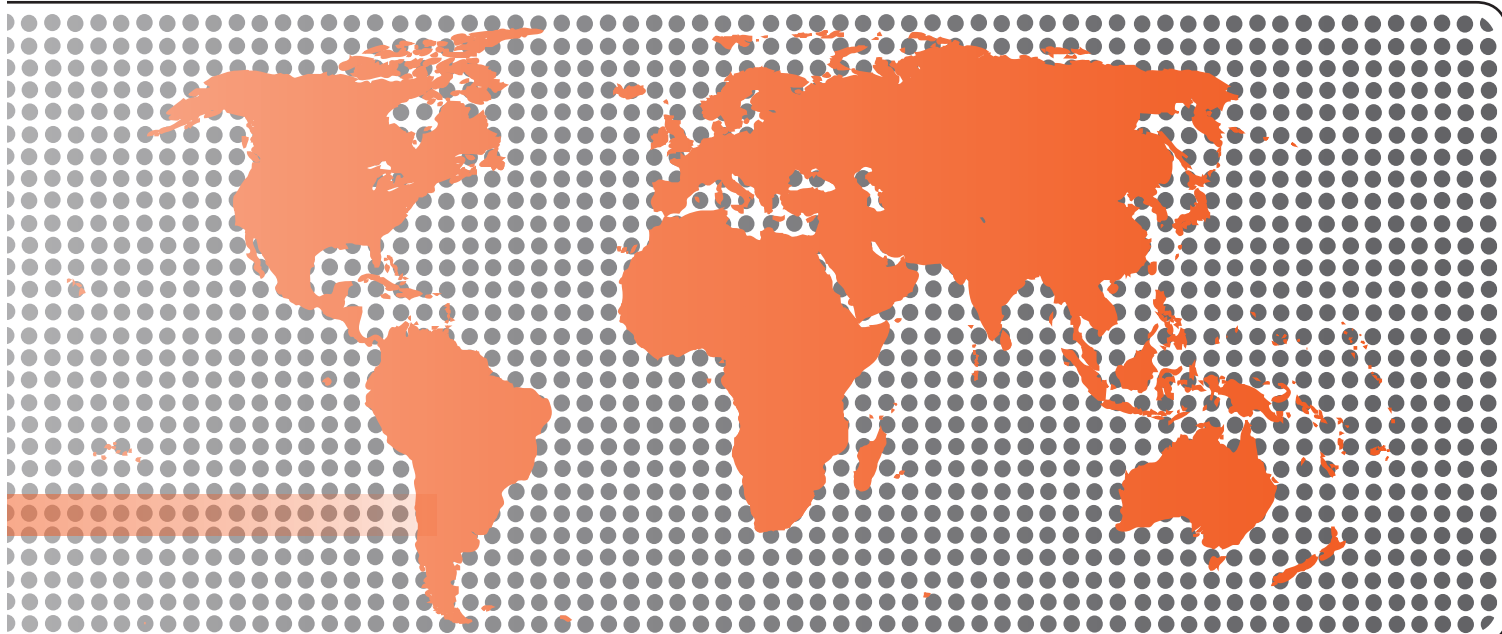


GTSS Global Adult Tobacco Survey (GATS)



Manuel de pondération de l'échantillon



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

Manuel de l'Enquêteur de terrain
Manuel du Superviseur de terrain
Manuel de cartographie et de listage

Gestion des données de l'enquête GATS

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

Assurance qualité GATS : Directives et documentation

Kit d'analyse et de rapport de l'enquête GATS

Modèle Fiche d'Information
Rapport du Pays : Plan et lignes directrices relatives aux tabulations
Définitions des indicateurs

Publication et diffusion des données de l'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 Weights 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
- CDC Foundation
- Johns Hopkins Bloomberg School of Public Health
- RTI International
- University of North Carolina Gillings School of Public Health
- 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
1. Introduction	1-1
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
2. Vue d'ensemble de la pondération de l'échantillon du GATS	2-1
3. Approche recommandée	3-1
3.1 Pondération de base	3-2
3.2 Ajustement pour la non-réponse d'une unité	3-5
3.3 Ajustement de l'étalonnage post-stratification	3-12
4. Assurer la qualité des pondérations de l'échantillon du GATS	4-1
5. Bibliographie	5-1

1. Introduction

Le tabagisme est une cause majeure évitable de mort prématurée et de maladies dans le monde. Environ 5,4 millions de personnes meurent chaque année en raison de maladies liées au tabac, un chiffre qui devrait dépasser les 8 millions par an d'ici à 2030. Si les tendances actuelles se confirment, le tabagisme sera responsable de la mort d'un milliard de personnes d'ici à la fin de ce siècle. On estime que plus des trois quarts de ces décès auront lieu dans les pays à faible et moyen revenu¹. Un mécanisme de surveillance efficace et systématique est indispensable pour suivre et gérer 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 lutte antitabac. Le GATS est une enquête représentative au niveau national, menée auprès de foyers d'adultes de 15 ans ou plus, en utilisant un questionnaire principal standard, une conception de l'échantillon, et des procédures de collecte 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 lutte antitabac.

Afin de maximiser l'efficacité des données recueillies par le biais du 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 le GATS. Ils visent également à apporter des conseils sur la façon dont un pays donné peut adapter les caractéristiques du protocole du GATS en vue de maximiser l'utilité des données dans le pays. Il est vivement recommandé de suivre le protocole standard, afin de maintenir la cohérence et la comparabilité entre les pays.

Les manuels du 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

Le 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 qui est leur lieu habituel de résidence.

Le 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 au GATS.

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

L'entretien de la GATS se compose de deux parties : le *Questionnaire Ménage* et le *Questionnaire Individuel*. Le *questionnaire Ménage* (pré-sélection du foyer) et le *questionnaire individuel* (entretien individuel) seront effectués en utilisant un dispositif électronique de recueil des données.

