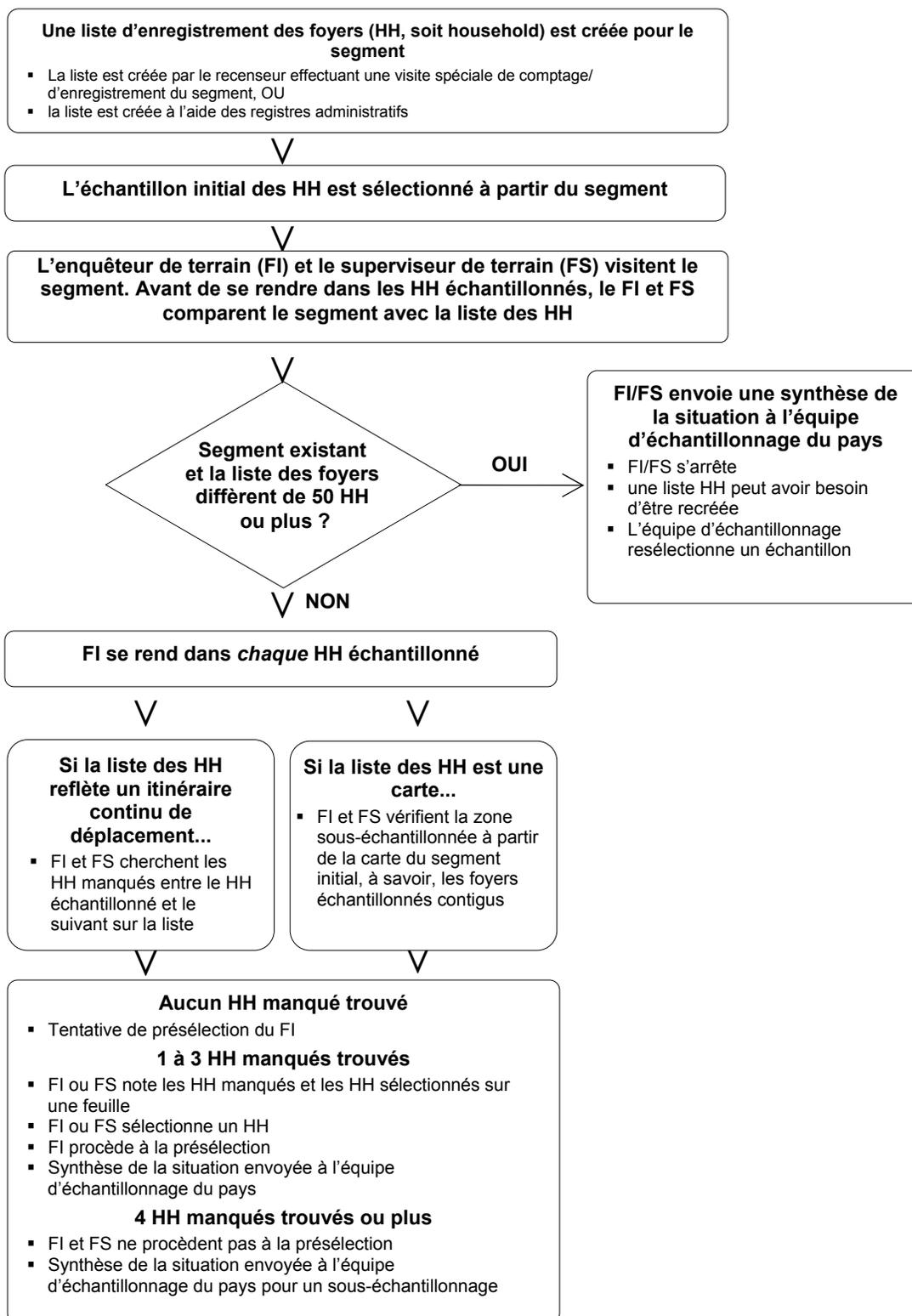
1. Identifier tous les grands ensembles de logements situés dans les limites du segment, mais qui ne sont pas inclus dans la liste des foyers à partir de laquelle l'échantillon de la GATS a été sélectionné. Si cela se produit, l'enquêteur ou le superviseur doit consigner les unités de logement manquées et envoyer l'information à l'équipe de conception de l'échantillon avant de se rendre dans les foyers. Bien que cela dépende de la zone échantillonnée, en général, si un enquêteur ou un superviseur trouve 50 unités de logement ou plus, il doit envoyer une liste de la zone mise à jour à l'équipe de conception de l'échantillon du pays avant de procéder à toute opération. Si un enquêteur ou un superviseur trouve moins de 50 unités de logement, l'envoi d'une liste mise à jour peut ne pas s'avérer nécessaire. La technique des intervalles semi-ouverts (décrite ci-dessous) peut être suffisante pour représenter moins de 50 foyers manqués dans un segment.

---

**Recommandation :**  
**si l'enquêteur et/ou le superviseur trouvent 50 foyers manqués ou plus dans un segment, la liste des foyers doit être recréée et l'échantillon resélectionné.**

---

**Document 8-1. Résumé du processus de résolution du problème des adresses de foyers manqués dans la liste d'enregistrement des foyers**



2. Pour chaque foyer échantillonné, l'enquêteur et le superviseur doivent employer la technique couramment utilisée des **intervalles semi-ouverts** pour identifier les foyers qui pourraient avoir été omis sur la liste des foyers de la zone. Cette méthode ne peut être utilisée que si la liste des foyers reflète une trajectoire continue de déplacement comme mentionné dans la **section 8.1**.

En résumé, pour chaque foyer échantillonné, l'enquêteur ou le superviseur doivent regarder si des foyers existent entre le foyer échantillonné et le foyer qui suit sur la liste des foyers. Si un à trois foyers sont manquants, l'enquêteur ou le superviseur doivent enregistrer l'unité échantillonnée et les différents foyers omis sur une feuille séparée. Ils doivent ensuite sélectionner un foyer au hasard dans cet ensemble. Le foyer sélectionné à l'origine peut être sélectionné à nouveau, ou l'un des foyers manqués peut être sélectionné. Si plus de trois foyers sont manquants, cette information doit être envoyée à l'équipe de conception de l'échantillon du pays avant de prendre tout contact avec un foyer de zone. L'équipe de conception de l'échantillon décidera si l'échantillonnage aléatoire d'une unité est suffisant pour résoudre le problème des foyers manqués, ou si la sélection d'un échantillon plus important se justifie.

Notez que si un à trois foyers sont omis dans la liste des foyers, la feuille que l'enquêteur et/ou le superviseur ont utilisée pour sélectionner un foyer doit être envoyée à l'équipe de conception de l'échantillon afin qu'elle puisse enregistrer cette information. Par ailleurs, le lien entre le questionnaire et le foyer nouvellement sélectionné peut avoir besoin d'être ajusté sur le fichier d'analyse finale dans les cas où le foyer initialement sélectionné n'a pas été de nouveau choisi.

3. Si une liste des foyers ne reflète pas un trajet continu de déplacement, et l'échantillon de foyers a été sélectionné en tant que grappe contiguë, l'enquêteur et le superviseur doivent identifier les foyers qui pourraient avoir été manqués dans les limites géographiques de la grappe contiguë choisie. Si un groupe contigu de un à trois foyers est manquant dans ces limites, l'enquêteur ou le superviseur doit identifier un sélectionné situé au plus près du groupe des un à trois foyers. Le foyer échantillonné le plus proche, ainsi que les foyers manqués, doivent être inscrits sur une feuille séparée et un foyer doit être choisi aléatoirement sur cette feuille. Ce foyer nouvellement sélectionné prendra la place du foyer échantillonné le plus proche. De nouveau, cette fiche doit être envoyée à l'équipe de conception de l'échantillon afin que les unités supplémentaires et le foyer remplacé puissent être documentés avec exactitude dans les dossiers d'analyse de

---

**Recommandation :**

**la technique des intervalles semi-ouverts doit être utilisée pour traiter les HH manqués quand la liste des foyers reflète une trajectoire continue de déplacement**

---

---

**Recommandation :**

**lorsque les enquêteurs ou les superviseurs trouvent un à trois HH omis, ils énuméreront les HH manqués sur une feuille et en sélectionneront un. Cette fiche doit être envoyée à l'équipe de conception du pays.**

---

---

**Recommandation :**

**l'enquêteur et/ou le superviseur doivent rechercher les foyers manqués dans les grappes contiguës sélectionnées dans le segment lorsque la liste des foyers est une carte.**

---

la GATS. Si un enquêteur ou un superviseur trouvent un groupe contigu de plus de trois foyers manqués alors ils doivent enregistrer tous les foyers manqués sur une feuille, et envoyer ces informations à l'équipe de conception de l'échantillon avant toute prise de contact avec un foyer échantillonné. Dans ce cas, l'équipe de conception de l'échantillon peut envisager de sous-sélectionner une zone dans le groupe contigu de foyers.

4. L'enquêteur et le superviseur doivent inspecter l'unité de logement choisie pour voir s'il semble y avoir plus d'un foyer vivant dans l'unité. Par exemple, une maison peut avoir été transformée en plusieurs appartements. Lorsque cela se produit, l'enquêteur ou le superviseur doit créer une liste de tous les foyers de l'unité de logement sélectionnée sur une feuille séparée et choisir au hasard l'un des foyers pour participer à l'enquête de la GATS. Cette étape de sous-échantillonnage ajoute une autre étape de sélection à la conception, et elle nécessite l'enregistrement du nombre de foyers à l'adresse de l'unité de logement, puisque la probabilité conditionnelle du sous-échantillonnage est égale à un divisé par le nombre de foyers à l'adresse sélectionnée. Encore une fois, si une seule unité de logement a plus de trois foyers manqués, alors l'information doit être envoyée à l'équipe de conception de l'échantillon du pays avant tout contact avec les foyers.

---

**EXIGENCE :**

**l'enquêteur et le superviseur doivent identifier une unité de logement sélectionnée avec plus d'un foyer et choisir l'un d'eux au hasard.**

---

Des recommandations plus détaillées sur la façon dont les enquêteurs et le superviseur doivent répondre à ces situations sont présentées dans le **Manuel de l'enquêteur de terrain de la GATS** et le **Manuel du superviseur de terrain de la GATS**.

Du point de vue de la conception de l'échantillon, ces types de problèmes sont fréquents dans les enquêtes de foyers et doivent être traités sur une base individuelle. Si un enquêteur ou un superviseur trouve un petit nombre d'unités de logement manquées ou trouve une unité de logement sélectionnée qui contient plusieurs foyers, ce sera plus efficace si l'enquêteur ou le superviseur prend un échantillon aléatoire des foyers manqués lors de la visite du segment. Si un grand nombre de foyers sont manqués, les enquêteurs et les superviseurs doivent transmettre cette information à l'équipe d'échantillonnage du pays et attendre des instructions. Dans ce dernier cas, on peut remédier à la situation de l'une des trois façons suivantes :

1. Les statisticiens peuvent simplement ajouter les unités de logement manquées à la liste des foyers et sélectionner à nouveau l'échantillon.
2. Les statisticiens peuvent simplement ajouter les unités de logement manquées à la fin de la liste des foyers et sélectionner un échantillon de foyers dans ceux qui ont été manqués, en utilisant le même taux d'échantillonnage que celui utilisé pour sélectionner l'échantillon initial des foyers dans le segment. Cette méthode est particulièrement avantageuse si un enquêteur a pris contact avec certains des ménages échantillonnés avant d'envoyer les informations à l'équipe d'échantillonnage du pays.
3. Si un nombre très élevé de foyers manqués est identifié, le statisticien peut diviser le segment en zones géographiques plus petites, choisir au hasard une zone, puis sélectionner un nouvel échantillon de foyers qui résident dans cette sous-zone. Ce processus est parfois appelé sous-segmentation.

Dans tous les cas, il est impératif que toute sous-sélection soit effectuée en utilisant un processus aléatoire, et que les probabilités de sélection au cours de ce processus de sous-sélection soient calculées et consignées dans le fichier d'analyse finale de l'enquête.

## 9. Sélectionner un résident admissible dans chaque foyer présélectionné

---

Une méthode de sélection aléatoire doit être utilisée pour sélectionner un individu admissible au hasard dans les foyers échantillonnés. (Voir les critères d'admissibilité des individus de la GATS dans la **section 3.1.**) En résumé, cette méthode de sélection procède comme suit :

1. Pour chaque foyer sélectionné dans l'échantillon de la GATS, un enquêteur frappera à la porte et essaiera d'identifier un individu du foyer qui est âgé de 18 ans ou plus et bien informé sur les résidents du foyer. Cette personne sera considérée comme le **répondant de la présélection**.
2. Un enquêteur posera plusieurs questions au répondant de la présélection sur les résidents du foyer. Entre autres choses, l'objectif est de créer une liste de tous les résidents du foyer n'habitant pas dans un établissement spécialisé, âgés de 15 ans ou plus, qui considèrent le foyer échantillonné comme leur résidence habituelle, et dont le sexe correspond à l'affectation au groupe masculin ou féminin du foyer échantillonné (si la randomisation des sexes a été appliquée à l'échantillon des foyers du pays). La liste finale des résidents admissibles du foyer de la liste doit être organisée du plus âgé au plus jeune.