Pour chaque adresse de l'échantillon, les enquêteurs de terrain administreront le *Questionnaire Ménage* à un adulte qui fait partie du foyer. Les objectifs du *questionnaire Ménage* sont de déterminer si le ménage sélectionné répond aux exigences d'admissibilité du GATS, et de faire une liste, ou un tableau, de tous les membres admissibles du ménage. 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 fumée, 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.

L'entretien du GATS comporte deux parties : le *Questionnaire Ménage* et le *questionnaire individuel*. Ces questionnaires sont administrés en utilisant un dispositif électronique de collecte de données.

1.2 Utilisation de ce manuel

Ce manuel présente les exigences auxquelles les pays doivent se conformer pour calculer les pondérations de l'échantillon pour le GATS. Ce premier chapitre fournit des informations générales sur le GATS. Les chapitres suivants sont résumés ci-dessous :

- Le **chapitre 2** définit la pondération de l'échantillon, indique le moment où les pondérations sont calculées dans le processus de l'enquête, donne un aperçu de la logique du processus de calcul et décrit comment les pondérations sont utilisées par l'analyste de données.
- Le **chapitre 3** décrit en détail chacune des étapes de la méthode recommandée pour calculer les pondérations. Chacune de ces étapes est illustrée par un exemple basé sur le type de répondant susceptible de se trouver dans un échantillon du GATS.
- Le **chapitre 4** recommande plusieurs mesures pour assurer la qualité élevée des pondérations.

Il convient de noter que bien que ce manuel offre un modèle étape par étape pour le calcul des pondérations de l'enquête, des modifications peuvent devoir être apportées à celui-ci en raison de circonstances propres à chaque pays liées à la conception de l'échantillon, des besoins en analyse et de la disponibilité de données additionnelles pour calculer les corrections. La collaboration entre les statisticiens du pays, les points centraux des CDC du pays et le comité d'examen de l'échantillon (SRC) est essentielle pour assurer la qualité des pondérations de l'échantillon de l'enquête et des corrections.

2. Vue d'ensemble de la pondération de l'échantillon du GATS

Une *pondération de l'échantillon* est une mesure statistique liée à un enregistrement de données pour tout répondant des échantillons de la population utilisant de bout en bout les méthodes de sélection aléatoire pour choisir l'échantillon. En termes généraux, une pondération individuelle de l'échantillon est généralement simplement l'inverse de la probabilité corrigée d'obtention des données du répondant. Dans la plupart des cas, cette probabilité est simplement la probabilité de sélection originale du répondant basée sur la conception de l'échantillon. La probabilité inverse, ou la pondération de base, (B_j) est généralement corrigée pour tenir compte des déséquilibres involontaires de l'échantillon au cours du processus de réalisation de l'enquête¹. Plus d'un ajustement par pondération peut être appliquée. Toutes sont multiplicatives. Sauf si une pondération est redimensionnée à des fins analytiques (par exemple, « étalonnée » pour faire la somme de la population totale couverte), sa valeur peut être interprétée comme une indication du nombre de membres de la population représentée par le répondant. Des séries distinctes de pondération peuvent être nécessaires lorsque les données à analyser sont réunies pour différents types d'éléments de données ou unités d'analyse liées au répondant. Par exemple, si les données dans une enquête de ménages sont rassemblées pour les ménages sélectionnés, et pour un résident choisi au hasard dans chacun de ces foyers, des séries distinctes de pondérations sont produites pour les données du ménage et pour les données du résident. Cependant, puisque seules les données au niveau individuel doivent être traitées dans le GATS, seules les pondérations au niveau individuel seront nécessaires.

Alors que le raisonnement statistique général qui sous-tend l'utilisation de pondérations pour l'estimation à partir d'échantillons de la population est bien établi (Horvitz et Thompson, 1952), aucun protocole universel n'existe pour les calculer. Cela est en partie dû à des variations de circonstances d'un échantillon à l'autre pour leur conception, la qualité de la documentation de la sélection de l'échantillon et les procédures de recrutement, la disponibilité des informations complémentaires sur l'échantillon et la population afin d'identifier et de traiter les déséquilibres dans l'échantillon causés par une couverture de la base de sondage différentielle et par les non-réponses. Ainsi, les étapes de calcul de pondération d'un échantillon peuvent varier selon les enquêtes. Toutefois, une combinaison des étapes ci-dessous est généralement suivie par la production d'une pondération pour chaque (c.-à-d. le *j-ème*) enregistrement de donnée d'un répondant, avec la pondération finale corrigée (W_j) étant le produit de la valeur générée à chaque étape, comme décrit plus en détail ci-dessous :

1. déterminer la pondération de base (B_j) pour tenir compte de toutes les étapes de sélection aléatoire qui ont conduit à l'échantillon de membres de la population,
2. Ajuster en fonction de la non-réponse (A_j^{nr}) pour compenser le déséquilibre de l'échantillon dû à la réussite différentielle du recrutement de l'échantillon,

¹Le concept de « déséquilibre », dans ce cas, indique seulement que la représentation démographique de la population échantillonnée est faussée d'une certaine manière par des forces liées à la sélection de l'échantillon et aux phases de recrutement de l'étude. En d'autres termes un échantillon par ailleurs représentatif, produit par une sélection aléatoire, devient un peu moins représentatif de la population.

3. en outre, corriger pour une couverture incomplète de la base de sondage ($A_j^{(cov)}$) pour compenser le déséquilibre dû à une base de sondage qui ne couvre pas entièrement la population ciblée par l'étude, et
4. procéder à un autre ajustement pour étalonner ($A_j^{(cal)}$) la série finale de pondérations corrigées en fonction de la distribution de la population suivant des caractéristiques étroitement corrélées aux principales mesures clés des résultats de l'étude (à savoir, le comportement tabagique dans le contexte du GATS).

La séquence réelle des étapes suivies pour la production de la série des pondérations statistiques pour l'échantillon de population est importante, puisque la pondération engendrée par une étape donnée quelle qu'elle soit dépendra des résultats de calcul de chaque étape précédente. La pondération finale de tout répondant est le produit des résultats de calcul pour le répondant à partir de toutes les étapes dans l'ordre :

$$W_j = B_j A_j^{(nr)} A_j^{(cov)} A_j^{(cal)} \quad (1)$$

Pour comprendre la logique du processus de production des pondérations pour les échantillons de l'enquête, où les résidents des ménages sélectionnés sont membres de la population cible, considérez d'abord qu'un membre de la population du GATS ne peut fournir des données que si l'ensemble des trois événements suivants (à savoir F, S, et R) se produit :

- **Événement F** : Le membre du foyer et le membre sont inclus dans les bases de sondage du ménage et des individus du foyer utilisées pour l'échantillonnage du GATS ;
- **Événement S** : si l'événement F se produit, le membre du ménage et le membre doivent être choisis au hasard pour participer au GATS, et
- **Événement R** : si l'événement S se produit, le membre du ménage et le membre doivent, à leur tour, accepter de devenir un participant du GATS et de répondre à un questionnaire valide.

La probabilité que les données du membre soient utilisées pour l'estimation de l'échantillon est le produit des probabilités d'observation de ces trois événements. Nous verrons plus tard que l'achèvement des étapes (1), (2) et (3) exige de déterminer ou d'estimer les probabilités des événements F, S et R. L'étalonnage de l'échantillon (ou « post-stratification ») sert à contrôler la perte de validité externe de l'échantillon due à des forces qui ne sont pas spécifiquement intégrées par les corrections pour la couverture et la non-réponse dans les étapes (2) et (3). Ces forces peuvent être des variations aléatoires dans la composition démographique de l'échantillon par rapport aux variables non utilisées pour définir les strates d'échantillonnage ainsi que la non-réponse différentielle et la couverture incomplète associées à des variables autres que celles qui définissent les cellules de correction utilisées pour produire les corrections dans les étapes (2) et (3). Comme il est souvent impossible de réaliser l'étape (3), la correction pour une couverture de base incomplète est généralement réalisée dans le cadre de l'étalonnage.

Les pondérations de l'échantillon sont importantes pour de nombreuses situations d'analyse. Par exemple, elles sont utilisées pour produire des estimations ponctuelles des caractéristiques de la

population (par exemple, un taux de prévalence de tabagisme actuel) ainsi que pour l'estimation de la variance de ces caractéristiques, bien que d'autres caractéristiques de la conception (à savoir la stratification, l'échantillonnage en grappes, la sélection sans remplacement) soient également importantes pour ces dernières. Ainsi, les pondérations sont importantes pour de nombreuses utilisations de données, y compris l'analyse descriptive et les essais comparatifs (par exemple, des différences significatives parmi les taux de prévalence des régions d'un pays). Elles sont aussi parfois utilisées pour la modélisation de la régression (par exemple, pour identifier les paramètres prédictifs du comportement tabagique). La pondération, par conséquent, est souvent une caractéristique de conception nécessaire, mais pas toujours suffisante à utiliser dans l'analyse des données d'enquête. Par exemple, si l'on utilise uniquement la pondération, mais que l'on ignore la stratification et l'échantillonnage par grappes en estimant la variance des estimations de l'enquête, les variances rapportées et les conclusions des tests d'hypothèse peuvent très bien être incorrectes.

Une série de pondérations est produite pour chaque ensemble d'unités d'observation dans l'étude. Des études ont souvent de multiples ensembles de ces unités d'observation. Par exemple, si l'étude du plan de recueil de données de l'étude implique la collecte des informations sur les ménages (par exemple, le revenu, SES, la distance par rapport aux professionnels de santé) et les membres de ces ménages (par exemple, le comportement tabagique, les informations démographiques, etc.), un ensemble distinct de pondérations peut être nécessaire pour chaque unité d'observation (à savoir les ménages et les personnes). Par ailleurs, la pondération peut être différente pour les ménages et leurs résidents répondants si l'échantillonnage est effectué dans le ménage pour choisir les individus fournissant les données au niveau des personnes. Ce genre de différence serait vrai dans le GATS si les données au niveau du ménage et individuel sont produites, puisqu'un résident du ménage admissible est sélectionné dans chaque ménage participant.

Le processus de production des pondérations a généralement lieu après la collecte des données et une fois que les données ont été traitées et nettoyées pour l'analyste. Elles ne peuvent être générées qu'après la fin du travail sur le terrain car elles sont appliquées à l'échantillon final de répondants et leur calcul s'appuie sur les informations des résultats finaux du recueil des données. Leurs calculs doivent également être terminés pour que l'analyse puisse commencer, puisque les analystes utilisant les données de l'enquête en auront besoin. Le personnel technique participant au choix de l'échantillon et au soutien du travail sur le terrain est généralement mieux équipé pour produire les pondérations en raison de ses connaissances de la population cible et de la façon dont l'échantillon a été créé et recruté. Au besoin, le personnel central du GATS peut fournir une assistance et un appui pour le processus de calcul des pondérations. Enfin, pour assurer la qualité des pondérations, le processus de leur calcul est entièrement revu par des experts externes pour s'assurer que les procédures appropriées sont suivies (voir **section 5 d'Assurance qualité du GATS : directives et documentation**).