3. Une fois que la liste de tous les résidents admissibles du foyer est entrée dans l'ordinateur de poche, celui-ci utilisera un processus aléatoire pour sélectionner un nom sur la liste et ce nom s'affichera sur l'écran. Plus précisément, l'appareil de poche générera un nombre aléatoire une fois que la liste des foyers est terminée. Supposons, par exemple, que le nombre aléatoire généré pour le foyer  $j$  est  $r_j$  et supposons que le nombre total de personnes inscrites

---

**Un résident admissible sera sélectionné au hasard à partir de la liste du foyer établie dans les foyers ayant un ou plusieurs résidents admissibles.**

---

sur la liste dans le foyer est  $\kappa_j$ . L'ordinateur de poche sera programmé pour sélectionner la personne correspondant à la partie entière de  $(r_j \cdot \kappa_j) + 1$ . Par exemple, si le nombre généré aléatoirement est 0,365789 et qu'il y a trois personnes admissibles sur la liste, le produit de 0,365789 par 3 est 1,097367 et en ajoutant un à ce résultat, on obtient 2,097367. Donc, l'individu en deuxième position sur la liste est sélectionné. Notez qu'en utilisant cette méthode de sélection, tous les répondants admissibles ont une chance égale de sélection et la probabilité de sélection pour les résidents choisis au hasard dans le foyer est l'inverse du nombre de résidents admissibles inscrits dans le foyer. Afin d'éviter un biais de sélection par ce processus, l'ordinateur de poche est programmé pour ne permettre la sélection du répondant qu'une fois par foyer (pour éviter le problème des enquêteurs appuyant sur la commande « select » [sélectionner] jusqu'à ce qu'un résident préféré soit choisi). Le résident sélectionné par l'ordinateur de poche et le résident répondant au *questionnaire individuel* doivent être la même personne.



## 10. Déterminer les tailles d'échantillon à chaque étape de la sélection et rapporter les taux de réponse définitifs

---

Dans ce chapitre, nous fournissons des recommandations sur la détermination de la taille de l'échantillon sélectionné nécessaire à chaque étape de la sélection de la GATS. Nous offrons également des directives sur le calcul des taux d'admissibilité et de réponse définitifs dans le contexte de la GATS. Les définitions du taux de réponse et d'admissibilité à employer sont particulièrement importantes, car l'utilisation de définitions uniformes de ces taux par tous les pays permettra à la GATS de comparer ces mesures essentielles de la qualité des données entre les différents pays.

### 10.1 Tailles de l'échantillon à chaque étape de la sélection

Rappelons que les PSU et/ou les segments doivent être stratifiés par degré d'urbanisation. Certains pays peuvent imposer une stratification supplémentaire explicite afin d'obtenir une taille d'échantillon spécifiée dans d'autres domaines d'intérêt, par exemple une stratification régionale. Dans cette section, nous avons fourni un exemple qui postule une taille de l'échantillon national de 8 000 répondants et où la randomisation des sexes est une caractéristique de la conception. Les pays avec d'autres tailles d'échantillon des répondants et caractéristiques doivent modifier leurs calculs en conséquence.

Dans chacune des strates géographiques, supposons que l'indice  $s$  se réfère à une strate géographique. Par ailleurs, supposons que :

$R_s^{Males}$  = Nombre total de répondants masculins désiré dans la strate. Si un pays ne forme des strates que par degré d'urbanisation, alors  $R_s^{Males} = 2,000$  pour  $s = \text{urbain}$  et  $s = \text{rural}$ . Il

$$\sum_s R_s^{Males} = 4,000$$

est nécessaire que

$R_s^{Females}$  = Nombre total de répondants féminins désiré dans la strate. Si un pays ne forme des strates que par degré d'urbanisation, alors  $R_s^{Females} = 2,000$ . Comme pour les hommes,

$$\sum_s R_s^{Females} = 4,000$$

il est nécessaire que

$\mathcal{E}_s^{HH}$  = Estimation du taux d'admissibilité des foyers Ce taux est défini comme le nombre total de foyers qui seront sélectionnés et déterminés admissibles à la GATS, divisé par le nombre total de foyers sélectionnés. Un foyer peut être déclaré non admissible s'il est vacant, n'a pas de résidents admissibles (des deux sexes, si la randomisation des sexes n'est pas utilisée, ou du sexe attribué si la randomisation des sexes est utilisée), ou n'existe pas au moment de la collecte de données. En général, nous nous attendons à ce que ce taux soit d'environ 90 %, bien que cela dépende du pays, de la probabilité des foyers où tous les résidents sont du même sexe (lorsque la randomisation des sexes est utilisée), et de l'ancienneté des données utilisées pour former la liste des foyers des zones.

$\tau_s$  = Estimation du taux de présélection des foyers. Il s'agit du pourcentage des foyers qui ont été présélectionnés et ont au moins une personne admissible à l'enquête. Ce taux prend en compte les foyers sans résident admissible âgé de 15 ans ou plus. Il tient également compte des foyers qui ont été affectés au groupe masculin, mais où aucun homme ne vit, lorsque la randomisation des sexes est utilisée. De même, il tient compte des foyers qui ont été affectés au groupe féminin, mais où aucune femme ne vit. En général, on s'attend à ce que la fraction des foyers qui ont au moins une personne admissible à la GATS soit très élevée. La plupart des pays doivent anticiper un taux de 95 % ou supérieur.

$\mathcal{E}_s^{Person}$  = Estimation du taux d'admissibilité au niveau individuel. Ce taux est défini comme le nombre total de personnes qui seront sélectionnées et déterminées admissibles à la GATS, divisé par le nombre total de personnes sélectionnées. Par exemple, une personne sera déclarée non admissible s'il s'avère qu'elle est âgée de 14 ans ou moins lors de l'entretien. En général, nous nous attendons à ce que ce taux soit très élevé, environ 98 % ou plus.

Remarque : Le  $\mathcal{E}_s^{Person}$  tient compte des personnes qui ont été sélectionnées et pour lesquelles il s'est avéré ultérieurement que celles-ci n'étaient pas admissibles à l'enquête. Le  $\tau_s$  tient compte des foyers, dont la liste n'a pas été achevée parce qu'aucune personne admissible à l'enquête n'a été identifiée par le résident du foyer répondant aux questions de présélection.

$\rho_s^{HH}$  = Estimation du taux de réponse des foyers. Il s'agit du nombre total de foyers avec une personne qui répondra aux questions de la liste, divisé par le nombre total de foyers *admissibles* sélectionnés. En général, nous nous attendons à ce que ce taux soit d'environ 98 % dans la plupart des pays.

$\rho_s^{Males}$  = Estimation du taux de réponse au niveau individuel pour les hommes. Ce taux est défini comme le nombre total d'hommes qui répondront à la GATS, divisé par le nombre total d'hommes *admissibles* sélectionnés. En général, nous nous attendons à ce que ce taux soit d'environ 80 % dans la plupart des pays.

$\rho_s^{Females}$  = Estimation du taux de réponse au niveau individuel pour les femmes. Ce taux est défini comme le nombre total de femmes qui répondront à la GATS, divisé par le nombre total de femmes *admissibles* sélectionnées. En général, nous nous attendons à ce que ce taux soit d'environ 90 % dans la plupart des pays.

Alors

$M_s^{Males}$  =  $\frac{R_s^{Males}}{\mathcal{E}_s^{Person} \cdot \rho_s^{Males}}$  est le nombre total d'hommes sélectionnés à partir des foyers présélectionnés, nécessaires dans la strate  $s$ .

$M_s^{Females}$  =  $\frac{R_s^{Females}}{\varepsilon_s^{Person} \cdot \rho_s^{Females}}$  est le nombre total de femmes sélectionnées à partir des foyers présélectionnés, nécessaires dans la strate  $s$ .

Et

$H_s$  =  $\frac{M_s^{Males} + M_s^{Females}}{\varepsilon_s^{HH} \cdot \tau_s \cdot \rho_s^{HH}}$  est le nombre total de foyers qui doivent être sélectionnés à partir de la strate  $s$ .

Nous recommandons en outre que la taille de l'échantillon au sein de la grappe se limite à environ 28 foyers sélectionnés par zone géographique finale (généralement segment) sélectionnée. En général, la taille optimale de la grappe de l'échantillon dépend de la mesure du résultat à l'étude et du degré d'homogénéité de la grappe.

---

**Recommandation :**  
sélectionnez 28 foyers par segment.

---

Par conséquent, si le nombre souhaité de foyers à sélectionner dans chaque zone géographique finale est de 28, alors le nombre total de zones à sélectionner à l'étape finale du processus de sélection

géographique est approximativement égale à  $\frac{H_s}{28}$ .

Notez que si un pays sélectionne des zones géographiques en une seule étape de sélection, le nombre total de PSU à sélectionner est  $\frac{H_s}{28}$ , qui devrait à peu près être égale à 400.

Comme indiqué précédemment, si un pays sélectionne des zones géographiques en plusieurs étapes de sélection géographique, le nombre total de PSU à sélectionner pourrait être d'environ 100 et le pays doit finalement choisir un échantillon d'environ quatre segments par PSU, soit environ 400 segments.

## 10.2 Exemple de calcul de la taille de l'échantillon

Pour illustrer les calculs de taille de l'échantillon mentionnés dans la section précédente, postulons que dans une strate un pays suppose ce qui suit :

Nombre total de répondants hommes dans la strate (Remarque : cette taille d'échantillon doit être égale à 4 000 ou plus sur l'ensemble des strates)	$R_s^{Male}$	2 000
Nombre total de répondants femmes dans la strate (Remarque : cette taille d'échantillon doit être égale à 4 000 ou plus sur l'ensemble des strates)	$R_s^{Female}$	2 000
Nombre total de répondants dans la strate (Remarque : cette taille d'échantillon doit être égale à 8 000 ou plus sur l'ensemble des strates)	$R_s^{Person}$	4 000
Taux d'admissibilité des foyers	$\varepsilon_s^{HH}$	90 %
Taux de réponse des foyers	$\rho_s^{HH}$	98 %
Pourcentage des foyers avec au moins un individu admissible à l'enquête	$\tau_s$	95 %
Taux d'admissibilité individuelle	$\varepsilon_s^{Person}$	98 %
Taux de réponse individuel	$\rho_s^{Person}$	85 %
Taux de réponse individuel des femmes	$\rho_s^{Female}$	90 %
Taux de réponse individuel des hommes	$\rho_s^{Male}$	80 %

Donc, le nombre total de personnes sélectionnées au sein des foyers présélectionnés avec succès est :

$$M_s^{Females} = \frac{R_s^{Females}}{\varepsilon_s^{Person} \cdot \rho_s^{Females}} = \frac{2,000}{.98 \cdot .90} = 2,268 \text{ femmes sélectionnées et}$$

$$M_s^{Males} = \frac{R_s^{Males}}{\varepsilon_s^{Person} \cdot \rho_s^{Males}} = \frac{2,000}{.98 \cdot .80} = 2,551 \text{ hommes sélectionnés.}$$

Et le nombre total de foyers à sélectionner à partir de cette strate est

$$H_s = \frac{M_s^{Males} + M_s^{Females}}{\varepsilon_s^{HH} \cdot \tau_s \cdot \rho_s^{HH}} = \frac{2,551 + 2,268}{.90 \cdot .95 \cdot .98} = 5,751 \text{ foyers sélectionnés.}$$

Notez que dans l'exemple ci-dessus la randomisation des sexes a été effectuée. Puisqu'il s'agit d'une enquête standard de la GATS où il n'y a pas de randomisation des sexes, le nombre total de personnes dans les foyers qui sont présélectionnées avec succès est :

$$M_s^{Person} = \frac{R_s^{Person}}{\varepsilon_s^{Person} \cdot \rho_s^{Person}} = \frac{4,000}{.98 \cdot .85} = 4,954 \text{ personnes sélectionnées.}$$

Et le nombre total de foyers à sélectionner à partir de cette strate est

$$H_s = \frac{M_s^{Person}}{\varepsilon_s^{HH} \cdot \tau_s \cdot \rho_s^{HH}} = \frac{4,954}{.90 \cdot .95 \cdot .98} = 5,912 \text{ foyers sélectionnés.}$$

### 10.3 Taux de réponses

Les taux de réponse et d'admissibilité doivent être calculés pour l'échantillon des foyers, l'échantillon individuel et l'échantillon combiné en utilisant les formules ci-dessous. Il est important d'utiliser ces formules pour que les taux de réponse et d'admissibilité puissent être comparés entre les pays. Ces taux doivent être calculés pour l'échantillon complet (à la fois du foyer et individuel), ainsi que par strates. Les strates peuvent être rurale/urbaine, masculine/féminine et toutes les autres strates qu'un pays utilise dans sa conception (par exemple, peut-être régionale).