3. Approche recommandée

Ce chapitre de ce manuel expose en détail la démarche en trois étapes correspondant aux principales composantes que le personnel technique du pays a pour recommandation de suivre lors de la production des pondérations pour l'échantillon du GATS. Ces étapes sont les suivantes : calculer une pondération de base pour chaque répondant de l'échantillon, corriger les pondérations de base en fonction de la non-réponse et étalonner les pondérations corrigées en fonction des populations totales connues. Une discussion plus générale de chacune de ces étapes est présentée dans les revues de Lessler et Kalsbeek (1992) et Kalton et Flores-Cervantes (2003). Notre approche, en présentant ces étapes, consiste à intégrer un exemple réaliste similaire au GATS dans la discussion. La discussion de chacune des trois étapes commence par énumérer ce qui est nécessaire pour exécuter l'étape, puis présente les formules spécifiques recommandées pour calculer la composante de la pondération dans le cadre de l'exemple. Des alternatives à certaines recommandations sont proposées pour examen par le personnel du GATS du pays, même si le choix de ces alternatives doit être pleinement justifié et fait avec l'entière collaboration des points focaux du GATS du pays et le comité d'examen de l'échantillon (SRC) avant de les mettre en application.

RECOMMANDATION :

La démarche suivante en trois étapes est recommandée pour chaque échantillon national [pays] du GATS :
(1) calculer une pondération de base pour chaque répondant de l'échantillon, (2) ajuster les pondérations de base en fonction de la non-réponse et (3) étalonner les pondérations ajustées en fonction des populations totales connues

Contexte de l'exemple

L'exemple utilisé pour illustrer le processus de calcul des pondérations du GATS se concentre sur le problème de la détermination de la pondération de l'échantillon final corrigé pour un individu fictif répondant du GATS (appelé \mathcal{R} par la suite) dans un échantillon du pays où la randomisation des sexes n'était pas requise. Nous supposons que la conception de l'échantillon qui a conduit à la sélection de cette personne est un échantillon de foyers stratifiés en trois étapes avec des unités de surface géopolitiques reconnues de taille variable comme unité primaire d'échantillonnage (PSU), un « segment » de taille approximativement égale, composé d'environ 200 foyers comme unités d'échantillonnage secondaires (SSU), et des ménages individuels sélectionnés au sein des segments échantillonnés. Dans cet exemple, nous supposons qu'aucun foyer n'a été manqué en établissant les listes des foyers du segment de \mathcal{R} et que l'attribution des sexes ne faisait pas partie du processus de sélection de \mathcal{R} , mais dans la pratique, ces questions pourraient avoir besoin d'être abordées (les méthodes appropriées pour répondre à ces situations sont incluses ultérieurement dans ce manuel)¹. Enfin, nous supposons que \mathcal{R} est choisi dans une liste des ménages qui a été entrée dans un ordinateur

EXEMPLE ILLUSTRATIF :

Renseignements généraux sur le répondant fictif de la GATS (\mathcal{R})

¹La sous-sélection des foyers manqués est nécessaire lorsque des foyers qui ne figurent pas dans la base de liste des foyers sont découverts grâce à une procédure d'unités de logement manquées (par exemple, la technique des intervalles « semi-ouverts ») et que le nombre de foyers découverts est suffisamment important pour justifier une sous-sélection aléatoire des foyers ne figurant pas sur la liste. La répartition des sexes est à faire avec les échantillons de la GATS lorsque, pour des raisons culturelles, les pays souhaitent faire correspondre le sexe des enquêteurs avec celui des répondants sélectionnés et utiliseront des équipes d'enquêteurs du même sexe, ou lorsqu'un groupe d'un sexe donné doit être suréchantillonné pour des raisons statistiques. Au moyen d'une attribution du sexe aléatoire, chaque foyer échantillonné sera un « foyer masculin » (seuls des résidents admissibles masculins seront listés pour la sélection des répondants), ou un « foyer de sexe féminin » (seules des résidentes admissibles seront inscrites sur la liste).

de poche, également utilisé pour effectuer l'entretien du GATS avec \mathcal{R} . D'autres détails admis pour la sélection et le recrutement de \mathcal{R} seront mentionnés dans la description de chaque étape.

3.1 Pondération de base

Nécessaire pour effectuer cette étape de calcul :

- Probabilités de sélection bien documentées pour chaque étape du processus de sélection de l'échantillon si, comme il est préférable, elles ont déjà été calculées.

OU

- Connaissances spécifiques sur la façon de calculer les probabilités correspondant aux méthodes de sélection aléatoire utilisées dans chaque étape du processus d'échantillonnage. Les étapes de sélection comprennent : pour les étapes d'échantillonnage, la répartition des sexes des ménages (si nécessaire) et la sous-sélection des ménages découverts grâce à une procédure de foyers manqués le cas échéant.
- Fiches de sélection et/ou code informatique pour le logiciel de sélection pour chaque étape du processus de sélection de l'échantillon, si les probabilités de sélection doivent être déterminées au moment où les pondérations sont calculées.

Calcul de la pondération de base de l'exemple :

La pondération de base d'un répondant dans un échantillon de probabilité est tout simplement un divisé par la probabilité de sélection globale du répondant en supposant que les étapes de sélection du répondant sont terminées. Le calcul de pondération de base pour un répondant du GATS exige donc que l'on réponde à la question : quelle était la probabilité statistique que la conception de l'échantillon conduise à la sélection du répondant ?