---

#### EXIGENCE :

**les taux d'admissibilité et de réponse définitifs doivent être calculés en utilisant les formules standard de la GATS spécifiées dans cette section.**

---

Il faut noter que la définition du taux de réponse tel que défini ci-dessous est cohérente avec une définition standard des taux de réponse telle que suggérée par l'Association américaine de recherche sur l'opinion publique (AAPOR). Pour la GATS, nous utilisons la définition que l'AAPOR désigne sous les termes de taux de réponse RR1 dans la version 2009 du rapport AAPOR qui peut être consultée sur le site [http://www.aapor.org/Standard\\_Definitions/1818.htm](http://www.aapor.org/Standard_Definitions/1818.htm). Quand l'admissibilité à l'étude d'une partie non négligeable de cas non-répondants de l'échantillon est inconnue et le pourcentage de personnes admissibles parmi ceux dont l'admissibilité est connue est relativement faible, il est recommandé d'utiliser le taux de réponse RR3 au lieu de RR1. Le calcul de ces taux est plus amplement décrit dans le **Manuel de pondération de l'échantillon de la GATS**.

Deux séries de codes de décision opérationnelle joueront un rôle dans le calcul des taux de réponse au niveau des foyers (pour les efforts visant à trouver quelqu'un dans le foyer pour répondre au *questionnaire du foyer*), et au niveau individuel (pour les efforts pour obtenir un résident sélectionné au hasard au sein du foyer pour répondre au *questionnaire individuel*). Les enquêteurs pourront avoir à faire de multiples tentatives pour mener à bien ces deux niveaux de réponse. L'un des codes de résultats de la série de codes au niveau du foyer, et au niveau individuel, respectivement, seront entrés dans le programme après chacune de ces tentatives (consultez le **Manuel de l'enquêteur de terrain de la GATS** pour une liste complète de ces deux séries de codes de résultats). Une fois la collecte des données terminée, un code de décision finale au niveau des foyers et au niveau individuel sera déterminé pour chaque foyer échantillonné (consultez **Assurance qualité de la GATS : directives et**

**documentation** pour plus d'informations sur l'attribution des codes de décision). Dans la plupart des cas, à l'un ou l'autre des deux niveaux d'effort, le code de décision finale correspondra au code de résultat enregistré à la dernière tentative pour compléter cette partie du recrutement de l'enquête.

Les codes de décision finale au niveau des mesures concernant le questionnaire du foyer et individuel sont définis dans les deux documents suivants :

#### Document 10-1. Codes des décisions finales du questionnaire des foyers

Code	Description
1	Questionnaire du foyer rempli, un individu sélectionné -Le questionnaire du foyer est considéré comme complet si la liste est complète
2	Questionnaire du foyer rempli, aucun individu sélectionné -Personne n'est âgé de 15 ans ou plus -Personne ne considère le foyer comme étant leur lieu habituel de résidence -Le foyer a été attribué aux hommes et aucun individu de sexe masculin ne réside dans le foyer ou le foyer a été affecté aux femmes et aucun individu de sexe féminin ne réside dans le foyer (si la randomisation des sexes a été utilisée)
3	Questionnaire du foyer partiellement complété, liste interrompue (entretien incomplet)
4	Questionnaire du foyer incomplet, le répondant approprié n'a pas pu être identifié à la présélection -Aucune personne âgé de 18 ou plus n'est au domicile -Le membre du foyer âgé de 18 ou plus est incompetent -Ces foyers n'ont peut-être pas de résidents admissibles à l'enquête
5	Personne au domicile
6	Refus du foyer
7	Foyer vacant
8	Adresse sélectionnée n'est pas un foyer
9	Autre non-réponse du foyer

#### Document 10-2. Codes des décisions finales du questionnaire individuel

Code	Description
11	Questionnaire individuel terminé -Questionnaire complété au moins jusqu'à la question E1
12	Incomplet -Questionnaire commencé, mais interrompu avant la question E1
13	L'individu sélectionné s'est avéré non-admissible à l'enquête ultérieurement -Âge < 15 -L'individu ne considère pas le foyer comme étant son lieu habituel de résidence -Erreur de sexe (si la randomisation des sexes a été utilisée)
14	Le répondant sélectionné n'est pas chez lui
15	Refus du répondant sélectionné
16	Répondant choisi incompetent

Supposons que [\*] est le nombre de cas dans la GATS avec le code de décision finale \*. Notez que chaque questionnaire peut avoir jusqu'à deux codes de décision finale, un qui reflète le statut final au niveau du foyer (présélection) et un second code qui reflète le statut final au niveau individuel (entretien). Plus particulièrement,

- Si un questionnaire a un code de décision finale du foyer de 1, le questionnaire doit aussi avoir un second code de décision finale qui lui est associé et qui reflète le statut final des tentatives d'entretien avec le résident sélectionné dans le foyer en utilisant le *questionnaire individuel*.
- Si un questionnaire a un code de décision finale du foyer de 2, 3, 4, 5, 6, 7, 8 ou 9, alors il ne recevra pas de code de disposition finale reflétant le statut du résident sélectionné.

En utilisant la notation figurant dans la **section 10.1**,

$$\begin{aligned} \varepsilon_s^{HH} &= \text{Estimation du taux d'admissibilité des foyers.} \\ &= \frac{[1] + [3] + [4] + [5] + [6] + [9]}{[1] + [2] + [3] + [4] + [5] + [6] + [7] + [8] + [9]} \end{aligned}$$

$$\begin{aligned} \rho_s^{HH} &= \text{Estimation du taux de réponse des foyers.} \\ &= \frac{[1]}{[1] + [3] + [4] + [5] + [6] + [9]} \end{aligned}$$

Notez que [3] n'est pas inclus dans le numérateur du taux de réponse des foyers. Par conséquent, un questionnaire de présélection des foyers qui n'est que partiellement complet (c.-à-d., la liste n'a pas pu être terminée) sera considéré comme un non-répondant à la GATS. Notez également que [2] ne doit être inclus ni dans le numérateur ni dans le dénominateur du taux de réponse, car les foyers sans résident admissible sont considérés comme non admissibles à l'enquête. Ceux-ci seraient des foyers n'ayant pas de résidents admissibles, âgés de 15 ans ou plus qui au moment de la présélection du foyer considèrent le foyer comme leur résidence habituelle. Si la randomisation des sexes est utilisée, le code [2] inclurait également les foyers dans un groupe d'un sexe particulier, qui n'ont pas de résidents admissibles dans le groupe du sexe en question.

$$\tau_s = \text{Estimation du pourcentage des foyers avec au moins un individu admissible à l'enquête.}$$

$$= \frac{[1]}{[1] + [2]}$$

$\epsilon_s^{Person}$  = Estimation du taux d'admissibilité au niveau individuel pour les hommes et les femmes confondus (si la randomisation des sexes n'est pas utilisée) ; calculée séparément pour chaque groupe de sexe des foyers répondants si la randomisation des sexes est utilisée.

$$= \frac{[1\ 1] + [1\ 2] + [1\ 4] + [1\ 5] + [1\ 6] + [1\ 7]}{[1\ 1] + [1\ 2] + [1\ 3] + [1\ 4] + [1\ 5] + [1\ 6] + [1\ 7]}$$

$\rho_s^{Person}$  = Estimation du taux de réponse au niveau individuel pour les hommes et les femmes confondus.

$$= \frac{[1\ 1]}{[1\ 1] + [1\ 2] + [1\ 4] + [1\ 5] + [1\ 6] + [1\ 7]}$$

D'après la description ci-dessus et le taux de réponse **RR1** de l'AAPOR, les entretiens « incomplets » (code 12) sont considérés comme non-répondants admissibles à la GATS.

Et le taux de réponse combiné pour la GATS, pour les hommes et femmes confondus, est :

$$\rho_s^{Total} = \rho_s^{HH} \cdot \rho_s^{Person}$$

Il faut noter que les taux de réponse définitifs pour un pays de la GATS peuvent être légèrement différents de ceux rapportés par un pays lors du recueil des données, car les codes de décision finale seront attribués aux codes de résultat finaux qui ont été saisis lors de la collecte des données.

---

**Le taux de réponse combinée pour la GATS est défini comme le produit des taux de réponse au niveau des foyers et au niveau individuel.**

---



## 11. Calcul de la pondération de l'échantillon final pour analyse

---

La pondération de l'échantillon est une mesure numérique essentielle à la production et l'évaluation des estimations à partir des données de l'échantillon de l'enquête. Elle est destinée à rendre compte de la probabilité de chaque répondant à être inclus dans l'échantillon et des effets différentiels de non-réponse, des bases de sondage imparfaites, et d'autres éléments qui affectent la composition de l'échantillon. La pondération pour les enquêtes comme la GATS est généralement calculée en trois étapes de base décrites ci-dessous.

---

**La pondération de l'échantillon est une mesure statistique importante qui doit être calculée et utilisée pour l'analyse des données de la GATS.**

---

1. **Pondération de base** — Une pondération de base est calculée pour chaque répondant comme l'inverse de la probabilité (inconditionnelle) avec laquelle le répondant a été sélectionné aléatoirement dans l'échantillon. Cette probabilité de sélection des répondants est à son tour définie comme étant le produit des probabilités pour les étapes de l'échantillonnage qui ont conduit à la sélection du répondant (par exemple, pour un échantillon de foyers de quatre étapes, sans randomisation de sexes des foyers, la probabilité de sélectionner en premier la PSU où réside le répondant, multipliée par la probabilité de choisir la SSU du répondant étant donné que sa PSU a été choisie, multipliée par la probabilité de choisir le segment du répondant étant donné que sa SSU a été choisie, multipliée par la probabilité de choisir le foyer du répondant, étant donné que son segment a été choisi, et multipliée par la probabilité de sélection du répondant au sein du foyer du répondant).
2. **Correction pour la non réponse** — La pondération de base est multipliée par l'inverse des taux de réponse au niveau du foyer et individuel pour un sous-ensemble des membres admissibles sélectionnés de l'échantillon qui sont similaires au répondant pour des caractéristiques qui (espérons-le) sont en corrélation avec les principales mesures de résultat de l'étude et la propension à répondre à la GATS.
3. **Étalonnage** — La pondération de base corrigée est ensuite multipliée par un facteur qui étalonne l'échantillon en fonction de la répartition démographique définie par des caractéristiques qui sont susceptibles d'être en corrélation avec les critères clés de l'étude (par exemple, l'âge, le sexe et le niveau d'éducation). Cette pondération étalonnée devient la pondération finale corrigée de l'échantillon qui doit être utilisée pour toutes les analyses des données de l'enquête de la GATS dans chaque pays participant.

La distribution pondérée de la pondération finale corrigée pour les variables démographiques utilisées pour l'étalonnage correspondra au dénombrement de la population pour ce qui est de ces variables. Le **Manuel de pondération de l'échantillon de la GATS** présente les calculs de chaque étape de ce processus en détail.

---

**La pondération de l'échantillon est calculée suivant un processus en trois étapes.**

---

Comme indiqué dans le **chapitre 4**, la probabilité de sélection pour chaque étape de la conception de l'échantillon doit être consignée dans le dossier d'analyse finale pour chaque foyer et répondant individuel sélectionnés. Ces facteurs doivent être calculés, stockés, et la procédure suivie pour les

calculer soigneusement documentée au moment où l'échantillon est choisi, à chaque étape de la sélection.

## 12. Bibliographie

---

### 12.1 Échantillonnage : simple et bref

- Henry, Gary T., *Practical Sampling*, Sage Publications, Newbury Park, 1990.
- Kalton, G., *Introduction to Survey Sampling*, Sage Publications, 1983.
- Stuart, A., *Basic Ideas of Scientific Sampling*, 2nd Ed., Griffin's Statistical Monographs and Courses, No. 4, Hafner Press, New York, 1976.

### 12.2 Échantillonnage : préliminaire

- Barnett, Vic., *Sample Survey Principles and Methods*, Oxford University Press, 2002.
- Chaudhuri, Arijit, *Survey Sampling: Theory and Methods*, M. Dekker, New York, 2005.
- Cochran, W.G., *Sampling Techniques*, 3rd Ed., John Wiley and Sons, New York, 1977.
- Foreman, E. K., *Survey Sampling Principles*, M. Dekker, New York, 1991.
- Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., and Tourangeau, R., *Survey Methodology*, Wiley and Sons, New York, 2004.
- Jaeger, Richard M., *Sampling in Education and the Social Sciences*, Longman, New York, 1984.
- Jessen, Raymond J., *Statistical Survey Techniques*, Wiley, New York, 1978.
- Kish, L., *Survey Sampling*, Wiley, New York, 1965.
- Konijn, H. S., *Statistical Theory of Sample Survey Design and Analysis*, American Elsevier Pub. Co., New York, 1973.
- Levy, P. S. and Lemeshow, S., *Sampling of Populations – Methods and Applications*, Wiley, New York, 1999.
- Lohr, S., *Sampling: Design and Analysis*, Duxbury Press, Pacific Grove, 1999.
- Raj, D., *Sampling Theory*, McGraw-Hill, Inc., New York, 1968.
- Scheaffer, R.L., Mendenhall, W., and Ott, L. *Elementary Survey Sampling* (6th Edition), Duxbury Press, 2006.
- Thompson, Steven K., *Sampling*, Wiley, New York, 2002.
- Yates, Frank, *Sampling Methods for Censuses and Surveys*, Macmillan, New York, 1981.

### 12.3 Échantillonnage : traitement approfondi

- Cassel, Claes-Magnus, Sarndal, Carl-Erik, and Wretman Jan H'akan., *Foundations of Inference in Survey Sampling*, Wiley, New York, 1977.
- Hansen, M.H., Hurwitz, W.N., and Madow, W.G., *Sample Survey Methods and Theory*, Vol. I (Concepts and Discussion) and Vol. II, 1953.
- Krewski, D., *Current Topics in Survey Sampling*, Academic Press, New York, 1981.
- Murthy, M.N., *Sampling Theory and Methods*, Statistical Publishing Society, Calcutta, 1967.
- Namboodiri, N. K., *Survey Sampling and Measurement*, Academic Press, New York, 1978.
- Sarndal, Carl-Erik, Swensson, Bengt, and Wretman, Jan H'akan, *Model Assisted Survey Sampling*, Springer-Verlag, New York, 1992.
- Sukhatme, P. V. et al., *Sampling Theory of Surveys with Applications*, Iowa State University Press, 1984.
- Sukhatme, P.V. and Sukhatme, B.V., *Sampling Theory of Surveys with Applications*. Revised 2nd Ed., Iowa State University Press, Ames, 1970.
- Wolter, K.M., *Introduction to Variance Estimation*, Springer-Verlag, New York, 1985.

### 12.4 Manuels pratiques

- Grosh, M. and Munoz, J. A Manual for Planning and Implementing the Living Standard Measurement Survey Study. LSMS Working Paper No. 126. World Bank, Washington, DC. 1996.
- Macro International. Sampling Manual. DHS Basic Documentation No. 6. Calverton, MD., 1996.

SAS Institute. SAS/STAT User's Manual, Version 8. Cary, NC, 1999.

United Nations. Household Sample Surveys in Developing and Transition Countries. New York, NY, 2005.

United Nations. Principles and Recommendations for Population and Housing Censuses, Revision 1. Series M, No. 67, Rev. 1 (United Nations publication, Sales No. E.98.XVII.1). (para. 2.61).

## 12.5 Autres publications

Chromy, J.R. (1979). "Sequential Sample Selection Methods." Proceedings of the American Statistical Association, Section on Survey Research Methods, pp. 401-406.

Folsom, R.E., F.J. Potter, and S.R. Williams (1987). "Notes on a Composite Size Measure for Self-Weighting Samples in Multiple Domains." Proceedings of the American Statistical Association, Section of Survey Research Methods, pp. 792-796.

Kavee, J., Liu, J., Iannacchione, V.G. (2001). "An Optimal Allocation Method for Two-Stage Sampling Designs with Stratification at the Second Stage". Proceedings of the American Statistical Association 2001, Section on Survey Research Methods.

Liu, J., Iannacchione, V.G., and Kavee, J.D. (1998). "Estimating Variance Components for a Two-Stage Design with Second-Stage Strata Nested Within PSUs". Proceedings of the American Statistical Association 1998, Section on Survey Research Methods, pp.657-661.

Sampford, M.R. (1967) On Sampling Without Replacement With Unequal Probabilities of Selection. Biometrika, 54, pp 499-513.

Vijayan, K. (1968), "An Exact  $\pi_{ps}$  Sampling Scheme: Generalization of a Method of Hanurav," Journal of the Royal Statistical Society, Series B, 30, 556 - 566.

## 12.6 Manuels de la GATS

Global Adult Tobacco Survey Collaborative Group. *Global Adult Tobacco Survey (GATS): Core Questionnaire with Optional Questions, Version 2.0*. Atlanta, GA : Centers for Disease Control and Prevention, 2010.

Global Adult Tobacco Survey Collaborative Group. *Global Adult Tobacco Survey (GATS): Field Interviewer Manual, Version 2.0*. Atlanta, GA : Centers for Disease Control and Prevention, 2010.

Global Adult Tobacco Survey Collaborative Group. *Global Adult Tobacco Survey (GATS): Field Supervisor Manual, Version 2.0*. Atlanta, GA : Centers for Disease Control and Prevention, 2010.

Global Adult Tobacco Survey Collaborative Group. *Global Adult Tobacco Survey (GATS): Mapping and Listing Manual, Version 2.0*. Atlanta, GA : Centers for Disease Control and Prevention, 2010.

Global Adult Tobacco Survey Collaborative Group. *Global Adult Tobacco Survey (GATS): Quality Assurance: Guidelines and Documentation, Version 2.0*. Atlanta, GA : Centers for Disease Control and Prevention, 2010.

Global Adult Tobacco Survey Collaborative Group. *Global Adult Tobacco Survey (GATS): Sample Weights Manual, Version 2.0*. Atlanta, GA : Centers for Disease Control and Prevention, 2010.

## Annexe A. Justification de la taille recommandée de l'échantillon de la GATS

---

Le **chapitre 4** de ce manuel indique que la taille minimale acceptable de l'échantillon des foyers dans chaque pays, pour chaque cycle de la GATS, est de 8 000.<sup>1</sup> Ce chiffre a été déterminé par les membres d'un comité de planification de la GATS durant la phase précoce de l'élaboration du protocole de recherche de la GATS. Il est basé uniquement sur les exigences de puissance et de précision pour les estimations clés de la GATS en fonction du niveau d'urbanisation (c'est-à-dire, zone urbaine ou rurale) et du sexe. L'annexe suivante présente une justification de la taille recommandée de l'échantillon.

### Estimations importantes de la GATS

Les mesures statistiques de la puissance et de la précision ont été examinées, puis développées dans le contexte des types d'estimations requis par la GATS. Deux types d'estimations ont été jugés importants dans l'analyse des données de la GATS. Un type indique les variations de diverses mesures du tabagisme au cours du temps (c'est-à-dire entre différentes paires de phases de l'enquête). Ces mesures de changement seront importantes puisque des cycles successifs de la GATS seront menés tandis que des programmes de lutte contre le tabagisme seront mis en place, et que leur efficacité devra être évaluée. Un deuxième type d'estimation sera utilisé pour établir un profil de tabagisme dans la population au moment où les phases individuelles de la GATS sont menées.

### Puissance et taille de l'échantillon pour détecter les changements au cours du temps

Les estimations de changement d'une phase à une autre ainsi que les estimations spécifiques des phases seront calculées pour divers taux de comportement/d'usage du tabac. Les trois taux suivants ont été considérés comme les plus importants pour la planification :

- **Taux de prévalence du tabagisme** — Le pourcentage de personnes qui utilisent actuellement des produits tabagiques ;
- **Taux de tabagisme passif** — Le pourcentage de personnes qui sont exposées à la fumée qui est produite par d'autres fumeurs ; et
- **Taux de sevrage** — Le pourcentage de fumeurs actuels qui ont essayé d'arrêter de fumer.

Le calcul des estimations du changement ou spécifiques de la phase de la consommation de tabac par sexe pour des segments urbains et ruraux de la population cible a été considéré suffisamment important pour que les tailles de l'échantillon soient sensiblement les mêmes pour chacune de ces quatre estimations de sous-groupes à chaque phase de la GATS. Alors que l'on s'attendait à ce que les tailles de l'échantillon des répondants masculins et féminins soient similaires, il a été décidé que la composition des échantillons de la GATS serait contrôlée par la stratification de sorte qu'environ la moitié des répondants masculins et féminins viennent des zones rurales et l'autre moitié des zones urbaines. Les méthodes pour atteindre l'égalité des tailles des échantillons de ces quatre importants sous-groupes sont présentées dans le manuel.

---

<sup>1</sup> Notez qu'un échantillon de 8 000 foyers répondants correspond à un échantillon de 8 000 individus répondants puisqu'un résident admissible sera sélectionné par foyer.

## Spécifications, hypothèses, et approche

La puissance statistique pour détecter les changements entre deux cycles de la GATS est déterminée par deux principales caractéristiques de la conception de l'échantillon, à savoir (i) l'effet multiplicateur combiné sur la variance des estimations en raison de l'échantillonnage en grappes, la stratification et la répartition disproportionnée entre les strates, tel que mesuré par l'effet de conception (ou *Deff*), et (ii) la sélection indépendante ou non des échantillons des deux cycles. Comme indiqué dans le **chapitre 4**, *Deff*, l'augmentation multiplicative de la variance des estimations de l'échantillon, est presque toujours supérieure à un (Kish, 1965). *Il a été admis que Deff* est égale à 2,00 pour les calculs de la puissance et de la précision ci-dessous. Par ailleurs, les calculs de précision et de puissance utilisent des formules standard pour la variance de la différence des estimations à partir d'échantillons indépendants. Une approximation du cumul de la distribution normale standard, proposée par Polya (1945) et décrite dans la section 13.5 de Johnson et Kotz (1970), a été utilisée pour calculer l'erreur de type II et la puissance.

La relation entre les choix des échantillons dans les différents cycles affecte également la puissance et la précision. Des échantillons conçus pour se chevaucher auront une plus grande puissance pour détecter les changements que celle des échantillons de la même taille, spécifiques du cycle, choisis de manière totalement indépendante. Le niveau d'augmentation de la puissance dépend de la manière dont les échantillons se chevauchent. Les échantillons forcés à se chevaucher dans les étapes tardives de la sélection et au-delà permettront généralement une augmentation supérieure de la puissance par rapport à ceux qui se chevauchent seulement dans les étapes précoces. Par exemple, si les échantillons de deux cycles de comparaison sont choisis dans les mêmes PSU mais les échantillons des foyers des PSU communes sont choisis indépendamment, le niveau d'augmentation de la puissance sera inférieur à celui des échantillons comprenant les mêmes foyers dans les deux cycles.<sup>2</sup> Comme, dans les premiers stades de la planification de la GATS, on ne savait pas s'il serait pratique de concevoir des échantillons qui se chevauchent, les calculs de puissance supposaient (de façon conservatrice) que l'échantillon à chaque cycle de la GATS serait choisi indépendamment de l'échantillon sélectionné dans les autres cycles. Donc, les figures de puissance présentées ci-dessous dans cette annexe seront un peu conservatrices, si la GATS décide finalement de chevaucher les échantillons au cours des cycles de quelque manière.

Pour les pays dont le rapport des estimations régionales n'est pas une priorité, les spécifications pour tous les cycles de la GATS exigent que les échantillons utilisés pour produire des estimations pour chacun des sous-groupes de population par degré d'urbanisation en fonction du sexe soient de taille suffisante pour détecter :

- **Les taux de prévalence du tabagisme** — Une baisse des taux de prévalence du tabagisme de 40 % à 34 % avec une puissance > 80 %, et à 33 % avec une puissance > 90 %, étant donné une erreur de type I de 0,05 et une hypothèse alternative bilatérale ;
- **Les taux de fumée passive** — Une baisse des taux de fumée passive (parmi les 60 % présumés de répondants non-fumeurs de l'échantillon) de 80 % à 70 % avec une puissance > 90 %, étant donné une erreur de type I de 0,05 et une hypothèse alternative bilatérale ;

---

<sup>2</sup> Notez que dans une population stable, deux cycles avec les mêmes foyers échantillonnés doivent avoir été choisis à partir des mêmes PSU, SSU, etc., pour qu'ils se chevauchent complètement dans toutes les étapes d'échantillonnage.

- **Taux d'abandon** — Une augmentation du taux d'abandon (parmi les 40 % présumés des répondants fumeurs de l'échantillon) de 10 % à 20 % avec une puissance > 90 %, étant donné une erreur de type I de 0,05 et une hypothèse alternative bilatérale.

Les exigences de précision de la GATS correspondantes pour les mêmes paramètres du pays étaient que les estimations spécifiques du cycle des taux de prévalence du tabagisme de 40 % doivent avoir une marge d'erreur de 3 points de pourcentage ou moins.

## Résultats

Les sept figures présentées ci-dessous contiennent les valeurs de puissance et de précision qui ont été utilisées pour déterminer la taille recommandée de l'échantillon des foyers de 8 000. Plus précisément, les résultats tirés de ces figures ont permis de déterminer la taille de l'échantillon nécessaire dans un pays de la GATS dont les principales priorités d'analyse étaient de produire des estimations conformes aux spécifications des sous-groupes suivant le degré d'urbanisation en fonction du sexe. Puisque des tailles d'échantillon égales pour ces quatre sous-groupes permettent de respecter cette exigence avec la taille de l'échantillon total la plus petite, quatre fois la taille de l'échantillon remplissant les spécifications de précision et de puissance de la GATS pour les estimations des quatre sous-groupes correspondrait à la taille recommandée de l'échantillon total. Parce qu'il a été montré (comme on le verra ci-dessous) qu'une taille d'échantillon de répondants d'environ 2 000 satisfait les trois exigences de puissance et l'exigence unique de précision, 8 000 = 4 x 2 000 a été adopté comme l'exigence de taille minimale de l'échantillon recommandée pour la GATS lorsque des estimations régionales ne sont pas nécessaires. Dans les pays où les estimations de qualité de la GATS sont requises pour les quatre sous-groupes au sein des régions, le chiffre de 8 000 a été recommandé comme taille minimale acceptable de l'échantillon *pour chaque région*.