Le **manuel de conception de l'échantillon du GATS (chapitre 11)** décrit les éléments suivants concernant la probabilité globale de sélection lorsque le pays suit la méthode d'échantillonnage recommandée en plusieurs étapes. Les indices α et k (conjointement pour la α -ème PSU et k -ème SSU, respectivement) dans cette description conjointement correspondant à la « zone » (segment) g , qui est choisie en deux étapes d'échantillonnage dans la sélection de \mathcal{R} :

$P_{\alpha k}^{(1)}$ = probabilité inconditionnelle de sélection de la α -ème PSU (unité de surface géopolitique dans laquelle \mathcal{R} vit) et de la k -e SSU (segment dans lequel \mathcal{R} vit) ;

$P_{\alpha k}^{(2)}$ = probabilité conditionnelle (en admettant les sélections de la PSU et de la SSU) de sélection du foyer dans lequel \mathcal{R} vit ;

$P_{\alpha ki}^{(3)}$ = pour un répondant fictif du GATS (\mathcal{R}), cette probabilité = 1, cependant, la probabilité conditionnelle (en admettant les sélections de la PSU, la SSU et du foyer) de l'assignation aléatoire du ménage de \mathcal{R} au groupe des ménages « féminins » est requise si la randomisation des sexes est utilisée ;

$P_{\alpha ki}^{(4)}$ = probabilité conditionnelle (en admettant les sélections de la PSU, la SSU, du ménage et la répartition des sexes) de sélectionner aléatoirement le foyer de \mathcal{R} , s'il n'était pas dans

la base des foyers d'origine pour le αk -ème segment et a été découvert dans le cadre d'une procédure de foyers manqués, sinon cette probabilité = 1 ; et

$p_{\alpha k i j}^{(5)}$ = probabilité conditionnelle (en admettant les sélections de la PSU, la SSU, les ménages, la répartition des sexes et la sélection des foyers manqués) de sélectionner aléatoirement \mathcal{R} à partir d'une liste des résidents admissibles du *foyer* de \mathcal{R} .

Notez que chacun des événements de sélection correspondant à ces probabilités doit se produire dans l'ordre pour sélectionner \mathcal{R} dans l'échantillon du GATS. La probabilité conjointe inconditionnelle de sélectionner \mathcal{R} (la $\alpha k i j$ -ème personne) dans l'échantillon du GATS est alors la suivante :

$$p_{\alpha k i j} = p_{\alpha k}^{(1)} \cdot p_{\alpha k}^{(2)} \cdot p_{\alpha k i}^{(3)} \cdot p_{\alpha k i}^{(4)} \cdot p_{\alpha k i j}^{(5)} \quad (2)$$

puisque la probabilité conjointe d'événements séquentiels est la probabilité inconditionnelle du premier événement de la séquence multipliée par les probabilités conditionnelles de chaque événement ultérieur au vu du résultat des événements antérieurs dans la séquence. Ainsi, la pondération de base associée à \mathcal{R} est la suivante :

$$B_{\alpha k i j} = \frac{1}{p_{\alpha k i j}} = \frac{1}{p_{\alpha k}^{(1)} \cdot p_{\alpha k}^{(2)} \cdot p_{\alpha k i}^{(3)} \cdot p_{\alpha k i}^{(4)} \cdot p_{\alpha k i j}^{(5)}} \quad (3)$$

Comme vu dans l'équation (3), nous devons déterminer chacune des composantes de la probabilité conjointe inconditionnelle de sélectionner \mathcal{R} afin de calculer la pondération de base de \mathcal{R} . En supposant qu'une certaine forme d'échantillonnage sans remplacement avec une probabilité proportionnelle à la taille (PPS) est utilisée pour sélectionner les PSU au sein des strates d'échantillonnage de la première étape, si $N_i = 2,462$ est la mesure de la taille (en nombre de ménages à la date du dernier recensement) pour la PSU de \mathcal{R} , $I = 2$ est le nombre de PSU choisies dans la strate d'échantillonnage à partir de laquelle la PSU de \mathcal{R} a été choisie, et la somme des mesures de taille pour toutes les PSU dans cette strate est $\sum_{\alpha} N_{\alpha} = 338,754$, la probabilité inconditionnelle de sélection pour la PSU de \mathcal{R} sera,

$$p_{\alpha}^{(1)} = \frac{I \cdot N_{\alpha}}{\sum_{\alpha} N_{\alpha}} = \frac{(2) \cdot (2,462)}{338,754} = 1.4536 \times 10^{-2} \quad (4)$$

Si la SSU de \mathcal{R} a $K_{\alpha} = 2$ segments choisis par échantillonnage aléatoire simple (sans remplacement) à partir de $S_{\alpha} = 12$ segments dans la PSU de \mathcal{R} , la probabilité conditionnelle (étant donnée la sélection des PSU) de la sélection de la SSU de \mathcal{R} est alors,

$$p_{k(\alpha)}^{(1)} = \frac{K_{\alpha}}{S_{\alpha}} = \frac{2}{12} = \frac{1}{6} \quad (5)$$

et la probabilité conjointe inconditionnelle de la sélection de la PSU de \mathcal{R} et de la SSU de \mathcal{R} est,

$$p_{\alpha k}^{(1)} = p_{\alpha}^{(1)} p_{k(\alpha)}^{(1)} = \left[\frac{I \cdot N_{\alpha}}{\sum_{\alpha} N_{\alpha}} \right] \cdot \left[\frac{K_{\alpha}}{S_{\alpha}} \right] = (1.4536 \times 10^{-2}) \cdot (1/6) = 2.4226 \times 10^{-3} \quad (6)$$

Le **Manuel de conception de l'échantillon du GATS** décrit deux méthodes de sélection des ménages à partir de segments ou de PSU sélectionnés. Si l'échantillonnage systématique est utilisé pour sélectionner les ménages, la probabilité conditionnelle de la sélection de chaque ménage est $1/K$ où K est l'intervalle de sélection. Si un échantillonnage aléatoire simple est utilisé pour sélectionner les ménages, cette probabilité est le rapport de la taille de l'échantillon des foyers dans chaque segment sur le nombre total de ménages de la base de liste pour chaque segment. En supposant que $H_{\alpha k} = 28$ ménages sont sélectionnés par échantillonnage aléatoire simple sans remplacement à partir de $L_{\alpha k} = 212$ foyers figurant dans le αk -ème segment dans lequel le ménage de \mathcal{R} est situé, la probabilité conditionnelle de la sélection du foyer de \mathcal{R} est alors,

$$p_{\alpha k}^{(2)} = \frac{H_{\alpha k}}{L_{\alpha k}} = \frac{28}{212} = 0.13208 \quad (7)$$