Les figures A.1 à A.3 montrent les tailles des échantillons de répondants nécessaires pour détecter des changements à l'aide d'un test de différence avec une erreur de type I de 0,05 et une alternative bilatérale entre les deux cycles de la GATS pour les trois principaux taux d'utilisation du tabac définis ci-dessus, où le « point A » se référant au premier cycle dans l'ordre et le « point B » se référant à un cycle ultérieur. Le niveau de puissance requis a été fixé à 80 % pour un taux de prévalence de l'usage du tabac, et à 90 % pour les taux de fumée passive et d'abandon. Dans chaque figure une série de taux possibles au point B (y compris celui fixé par les spécifications de puissance de la GATS) est tracée en comparaison avec les spécifications des taux correspondants au point A, faisant ainsi varier la taille de la différence des taux détectables.

Trois principales conclusions de ces trois premières figures ont conduit à la taille recommandée de l'échantillon des répondants de 2 000 pour chacun des quatre sous-groupes suivant le degré d'urbanisation en fonction du sexe. En premier lieu, remarquez dans la figure A.1 que la taille de l'échantillon des répondants nécessaire pour une baisse à un taux de prévalence du tabagisme au point B de 34 % avec une puissance de 80 % est de 2 024. Par ailleurs, la taille de l'échantillon des répondants requise dans la figure A.2 pour une baisse à un taux de fumée passive au point B de 70 % avec une puissance de 90 % est de 1 295, alors que dans la figure A.3 la taille de l'échantillon des répondants nécessaires pour une augmentation du taux d'abandon au point B de 20 % avec une puissance de 90 % est de 1 317. Ainsi, une taille d'échantillon de 2 000 fournira la puissance quasi adéquate pour les

comparaisons de la prévalence du tabagisme, surtout si  $Deff$  s'avère être inférieur à 2,00 et étant donné l'hypothèse plus conservatrice de l'indépendance de l'échantillon et une hypothèse de différence alternative bilatérale. Avec une taille d'échantillon de 2 000 pour chacun des quatre sous-groupes principaux, une puissance clairement suffisante est indiquée pour détecter les différences entre les cycles pour les taux de fumée passive et d'abandon. Pour ces raisons, une taille d'échantillon de répondants de 2 000 a été jugée suffisante pour satisfaire les exigences de puissance pour chacun des sous-groupes de degré d'urbanisation en fonction du sexe.

Les figures A.4 à A.6 contiennent des tracés du niveau de puissance statistique correspondant à la taille de l'échantillon des répondants de 2 000 pour chacun des sous-groupes de degré d'urbanisation en fonction du sexe, dans son ensemble. Par exemple, notez dans la figure A.4 qu'une taille d'échantillon de répondants de 2 000 permettrait de détecter une diminution de la prévalence du tabagisme de 40 % au point A à 34 % au point B avec une puissance de 79,52 %, et un changement de 40 % à 33 % avec une puissance de 90,50 %, ce qui suggère que les exigences de puissance de la GATS pour ces comparaisons seront respectées. En outre, la figure A.5 montre qu'avec un échantillon de 1 200 non-fumeurs sur une taille d'échantillon totale des sous-groupes de 2 000, la puissance pour détecter une diminution du taux de fumée passive de 80 % au point A à 70 % au point B est 98,26 %, ce qui est bien au-dessus des 90 % requis. De plus, la figure A.6 illustre qu'avec un échantillon de 800 fumeurs sur une taille d'échantillon totale des sous-groupes de 2 000, la puissance pour détecter une augmentation du taux d'abandon de 10 % au point A à 20 % au point B (98,11 %) est bien au-dessus des 90 % requis. Ces résultats confirment ainsi davantage la capacité d'une taille d'échantillon d'un sous-groupe de 2 000 à répondre aux exigences de puissance de la GATS.

### **Précision et taille de l'échantillon pour estimer les taux de prévalence à chaque cycle**

Enfin, la figure A.7 indique les exigences de taille de l'échantillon de répondants pour estimer les taux de tabagisme entre 10 % et 90 % avec une marge d'erreur (MOE) basée sur un niveau de confiance de 95 %. Des tracés séparés sont présentés pour les MOE variant de deux à six points de pourcentage. Selon les exigences de précision de la GATS, les MOE des estimations spécifiques du cycle de la prévalence du tabagisme de 40 % pour les sous-groupes de degré d'urbanisation en fonction des sexes ne doivent pas être supérieures à 3 %. Un échantillon de 2 049 répond précisément à cette exigence, avec un échantillon de 2 000 impliquant une marge d'erreur de 3,04 %.

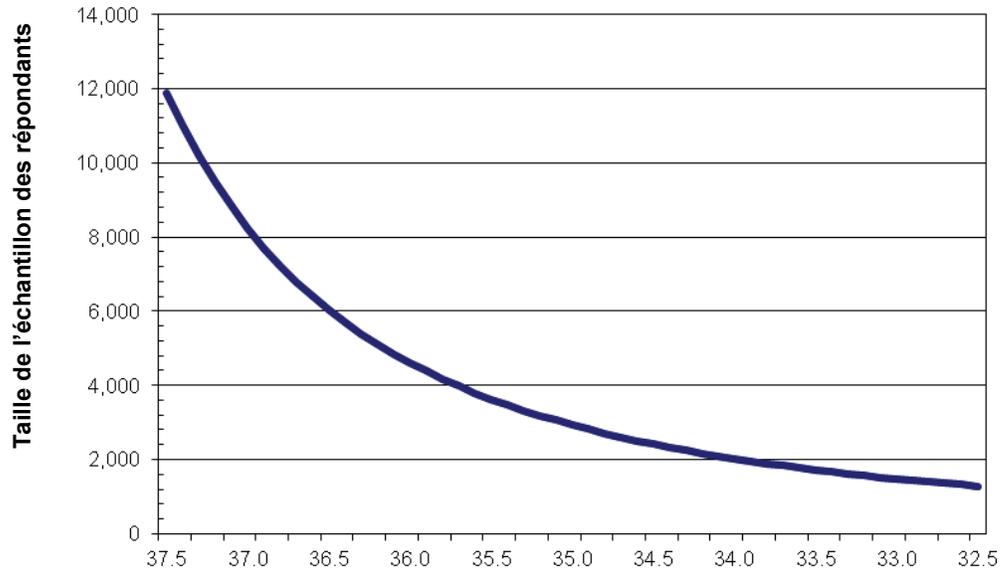
### **Références**

1. Johnson, N.L. and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions – 1*, Houghton Mifflin Company, Boston, MA.
2. Kish, Leslie (1965) *Survey Sampling*, John Wiley and Sons, New York.
3. Polya, G. (1945). "Remarks on computing the probability integral in one and two dimensions," *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability*, pp.63-78.

**FIGURE A.1**

**Taille de l'échantillon global à chaque point temporel pour détecter une diminution de la prévalence du tabagisme à partir de 40 % au point A avec une puissance de 80 %**

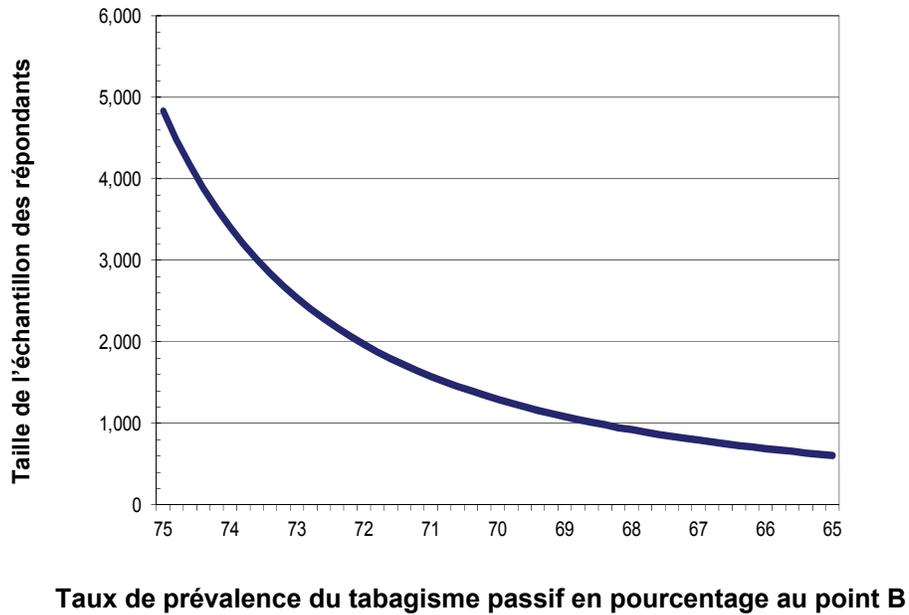
Deff = 2,00 ; alternative bilatérale



**Taux de prévalence du tabagisme actuel en pourcentage au point B**

**FIGURE A.2**

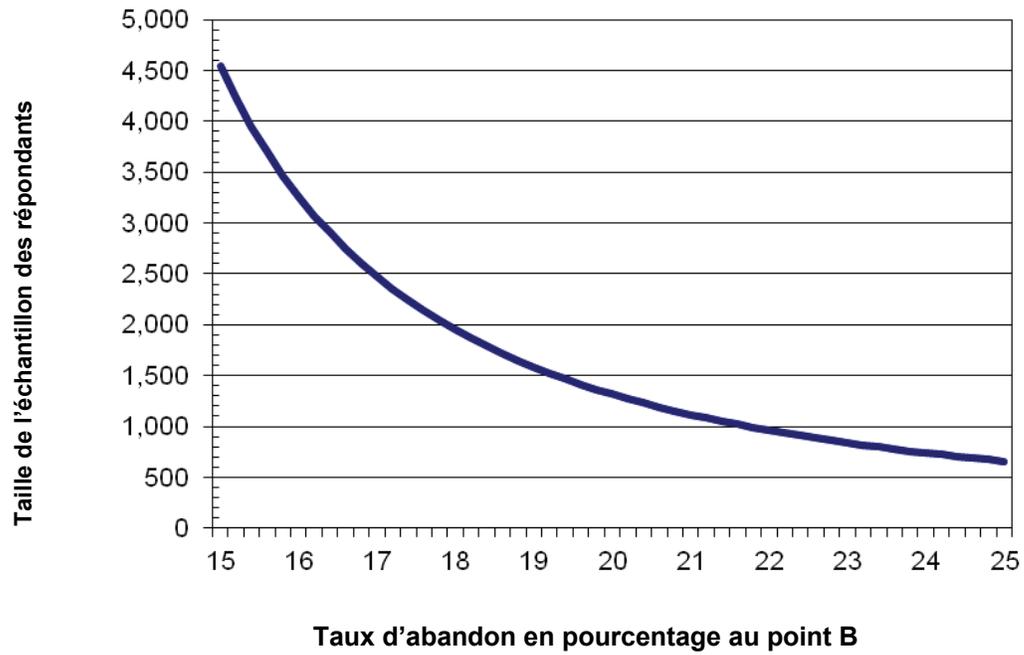
**Taille de l'échantillon global à chaque point temporel pour détecter une diminution du taux de tabagisme passif à partir de 80 % au point A avec une puissance de 90 %**  
Deff = 2,00 ; alternative bilatérale



**FIGURE A.3**

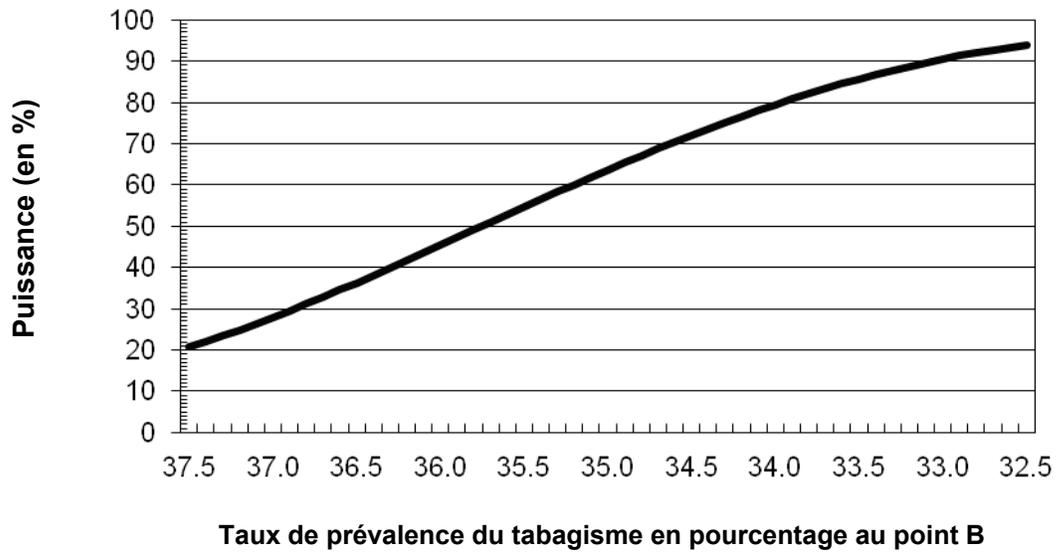
**Taille de l'échantillon global à chaque point temporel pour détecter une augmentation du taux d'abandon à partir de 10 % au point A avec une puissance de 90 %**