Puisque le répondant fictif se trouve dans un pays où la randomisation des sexes n'a pas été nécessaire, la probabilité de sélection du répondant,

$$p_{\alpha k i}^{(3)} = 1 \quad (8)$$

Notez qu'en général, lorsqu'un sexe doit être attribué aux ménages, et $M_{\alpha k}$ et $F_{\alpha k}$ sont, respectivement, les nombres de $H_{\alpha k} = M_{\alpha k} + F_{\alpha k}$, ménages sélectionnés dans la αk -ème « zone » (segment) attribuée aux groupes « masculin » et « féminin », alors pour les répondants de sexe masculin,

$$p_{\alpha k i}^{(3)} = \frac{M_{\alpha k}}{H_{\alpha k}} \quad (9)$$

et pour les répondants de sexe féminin,

$$p_{\alpha k i}^{(3)} = \frac{F_{\alpha k}}{H_{\alpha k}} \quad (10)$$

En général, si un répondant est sélectionné dans un foyer qui a été découvert par le biais d'une procédure de foyers manqués (par exemple, la technique des intervalles semi-ouverts) et un taux d'échantillonnage ($f_{\alpha k i}$) est appliqué en sous-sélectionnant le ménage par échantillonnage aléatoire simple, la probabilité de sous-sélection pour tout (à savoir, le $\alpha k i$ -e) ménage choisi de cette façon sera alors,

$$p_{\alpha k i}^{(4)} = f_{\alpha k i} \quad (11)$$

Comme nous supposons que le ménage de \mathcal{R} n'a pas été sélectionné dans le cadre d'une procédure de ménages manqués,

$$p_{aki}^{(4)} = 1 . \quad (12)$$

Pour le cas spécifique de la détermination de la probabilité de sélection de \mathcal{R} au sein du ménage, nous remarquons que le répondant a été choisi au hasard parmi les $R_{aki} = 4$ membres du ménage du répondant qui ont été inclus sur la liste du foyer Ainsi, pour le ménage,

$$p_{akij}^{(5)} = \frac{1}{R_{aki}} = \frac{1}{4} = 0.25 \quad (13)$$

En combinant toutes les probabilités pour le processus de sélection qui a conduit à choisir \mathcal{R} , nous obtenons, en résumé, que la probabilité de sélection globale inconditionnelle de \mathcal{R} est

$$P_{akij} = p_{ak}^{(1)} \cdot p_{ak}^{(2)} \cdot p_{aki}^{(3)} \cdot p_{aki}^{(4)} \cdot p_{akij}^{(5)} = (2.4226 \times 10^{-3}) \cdot (0.13208) \cdot (1) \cdot (1) \cdot (0.25) = 7.9992 \times 10^{-5} , \quad (14)$$

et la pondération de base du répondant est,

$$B_{akij} = \frac{1}{p_{akij}} = \frac{1}{7.9992 \times 10^{-5}} = 12501.3081 , \quad (15)$$

et que la pondération de base pour le ménage est

$$B_{aki} = p_{akij}^{(5)} / p_{akij} = (0.25) \cdot (12501.3081) = 3125.2370 \quad (16)$$

Remarquez que B_{akij} est notée B_j dans les sections suivantes du manuel.

3.2 Correction pour la non-réponse d'une unité

Nécessaire pour effectuer cette étape de calcul :

- La pondération de base pour chaque répondant du GATS.
- La décision finale de recrutement (par exemple, réponse, refus, absent, etc.) **pour tous les ménages sélectionnés** à utiliser dans le calcul des taux de réponse des ménages parmi les foyers admissibles de l'échantillon dans chaque « segment » de l'échantillon. Cela signifie que le dénombrement des ménages participants sélectionnés, admissibles à l'étude sera nécessaire pour chaque segment de l'échantillon. Pour les conceptions d'échantillon qui sélectionnent directement les ménages dans les unités d'échantillonnage de la première étape, les décisions finales de recrutement dans chaque PSU seront utilisées pour l'étalonnage en fonction de la non-réponse des foyers.
- La décision de recrutement finale ainsi que les informations sur le sexe, l'âge et le statut tabagique (fumeur actuel ou non) à partir des listes des foyers **pour tous les résidents sélectionnés** de tous les foyers participants. Des informations complémentaires disponibles à partir de la sélection de la PSU, telles que l'environnement rural/urbain et la région peuvent également être utilisées pour une correction de la non-réponse. Ces informations complémentaires sur l'échantillon sélectionné seront utilisées pour produire les taux de réponse selon certaines variables de la classe de pondération sélectionnée. Il est essentiel qu'aucune des variables utilisées pour les calculs de taux de réponse ne soit manquante. Dans le cas où

le statut tabagique serait manquant ou inconnu, le membre de la liste doit se voir attribuer le statut de « non-fumeur » en ajustement de la pondération pour la non-réponse au niveau individuel.

Calcul des corrections de la non-réponse de l'exemple :

Le biais de non réponse des estimations des caractéristiques simples de la population, comme les moyennes, les totaux et les proportions est basé uniquement sur les données des répondants et la pondération de base est en partie déterminée par la covariance au niveau des membres de la population ($\sigma_{\pi y}$) entre la propension (à savoir, la probabilité dans un sens stochastique) du j -ème membre à répondre (π_j) et les mesures d'enquête (y_j) pour ce qui est estimé à partir des données de l'enquête (Lessler et Kalsbeek, 1992)². Notez que nous considérons provisoirement tous les individus comme des membres de la population au sens large et, donc, nous abandonnons l'étiquette du ménage (i) dont l'individu est membre.

Pour des populations (cibles) de taille N (supposée connue), nous pouvons écrire le biais dû à la non-

réponse d'une estimation non corrigée ($\hat{y}_r = \frac{\hat{t}_r}{N} = \frac{\sum_{j=1}^N s_j r_j y_j / p_j}{v}$) de la moyenne de la population

($\bar{Y} = t / N$) comme

$$Bias(\hat{y}_r) = \frac{\sigma_{\pi y}}{\bar{\pi}}, \quad (17)$$

où, pour le j -ème membre de la population, s_j est l'indicateur 0/1 pour l'événement S, p_j est la probabilité de sélection globale (à savoir, $Pr\{s_j = 1\}$), r_j est l'indicateur 0/1 pour l'événement R et $\bar{\pi}$ est la moyenne de toutes les propensions à répondre (π_j) dans la population. Ajuster la pondération de base d'un répondant à l'enquête pour la non-réponse nécessite une estimation de la propension à répondre du répondant. L'inverse de (c.-à-d., 1 divisé par) cette propension estimée devient la correction pour la non-réponse, qui est multipliée par la pondération de base de l'étape précédente pour produire une pondération ajustée en fonction de la non-réponse du répondant.

Plus précisément, ajuster les pondérations d'échantillon pour la non-réponse exige que chaque valeur de π_j soit estimée empiriquement sur la base de l'expérience de la non-réponse dans l'échantillon. La correction multiplicative au niveau individuel pour la non-réponse est alors simplement l'inverse de la propension de réponse estimée du répondant :

$$A_j^{(nr)} = \hat{\pi}_j^{-1} \quad (18)$$

et la pondération corrigée pour la non-réponse est

² Le terme « propension » plutôt que « probabilité » est utilisé dans le cadre de la non-réponse à l'enquête, puisque le comportement stochastique présumé, plutôt que la randomisation explicite, détermine le résultat du processus déterminant si oui ou non un membre d'un échantillon de l'enquête répond.

$$W_j^{(2)} = B_j A_j^{(nr)} \quad (19)$$

Une question clé dans la production de cette correction est alors de savoir comment estimer les propensions à répondre individuellement. Les taux de pondération de classe de réponse et de propension anticipée à répondre à partir d'un modèle de logistique ajusté sont deux approches qui ont été utilisées pour corriger la non-réponse unitaire. Puisque la méthode de classe de pondération est plus simple à mettre en œuvre et l'approche de propension anticipée n'a généralement pas été jugée supérieure dans sa capacité à contrôler (**pas à éliminer**) le biais de non-réponse, l'approche recommandée pour le GATS est la correction de la classe de pondération.

Estimer les propensions est un peu plus compliqué dans la GATS, puisque la non-réponse peut se produire à la fois au niveau du foyer et au niveau individuel de l'échantillonnage pour le recrutement des répondants. La propension de réponse combinée pour le *j*-ème répondant (π_j) a donc deux composantes multiplicatives qui doivent être estimées séparément, l'une pour intégrer la propension du foyer du répondant à répondre en complétant la liste du foyer ($\pi_j^{(HH)}$) et l'autre afin de refléter la propension du répondant à répondre en effectuant l'entretien du GATS, une fois que le répondant a été choisi sur la liste du foyer ($\pi_j^{(person)}$).

Selon la méthode recommandée pour corriger la non-réponse, chaque composante de propension est estimée pour le *j*-ème répondant comme le taux de réponse pour les membres du sous-groupe (c'est-à-dire la « classe ») des membres de l'échantillon sélectionné avec des caractéristiques et des tendances de réponse similaires à celle du répondant. Le calcul du taux de « réponse » pour les composantes au niveau du foyer et individuel doit suivre les directives pour les versions pondérées ou non pondérées des taux de réponse RR1 telles que définies par l'AAPOR (Association américaine de recherche sur l'opinion publique, 2009) et présentées dans le **Manuel de conception de l'échantillon du GATS**. Les taux de réponse pondérés à l'aide des pondérations de base sont préférables à des taux de réponse non pondérés. En définitive, il est du ressort des pays de décider lesquels utiliser. En utilisant les codes de décision des foyers, le taux de réponse des foyers calculé séparément pour chaque segment de l'échantillon (voir description ci-dessous) est calculé comme

$$\text{Taux de réponse au niveau des ménages} = \frac{[1]}{[1] + [3] + [4] + [5] + [6] + [9]} \quad (20)$$

où

- 1 = Questionnaire Ménage rempli, une personne sélectionnée
- 2 = Questionnaire Ménage rempli, personne n'a été sélectionné
- 3 = Questionnaire Ménage partiellement rempli, liste du ménage interrompue (entretien incomplet)
- 4 = *Questionnaire Ménage* incomplet, le répondant approprié n'a pas pu être identifié à la présélection
- 5 = Personne au domicile
- 6 = Refus du ménage
- 9 = Autre non-réponse du ménage.

Notez que le code 2 de décision finale du foyer est exclu à la fois du numérateur et du dénominateur du taux de réponse des ménages, puisque ces foyers sont considérés comme n'étant pas admissibles, que la randomisation des sexes ait été utilisée ou non.