Deff = 2,00 ; alternative bilatérale



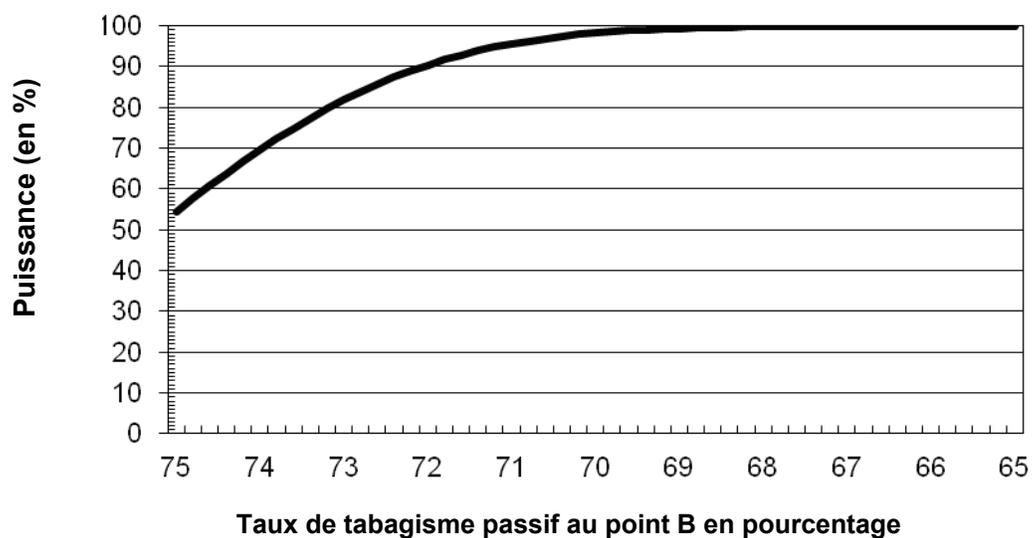
**FIGURE A.4**  
**Puissance (en %) pour détecter une diminution à partir du point temporel A du taux de prévalence du tabagisme de 40 %**

Tailles de l'échantillon point A et point B = 2,00 ; alternative bilatérale



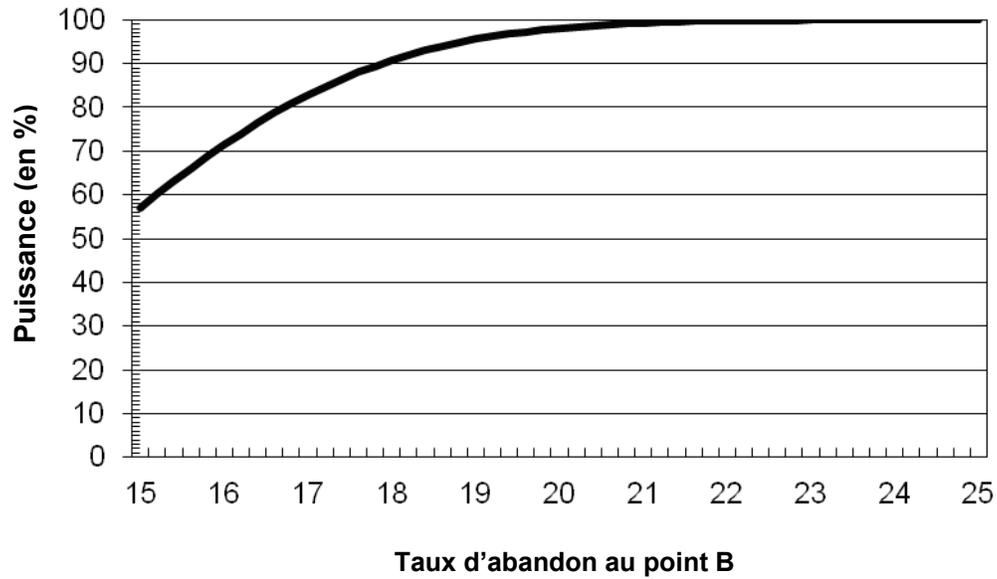
**FIGURE A.5**  
**Puissance (en %) pour détecter une diminution à partir du point A du taux de tabagisme passif de 80 %**

Tailles de l'échantillon non-fumeur point A et point B = 1 200 ;  $D_{eff} = 2,00$  ;  
alternative bilatérale

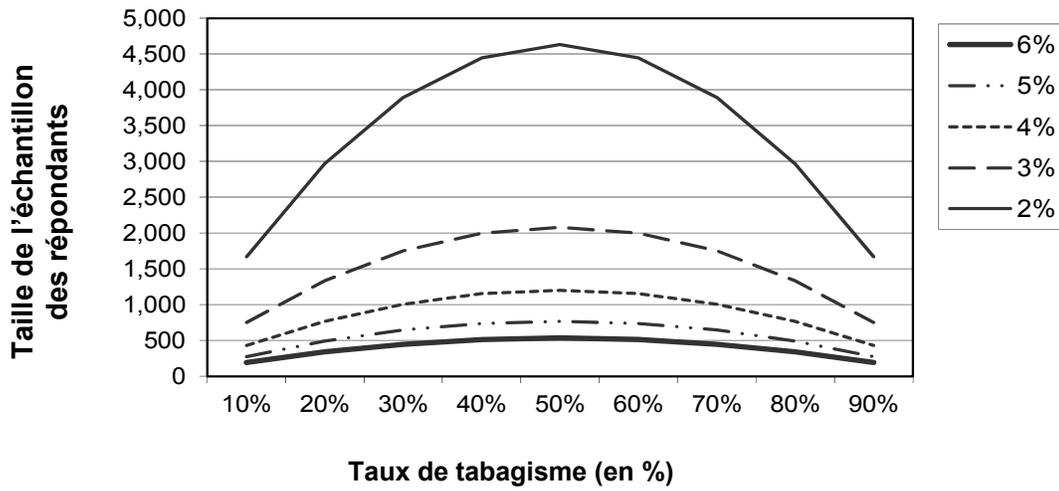


**FIGURE A.6**  
**Puissance (en %) pour détecter une augmentation du taux d'abandon à partir du point A de 10 %**

Tailles de l'échantillon fumeur point A et point B = 800 ;  $D_{eff} = 2,00$  ; alternative bilatérale



**FIGURE A.7**  
**Taille de l'échantillon nécessaire pour obtenir une**  
**marge d'erreur spécifique de 95 % (en %)**  
 Deff = 2,00 ; valeurs tracées de MOE de 2 % à 6 %





## Annexe B. Options de conception de l'échantillon pour les nouveaux pays de la GATS

---

Cette annexe décrit trois options de conception (OPTIONS N1 à N3) qui pourraient être envisagées en conjonction avec le développement de méthodes d'échantillonnage pour les pays mettant en œuvre la GATS pour la première fois ou comme base de référence par rapport à aux pays qui ont déjà réalisé au moins une fois la GATS. Une courte description de chaque option, avec une liste de certains des avantages et des inconvénients, est présentée ci-dessous. Le personnel de la GATS travaillera avec le personnel technique du pays pour choisir l'une de ces options, en cherchant à déterminer ce qui facilitera plus efficacement le développement de plans d'échantillonnage reconnus.

Les considérations générales suivantes portent sur le développement de l'ensemble des options de conception pour des échantillons de nouveaux pays de la GATS :

- Toutes les conceptions doivent respecter les spécifications mentionnées dans les sections précédentes de ce manuel.
- Plus précisément, toutes les conceptions de l'échantillon doivent utiliser une sélection aléatoire de sorte que la probabilité de sélection pour tous les membres de l'échantillon résultant à chaque étape de sélection puisse être déterminée. Ce faisant, la conception produira un **échantillon de probabilité** des répondants de la GATS.
- Les PSU doivent être des unités territoriales géopolitiques reconnues, au nombre d'au moins 1 000 dans le pays, et au moins 100 PSU doivent être échantillonnées dans la première étape de sélection globalement, ou pour chaque région du pays pour lesquelles des estimations de la qualité de la GATS sont désirées.
- Pour les conceptions permettant la production des estimations de la qualité de la GATS à la fois par sexe et par degré d'urbanisation, au moins 8 000 répondants sont recommandés.
- L'échantillon des foyers doit être choisi en un minimum de deux étapes d'échantillonnage.
- Intégrer la conception de l'échantillon de la GATS dans un autre échantillon d'enquête nationale reconnue est acceptable et encouragé, surtout si cela permettra d'économiser sur les coûts de collecte des données.

### OPTION N1 : Conception autonome

Une conception autonome de la GATS (c'est-à-dire, un échantillon est conçu et sélectionné pour la GATS uniquement), avec la conception de l'échantillon suivant le protocole d'échantillonnage standard de la GATS pour la structure et la taille. Une taille globale de l'échantillon des répondants de  $n = 8\,000$  est recommandée, puisque  $n = 2\,000$  sont nécessaires pour répondre aux normes de qualité statistique de la GATS (afin d'estimer les taux de tabagisme spécifiques aux cycles et de détecter les modifications de ces taux d'un cycle à l'autre) pour chacun des quatre sous-groupes de population classés par sexe en fonction du degré d'urbanisation.

- Les taux de prévalence masculin et féminin de tabagisme sont présumés être relativement similaires.

- n = 8 000 répondants par région, si des estimations de qualité régionales de la GATS sont nécessaires.

**Avantages :**

- La méthode la plus commune, puisque c'est la conception décrite dans le **Manuel de conception de l'échantillon de la GATS** et fréquemment utilisée dans les pays en phase I.
- Elle constitue la meilleure approche pour assurer la comparabilité générale de conception des résultats *obtenus dans les autres pays de la GATS*.

**Inconvénients :**

- La taille de l'échantillon peut rendre la méthode inabordable, si les fonds disponibles pour la collecte des données sont limités.
- Nécessite une ou des organisations partenaires dans le pays ayant une solide expérience de l'échantillonnage et des enquêtes. Davantage d'assistance technique peut être nécessaire si l'expérience préalable de la conduite des enquêtes de foyers, en personne, à l'échelle nationale est limitée.

**OPTION N2 : Conception autonome à échelle réduite**

GATS autonome suivant le protocole d'échantillonnage standard de la GATS pour la structure, mais ayant une taille d'échantillon plus petite suffisante pour produire des estimations de qualité acceptable pour la GATS pour les deux groupes de sexe et les deux sous-groupes de degré d'urbanisation séparément. Une taille globale d'échantillon de répondants de n = 4 000 est recommandée car les tailles des échantillons acceptables pour la GATS de n = 2 000 peuvent être obtenues marginalement pour chacun des sous-groupes clés des deux sexes ou des degrés d'urbanisation.

- Il n'est pas nécessaire de produire des estimations d'une précision suffisante pour les quatre sous-groupes de population suivant le sexe en fonction du degré d'urbanisation.
- Les taux de prévalence masculin et féminin de tabagisme sont présumés être relativement similaires.
- L'échantillon peut avoir besoin d'être sélectionné de manière disproportionnée par degré d'urbanisation si un partage 50/50 est nécessaire pour la taille de l'échantillon urbain/rural.
- Consultez la **section 7** de ce manuel pour préciser les exigences et les recommandations sur la taille de l'échantillon pour chaque étape de l'échantillonnage, notant en particulier le nombre minimal de PSU de l'échantillon.
- n = 4 000 est recommandé par région, si des estimations régionales de qualité de la GATS sont nécessaires.

**Avantages :**

- Phase de recueil de données plus courte et moins coûteuse en raison de la taille plus petite de l'échantillon.

- Une approche adéquate pour assurer la comparabilité avec les résultats provenant des autres pays de la GATS.

### ***Inconvénients :***

- Qualité plus limitée de l'ensemble des résultats, urbains-ruraux, et des autres plus petits sous-groupes comparé à l'OPTION N1 si l'échantillonnage n'est pas pratiqué de manière disproportionnée par degré d'urbanisation.
- Nécessite une ou des organisations partenaires dans le pays ayant une solide expérience de l'échantillonnage et des enquêtes. Davantage d'assistance technique peut être nécessaire si l'expérience préalable de la conduite des enquêtes de foyers, en personne, à l'échelle nationale est limitée.
- Il est moins probable (par rapport à l'OPTION N1) que les estimations régionales approchent une qualité acceptable si la taille de l'échantillon global est limitée à  $n = 4\ 000$  répondants.

### **OPTION N3 : Conception intégrée**

Intégrer l'échantillon de la GATS avec un échantillon de foyers national existant et hautement apprécié. L'échantillon existant peut être un échantillon maître d'un pays ou une autre enquête (par exemple, DHS, une importante enquête sociale en cours menée par le principal organisme statistique national, une enquête de santé nationale hautement considérée, etc.). La taille des échantillons peut varier selon les besoins en termes d'analyse (voir OPTIONS N1 et N2).

- Définition : l'intégration de la conception de l'échantillon se produit lorsque deux ou plusieurs échantillons partagent au moins une partie de leurs unités d'échantillonnage sélectionnées d'une ou plusieurs de leurs étapes de sélection.
- La GATS aura généralement la conception « bénéficiaire » de l'échantillon provenant de la conception de l'échantillon existant, ce qui signifie que tout l'échantillon ou un sous-échantillon aléatoire des foyers de l'échantillon existant est choisi pour la GATS.
- L'intégration des conceptions d'échantillons peut impliquer que la collecte de données de deux enquêtes devra être effectuée par un organisme d'enquête plus ou moins simultanément.
- Comme une conception autonome, cette option de conception peut être d'une taille plus ou moins grande selon les sous-groupes de population (définis par sexe et/ou de degré d'urbanisation) pour lesquels des estimations suffisamment précises sont nécessaires. La taille de l'échantillon de la GATS proposée sous cette option dépendra des ressources disponibles et de la taille de l'échantillon des foyers avec lequel l'échantillon de la GATS est intégré.

### ***Avantages :***

- De plus en plus de pays ont des échantillons de foyers maîtres et les utilisent pour faire diverses enquêtes nationales pour répondre à leurs besoins croissants d'informations produites par des enquêtes. Par ailleurs, plusieurs de ces échantillons maîtres sont conçus suivant des principes établis d'échantillonnage probabiliste de surface, et sont élaborés en consultation avec des personnes expérimentées dans l'application de ces principes.

- Le coût moyen par répondant pour intégrer l'échantillonnage de la GATS et le recueil de données avec un échantillon de foyers maître/existant et son activité associée de collecte de données peut être significativement moindre en raison du rendement du processus (à savoir, une infrastructure de collecte de données d'enquête pour rassembler de multiples ensembles de données, peut-être simultanément).
- Le pays peut être plus à même de soutenir le système d'enquête de la GATS puisqu'en choisissant cette option, la GATS travaillera avec des personnes capables de concevoir et mettre en œuvre un système de données gouvernemental établi et vraisemblablement de haute qualité.
- Nécessite la pleine coopération de ceux qui ont produit l'échantillon existant (pour élaborer un plan réalisable d'intégration, calculer la pondération de l'échantillon final de la GATS, documenter de manière efficace l'échantillon de la GATS, rendre les deux plans d'échantillonnage compatibles pour l'analyse, etc.).
- La qualité de l'échantillon final de la GATS dépend de la qualité de la conception et de la sélection de l'échantillon existant.

### **Inconvénients :**

- Si un sous-échantillonnage est impliqué, il peut être difficile de concevoir l'approche de sous-sélection afin qu'elle génère un sous-échantillon qui respecte les normes de la GATS (par exemple, la difficulté rencontrée dans le développement de l'approche de sous-échantillonnage des PSU au Viet Nam, pour que l'effet final soit de produire un échantillon PPS de PSU).
- Impossibilité de produire une pondération de base au cours du processus de pondération de l'échantillon, si les probabilités de sélection sont inconnues pour l'échantillon existant.
- Besoin probable de plus (voire beaucoup plus) d'assistance technique et de soutien, si le personnel technique du pays est inexpérimenté ou moins équipé pour gérer les complexités d'un échantillonnage à deux phases.
- La GATS est limitée par les caractéristiques de conception et le calendrier de recueil des données de l'échantillon existant (par ex., les définitions des strates).
- La comparabilité avec d'autres pays de la GATS sera possible dans la mesure où la conception de l'enquête de l'échantillon existant est conforme aux normes établies par le **Manuel de conception de l'échantillon de la GATS**.
- Les analystes des données de la GATS doivent faire face au défi technique qui consiste à adapter la conception de l'échantillon en deux phases, si l'échantillon de la GATS est un sous-échantillon aléatoire de l'échantillon existant.
- Nécessite la pleine coopération de ceux qui ont produit l'échantillon existant (pour élaborer un plan réalisable d'intégration, calculer la pondération de l'échantillon final de la GATS, documenter de manière efficace l'échantillon de la GATS, rendre les deux plans d'échantillonnage compatibles pour l'analyse, etc.).
- La qualité de l'échantillon final de la GATS dépend de la qualité de la conception et de la sélection de l'échantillon existant.

## Annexe C. Options de conception de l'échantillon pour les pays effectuant la GATS une nouvelle fois

---

Cette annexe décrit les orientations possibles de la conception de l'échantillon que les cycles subséquents de la GATS pourraient prendre dans les pays répétant la GATS. Bien que cette discussion s'applique à toute paire de cycles dans un pays où plusieurs cycles de la GATS sont terminés, le cycle de « référence » désigne généralement le premier cycle dans l'ordre chronologique, et « répétition » se réfère à un cycle ultérieur.

Trois options (R1 à R3) de conception de l'échantillon pour un cycle de répétition de la GATS sont présentées et brièvement discutées. Ces options présument que l'une des options N (ou une variante de celles-ci) mentionnées dans l'**annexe B** a été utilisée dans le cycle de référence. Il deviendra évident que les options R présentées ici varient suivant le degré avec lequel la composition de l'échantillon de la répétition dépend de la composition de l'échantillon de référence. Comme pour les nouveaux pays, le personnel de la GATS collaborera avec le personnel technique du pays organisant la répétition et le comité d'examen de l'échantillonnage de la GATS pour choisir l'une des options indiquées.

Voici quelques points à considérer lors du développement de l'ensemble des options de conception de l'échantillon pour les échantillons de pays répétant la GATS :

- Comme dans tous les cycles de la GATS dans les pays participants, la conception doit respecter les spécifications, comme décrites dans le **manuel de conception de l'échantillon de la GATS** qui demande généralement les caractéristiques de conception suivantes :
  - *Sélection aléatoire dans toutes les étapes d'échantillonnage pour produire des probabilités de sélection connues pour tous les membres de l'échantillon (c'est-à-dire que la conception de l'échantillon doit produire un « échantillon de probabilité » des répondants de la GATS).*
  - *Un minimum de 100 PSU.*
  - *Au moins deux étapes d'échantillonnage sont nécessaires pour choisir les foyers.*
  - *Intégrer la conception de l'échantillon à un autre plan d'échantillonnage national de renom est encouragé, surtout si cela a bien fonctionné dans le cycle de référence de la GATS.*
  - *La taille globale de l'échantillon des répondants doit être d'au moins 8 000, si des estimations de qualité de la GATS sont requises pour les sous-groupes de population définis conjointement par sexe et par catégories urbaines et rurales.*
  - *Si les ressources sont insuffisantes pour produire cette taille d'échantillon, les implications sur la qualité d'une taille d'échantillons de répondants plus petite doivent être discutées avec les membres du comité d'examen de l'échantillonnage de la GATS de sorte qu'il y ait une compréhension claire de ce que l'échantillon créé produira.*
- La discussion des options doit commencer par un examen détaillé de la conception de l'échantillon qui a été utilisée dans le cycle de référence de la GATS. Ceci est nécessaire pour orienter la discussion de conception vers ce qui est possible de réaliser dans le cycle de répétition.

- La comparabilité entre les cycles de recueil des données de la GATS est obtenue dans un pays si les conceptions à chaque cycle répondent aux normes de base du **Manuel de conception de l'échantillon de la GATS**, ce qui permet une certaine latitude dans les détails de la manière dont chaque échantillon probabiliste randomisé de la GATS est choisi. Ainsi, les aspects spécifiques de la conception pourront varier entre les cycles (par exemple, les définitions des strates, la taille des échantillons, le choix des unités d'échantillonnage, etc.). Il y a, cependant, certains avantages statistiques et pratiques à suivre exactement la même conception de l'échantillon à chaque cycle. Par exemple, utiliser le même échantillon de PSU à chaque cycle : (i) crée un chevauchement planifié de l'échantillon qui produira des estimations un peu plus précises des différences d'un cycle à l'autre (par rapport à des échantillons choisis indépendamment), et (ii) rend le recrutement des enquêteurs de terrain plus facile, puisque les mêmes enquêteurs utilisés au cours du premier cycle pourraient être embauchés pour le second.

La méthode d'échantillonnage la plus statistiquement utile, mais aussi la plus difficile à réaliser dans un cycle de répétition de la GATS consiste à revenir dans *tous les* emplacements *des foyers de l'échantillon déjà sélectionnés* (à savoir, les adresses résidentielles) qui ont été attribués au personnel de terrain dans le cycle de référence, permettant alors à un échantillon de cohorte complète de foyers choisis de devenir le résultat statistique final pour les deux cycles. Dans une approche de « cohorte complète », les enquêteurs reviennent à tous les emplacements *des foyers sélectionnés* lors du cycle de référence (ou premier cycle), y compris ceux qui : (i) ont conduit à des entretiens de répondant complets, (ii) ont été jugés non admissibles à l'étude, (iii) sont devenus des foyers non-répondants admissibles, et (iv) ont répondu mais où il n'y a pas eu d'entretien du répondant. Les résidents des foyers trouvés aux endroits échantillonnés au moment de la collecte de données pour le cycle de répétition en vertu de cette approche sont recrutés pour compléter un entretien de la GATS pour ce cycle, même si l'ensemble des résidents admissibles du foyer a changé depuis la référence, ou si un tout nouveau groupe de résidents a emménagé. Ainsi, seule une portion d'un échantillon chevauchant complètement des lieux de résidence pour l'échantillon de répétition impliquerait les mêmes foyers et individus que ceux du cycle de référence.

Le chevauchement complet des échantillons sélectionnés dans une enquête en plusieurs cycles est statistiquement avantageux, car cela améliore la qualité de l'estimation des différences d'un cycle à l'autre. La caractéristique de cette stratégie de conception multicycle qui la rend avantageuse est le chevauchement planifié (par opposition à non planifié ou aléatoire) des échantillons de foyers entre les deux cycles.<sup>1</sup>

Deux échantillons avec un chevauchement planifié (par opposition à non planifié, ou aléatoire) produisent généralement des estimations plus précises des différences entre les deux estimations spécifiques des échantillons qu'avec deux échantillons choisis indépendamment. Par ailleurs, plus le pourcentage de chevauchement planifié est élevé, plus le gain de précision des estimations de différences est grand. Plus précisément, si  $B$  et  $R$  sont les estimations des taux du cycle de référence et de répétition, respectivement, alors la variance de leur différence est  $Var(R - B) = Var(R) + Var(B) - 2\theta Cov(R, B)$ , où  $0 \leq \theta \leq 1$  est une mesure de l'effet du chevauchement planifié pour les deux plans d'échantillonnage.

---

<sup>1</sup>Un chevauchement non planifié est le fruit du hasard seul (autrement dit, il se trouve que des membres de la population ont été choisis dans les deux échantillons). En comparaison, un chevauchement planifié survient en désignant une partie de l'échantillon de référence comme membres de l'échantillon du cycle de répétition. Les chevauchements planifiés parmi les membres des deux échantillons peuvent être complets (résultant en un échantillon de cohorte) ou partiels.

Puisque  $Cov(R,B)$  est généralement positive et  $\theta$  commencerait peut-être à approcher 1 dans une méthode de chevauchement complet, si les conceptions du cycle de référence et de répétition ne diffèrent que par le niveau de chevauchement planifié,  $Var(R - B)$  sera inférieure suivant cette approche par rapport à toute autre option de conception où le pourcentage de chevauchement planifié est moindre (à savoir, les options R1 à R3 décrites ci-après).<sup>2</sup>

En supposant des taux de réponse similaires pour le cycle de référence et la répétition, cette approche aboutirait à peu près à la même taille d'échantillon de répondants pour le cycle de répétition que pour le cycle de référence. Si une taille d'échantillon de répondants supérieure était nécessaire pour le cycle de répétition, l'échantillon de référence pourrait être complété au besoin. De même, si un plus petit échantillon était nécessaire, une partie de l'échantillon pourrait être aléatoirement supprimée (voir OPTION R1, ci-dessous).

En raison de la dépendance de cette approche vis-à-vis de la disponibilité des dossiers d'informations de localisation des foyers sélectionnés qui ont été choisis pour le cycle de référence (qui, au moment où le cycle de répétition commence, peuvent dater de plusieurs années), et la nécessité de gérer les nouvelles constructions résidentielles et le changement de la composition de la population cible des individus de 15 ans ou plus, il semblait impossible d'envisager cette méthode pour l'échantillon dans un cycle de répétition. Pour ces raisons, il a été décidé (en consultation avec le personnel des CDC) de ne pas considérer l'échantillon de cohorte complète de foyers sélectionnés comme une des options à envisager pour la répétition des pays en phase I.

### **OPTION R1 : Conception de chevauchement partiel**

Dans cette option de conception, le plan est de revenir à une *partie de* l'échantillon *sélectionné* (le même) d'emplacements de foyers qui ont été attribués au personnel de terrain lors du cycle de référence, produisant ainsi un échantillon de cohorte partielle de foyers sélectionnés comme résultat statistique final. Cette méthode, sans aller jusqu'à utiliser des échantillons choisis indépendamment à chaque cycle, est une alternative qui semble plus pratique pour les échantillons de référence et de répétition qui se chevauchent complètement. Comme pour les échantillons se chevauchant complètement, cette option peut produire des estimations de différences plus précises dans le temps que les échantillons choisis indépendamment. Toutefois, ses gains en précision par rapport aux échantillons choisis indépendamment ne sont pas aussi importants que les échantillons qui se chevauchent complètement.

- *Le chevauchement partiel* de deux échantillons se produit quand un sous-ensemble aléatoire de membres du premier échantillon est désigné pour faire partie des membres du deuxième échantillon.
- Cette option de conception peut se produire dans les situations suivantes : (i) le pays a développé son propre échantillon autonome pour la référence et veut économiser sur les coûts de recueil de données en utilisant le même échantillon de PSU que la référence, mais il veut/a besoin de sélectionner un nouvel échantillon de foyers avec chaque PSU pour le cycle de répétition, (ii) le pays veut une nouvelle fois intégrer l'échantillon de la GATS à une autre

---

<sup>2</sup>  $\theta$  ne sera pas égal à 1, même dans une cohorte complète d'emplacements des foyers sélectionnés, puisque la composition du foyer et/ou le résident choisi pour l'entretien seront différents dans certains des endroits qui se chevauchent.

enquête et une partie de l'échantillon maître (donneur) doit être re-sélectionnée, ou (iii) le pays veut une plus petite taille d'échantillon de répondants et utiliser le même échantillon de foyers sélectionnés que celui du cycle de référence.

- Le chevauchement peut s'appliquer à une ou à toutes les étapes d'échantillonnage jusqu'à la sélection des foyers, notant que son application complète à toutes ces étapes (soit 100 % de chevauchement dans toutes les étapes d'échantillonnage) produit un échantillon de foyers sélectionnés qui se chevauchent complètement. Ainsi, une conception de chevauchement partielle a un chevauchement  $< 100\%$  au cours d'une étape au moins.
- Le chevauchement dans les phases tardives de l'échantillonnage (par exemple, au niveau du segment ou du foyer) est statistiquement plus efficace que les chevauchements au stade de la PSU de l'échantillonnage, puisqu'on tend à trouver plus de variation des principales variables des critères d'étude dans les unités d'échantillonnage utilisées dans les étapes tardives de sélection.
- Notez que la comparabilité des estimations du cycle de référence et de répétition *ne dépend pas* des changements qui peuvent être apportés à la méthode d'échantillonnage probabiliste pour le cycle de répétition, mais en revanche *dépend fortement* de la cohérence des éléments clés de la conception de l'enquête proposée pour le cycle de répétition (par exemple, la formulation des questionnaires et le contexte, la source et la qualité des listes de foyers utilisées pour les foyers échantillonnés, le protocole de recrutement de l'échantillon, le plan de contrôle de la qualité, le recrutement du personnel de terrain et la méthode de supervision).

#### **Avantages :**

- Cette option peut produire des estimations de différences entre le cycle de référence et le cycle de répétition statistiquement plus précises ; plus le pourcentage de chevauchement planifié est élevé, plus le gain de précision des estimations de différences est grand. Plus précisément,  $Var(R - B)$  pour l'OPTION R1 serait inférieure à  $Var(R - B)$  pour toute option de conception impliquant des échantillons choisis indépendamment (c'est-à-dire, les OPTIONS R2 ou R3), puisque  $\theta = 0$  pour des échantillons choisis indépendamment et donc  $Var(R - B) = Var(R) + Var(B)$  ce qui manque l'avantage de réduction de la variance produit par la covariance découlant du chevauchement planifié. Si les tailles d'échantillons ne varient pas beaucoup, la qualité des estimations transversales au cours du cycle de répétition sera presque la même qu'à la référence.

#### **Inconvénients :**

- Une limitation de cette option est que, si une partie des foyers échantillonnés se chevauchent en raison de la conception, le personnel de l'enquête doit être en mesure de localiser et de trouver des individus acceptant de répondre à la même adresse. Dans la plupart des cas, ils verront les mêmes foyers et les mêmes résidents que précédemment, donc il faudra convaincre les mêmes individus d'accepter de participer à l'entretien de la GATS une seconde fois.
- Le personnel de terrain devra être encadré et formé à vérifier avec le plus grand soin que chaque adresse résidentielle attribuée dans la portion de chevauchement de l'échantillon du

cycle de répétition est la même que l'adresse de l'échantillon correspondant de l'échantillon de référence.

- L'analyse pour produire des estimations de variance des estimations de différences entre le cycle de répétition et de référence sera plus complexe qu'avec les autres options, comme une partie de l'échantillon a une covariance du chevauchement planifié à intégrer tandis que le reste des échantillons sont choisis indépendamment. L'alternative serait d'ignorer ce terme de covariance mais cela entraînerait la surévaluation des variances de différences.
- Si les tailles d'échantillons ne varient pas beaucoup, la qualité des estimations transversales au cycle de répétition sera presque la même qu'à la référence.

### **OPTION R2 : Structure identique, mais la conception de sélection indépendante**

Dans cette option de conception, la structure (à savoir, les unités d'échantillonnage et les strates) de l'échantillon de répétition est identique à celle de l'échantillon de référence, mais les échantillons dans chacune des étapes de sélection sont établis de manière indépendante pour le cycle de répétition. Selon les souhaits du pays pour le cycle de répétition, les tailles de l'échantillon à chaque étape de sélection de la conception de l'échantillon du cycle de répétition peuvent ou non différer de la conception de l'échantillon de référence.

- Cette option de conception peut se produire dans les situations suivantes : (i) le pays souhaite utiliser la même conception d'échantillon que pour la référence mais veut/doit sélectionner un échantillon entièrement nouveau pour le cycle de répétition en utilisant les mêmes bases de sondage spécifiques de l'étape d'échantillonnage que celles utilisées pour sélectionner l'échantillon de référence, ou (ii) le pays souhaite utiliser la même conception de l'échantillon que pour la référence mais veut sélectionner un échantillon entièrement nouveau pour le cycle de répétition en utilisant des bases de sondage spécifiques à l'étape qui ont été mises à jour à partir de (et sont donc vraisemblablement plus à jour que) celles qui ont été utilisées pour la référence.
- Selon les souhaits du pays pour le cycle de répétition, les tailles de l'échantillon à chaque étape de sélection de la conception de l'échantillon du cycle de répétition peuvent ou non différer de la conception de l'échantillon de référence.
- Si le pays a produit un tableau récapitulatif de conception pour la conception de l'échantillon de la GATS au cours du cycle de référence (c'est-à-dire, un tableau récapitulatif des unités d'échantillonnage, l'utilisation de stratification, les méthodes de sélection, et les tailles d'échantillon pour chaque étape de l'échantillonnage), ce tableau pour le cycle de répétition aura exactement le même aspect.
- Même si la structure de conception de l'échantillon (à savoir, le nombre d'étapes, les unités d'échantillonnage, et l'utilisation de la stratification) est la même pour les deux cycles suivant cette option, les échantillons de répondants des cycles de référence et de répétition seront statistiquement indépendants puisque les échantillons à chaque étape sont choisis indépendamment pour chaque cycle.
- Notez que la comparabilité des estimations du cycle de référence et de répétition *ne dépend pas* des changements qui peuvent être apportés à la méthode d'échantillonnage probabiliste

pour le cycle de répétition, mais en revanche *dépend fortement* de la cohérence des éléments clés de la conception de l'enquête proposée pour le cycle de répétition (par exemple, la formulation des questionnaires et le contexte, la source et la qualité des listes de foyers utilisées pour les foyers échantillonnés, le protocole de recrutement de l'échantillon, le plan de contrôle de la qualité, le recrutement du personnel de terrain et la méthode de supervision).

### **Avantages :**

- Les échantillons choisis indépendamment n'ont aucun terme de covariance à intégrer dans l'estimation des variances des différences estimées entre le cycle de répétition et le cycle de référence, donc l'analyse est moins compliquée (c'est-à-dire, la variance de la différence estimée est simplement la somme de la variance de l'estimation du cycle de référence plus la variance de l'estimation du cycle de répétition, ou  $Var(R - B) = Var(R) + Var(B)$ ).
- Si le nombre d'étapes d'échantillonnage, les définitions des unités d'échantillonnage à chaque étape, et la taille des échantillons pour chaque étape sont les mêmes que dans la conception de l'échantillon pour le cycle de référence, la qualité des estimations transversales lors du cycle de répétition sera pratiquement la même qu'au cycle de référence.

### **Inconvénients :**

- Si le nombre d'étapes d'échantillonnage, les définitions des unités d'échantillonnage à chaque étape, et la taille des échantillons pour chaque étape sont les mêmes que dans la conception de l'échantillon pour le cycle de référence, la qualité des estimations transversales lors du cycle de répétition sera pratiquement la même qu'au cycle de référence.
- La précision statistique de l'estimation des différences ne sera pas aussi élevée que pour celle de l'OPTION R1. Plus précisément,  $Var(R - B)$  pour l'OPTION R1 serait inférieure à  $Var(R - B)$  pour toute option de conception de répétition impliquant des échantillons choisis indépendamment, puisque  $\theta = 0$  pour des échantillons choisis indépendamment et donc  $Var(R - B) = Var(R) + Var(B) - 2\theta Cov(R, B) = Var(R) + Var(B)$  pour cette option, donc manquant l'avantage de réduction de la variance produit par la covariance découlant du chevauchement planifié.

### **OPTION R3 : Conception de sélection totalement indépendante**

Dans cette option de conception, le pays repose sur une conception de l'échantillon complètement différente dans le cycle de répétition de celui utilisé dans le cycle de référence. Cela signifie que des unités d'échantillonnage différentes peuvent être utilisées dans des étapes de sélection précoces et une autre personne et/ou organisation statistique fournira l'échantillon.

- Cette option de conception peut se produire dans les situations suivantes : (i) une nouvelle organisation de statistique fournira l'expertise d'échantillonnage pour sélectionner l'échantillon du cycle de répétition, autre que celle utilisée pour le cycle de référence, (ii) l'échantillon d'une autre enquête nationale est utilisé pour produire un échantillon pour la GATS qui est intégré à l'échantillon d'une autre enquête ou échantillon maître, ou (iii) l'échantillon de la GATS pour le cycle de référence a été intégré à une autre enquête ou échantillon maître, mais le pays décide

qu'il souhaite développer une conception autonome pour le cycle de répétition et les cycles subséquents.

- Selon les souhaits du pays pour le cycle de répétition, les tailles de l'échantillon à chaque étape de sélection de la conception de l'échantillon du cycle de répétition peuvent ou non différer de la conception de l'échantillon de référence.
- Notez que la comparabilité des estimations du cycle de référence et de répétition ne dépend pas des changements qui peuvent être apportés à la méthode d'échantillonnage probabiliste pour le cycle de répétition, mais en revanche dépend fortement de la cohérence des éléments clés de la conception de l'enquête proposée pour le cycle de répétition (par exemple, la formulation des questionnaires et le contexte, la source et la qualité des listes de foyers utilisées pour les foyers échantillonnés, le protocole de recrutement de l'échantillon, le plan de contrôle de la qualité, le recrutement du personnel de terrain et la méthode de supervision).

### **Avantages :**

- Les échantillons choisis indépendamment n'ont aucun terme de covariance à intégrer dans l'estimation des variances des différences estimées entre le cycle de répétition et le cycle de référence, donc l'analyse est moins compliquée (c'est-à-dire, la variance de la différence estimée est simplement la somme de la variance de l'estimation du cycle de référence plus la variance de l'estimation du cycle de répétition, ou  $Var(R - B) = Var(R) + Var(B)$ ).
- Si le nombre d'étapes d'échantillonnage et la taille des échantillons pour chaque étape sont au moins approximativement les mêmes que dans la conception de l'échantillon pour le cycle de référence, la qualité des estimations transversales lors du cycle de répétition sera pratiquement la même qu'au cycle de référence.

### **Inconvénients :**

- Si le nombre d'étapes d'échantillonnage et la taille des échantillons pour chaque étape sont au moins approximativement les mêmes que dans la conception de l'échantillon pour le cycle de référence, la qualité des estimations transversales lors du cycle de répétition sera pratiquement la même qu'au cycle de référence.
- Si différents organismes de statistiques, avec une qualité variable de base de sondage, stratégies de recrutement de l'échantillon, et mesure de l'enquête, sont impliqués dans la production des échantillons pour les deux cycles, un inconvénient possible de cette option est que la comparaison des estimations de l'enquête peut être faussée par ces différences de qualité.
- La précision statistique de l'estimation des différences ne sera pas aussi élevée que pour celle de l'OPTION R1. Plus précisément,  $Var(R - B)$  pour l'OPTION R1 serait inférieure à  $Var(R - B)$  pour toute option de conception de répétition impliquant des échantillons choisis indépendamment, puisque  $\theta = 0$  pour des échantillons choisis indépendamment et donc  $Var(R - B) = Var(R) + Var(B) - 2\theta Cov(R, B) = Var(R) + Var(B)$  pour cette option, donc manquant l'avantage de réduction de la variance produit par la covariance découlant du chevauchement planifié.







GLOBAL TOBACCO SURVEILLANCE SYSTEM (GTSS)