Le taux de réponse au niveau des ménages rempli expliqué ci-dessus exclut les ménages inadmissibles du dénominateur et suppose que tous les foyers sélectionnés dont l'admissibilité est inconnue (codes de décision finale 3, 4, 5, 6, 9) sont admissibles à participer à la GATS. Cela peut conduire à sous-estimer les taux de réponse au niveau des ménages, si les foyers dont l'admissibilité est inconnue sont vraiment non admissibles. Il est recommandé aux pays d'estimer cette proportion (e) en calculant le taux d'admissibilité connu, correspondant au nombre de personnes admissibles connues (code de décision 1), divisé par le nombre de personnes admissibles connues plus les personnes non admissibles connues (code de décision 2) :

$$e^{household-level} = \frac{[1]}{[1] + [2]} \quad (21)$$

Si cette proportion pour l'ensemble de l'échantillon est inférieure à 0,90, les pays doivent ajuster la composante inconnue du taux de réponse du foyer en multipliant les inconnues (code de décision finale 3, 4, 5, 6, 9) par ce rapport (e). La formule suivante est conforme au taux de réponse RR3 (taux de réponse 3) de l'AAPOR (2009) :

$$\text{Household-level response rate} = \frac{[1]}{[1] + e^{household-level} [3 + 4 + 5 + 6 + 9]} \quad (22)$$

La correction correspondant à la classe de pondération au niveau des foyers ($A_{hi}^{(nr, HH)}$) se calcule en divisant un par le taux de réponse des ménages pondéré pour chaque segment de l'échantillon.

Le taux de réponse au niveau individuel est calculé dans des sous-groupes formés de façon stratégique (voir description ci-dessous) :

$$\text{Person-Level RR} = \frac{[11]}{[11] + [12] + [14] + [15] + [16] + [17]} \quad (23)$$

où

- 11 = Questionnaire individuel *terminé*
- 12 = Entretien incomplet
- 13 = L'individu sélectionné s'est avéré inadmissible à la GATS ultérieurement
- 14 = Le répondant sélectionné n'est pas chez lui
- 15 = Refus du répondant sélectionné
- 16 = Répondant choisi incompetent
- 17 = Autre non-réponse individuelle.

Le taux de réponse individuelle ci-dessus exclut du dénominateur les individus non admissibles (code de décision finale 13) et présume que toutes les personnes sélectionnées dont l'admissibilité est inconnue (code de décision finale 14) sont admissibles à la GATS. Cela peut conduire à sous-estimer les taux de réponse individuels si les enquêteurs choisissent souvent des répondants qui se trouvent être non admissibles à l'enquête une fois que l'entretien a commencé. Il est recommandé aux pays d'estimer la proportion de ces répondants choisis dans la liste qui sont vraiment admissibles à répondre à l'enquête de la GATS (e) en utilisant les fréquences de codes de décision pondérées comme :

$$e^{person-level} = \frac{[11] + [12] + [15] + [16] + [17]}{[11] + [12] + [13] + [15] + [16] + [17]} \quad (24)$$

Si cette proportion pour l'ensemble de l'échantillon est inférieure à 0,90, les pays doivent ajuster la composante inconnue du taux de réponse individuel en multipliant les inconnues (code de décision finale 14) par ce rapport (e). La formule suivante est conforme au taux de réponse RR3 (taux de réponse 3) de l'AAPOR (2009) :

$$\text{Person-level response rate} = \frac{[11]}{[11] + [12] + [e^{person-level} [14]] + [15] + [16] + [17]} \quad (25)$$

La composante de correction au niveau individuel ($A_{ij}^{(nr, person)}$) de l'ajustement de la classe de pondération pour la j -ème personne dans le i -ème foyer se calcule en divisant un par le taux de réponse pondéré au niveau individuel.

Les corrections basées sur de petites tailles de sous-groupe ou de grappe peuvent souffrir de variations considérables et entraîner des valeurs de correction trop grandes. Par conséquent, une limite supérieure de 3,00 sera fixée pour toutes les corrections calculées au niveau des foyers et individuel des classes de pondération. Les valeurs supérieures à 3,00 dans les deux composantes de la pondération de l'ajustement seront plafonnées à 3,00.

RECOMMANDATION :
toutes les composantes de correction au niveau des foyers ou individuel qui dépassent 3,00 doivent être fixées à 3,00.

Enfin, les composantes correspondant au niveau des foyers ($A_{hi}^{(nr, HH)}$) et la composante au niveau individuel ($A_{ij}^{(nr, person)}$) de la correction pour la j -ème personne dans l' i -ème foyer doivent être multipliées pour produire la correction combinée de non-réponse pour ce répondant ($A_j(nr)$).

Le choix des caractéristiques à utiliser dans la définition des sous-groupes (à savoir, les « classes de pondération ») pour chaque composante est stratégiquement important, puisque la réduction du biais associé à cette approche est directement liée à l'ampleur de la corrélation entre le taux de réponse et le paramètre d'intérêt pour ces classes (Kalton, 1983). Les classes de pondération dans la plupart des situations sont définies par la classification croisée de plusieurs variables catégorielles (par exemple, ou des variables continues qui ont été classées, telles que l'âge en années individuelles étant divisé en plusieurs catégories définies par groupes d'âge). Ces sous-groupes peuvent être des strates d'échantillonnage, des grappes de l'échantillon ou définis par d'autres informations connues pour tous les membres sélectionnés de l'échantillon. Pour les échantillons de la GATS, nous recommandons que les classes de pondération pour la composante du foyer de la correction pour la non-réponse soient définies par l'ensemble des foyers sélectionnés dans les « segments » de l'échantillon. Dans les pays qui ont choisi les grappes géographiques pour effectuer la liste des foyers en une seule étape, les PSU de l'échantillon seront les classes de pondération pour la composante foyer. Pour la composante au niveau individuel de la correction pour la non-réponse, nous recommandons que les classes de pondération soient définies séparément pour chaque région qui nécessite 8 000 (4 000 selon le cas) répondants suivant les estimations de la GATS. En outre, la région suivant une combinaison ou l'ensemble des variables des classes de pondération suivantes au niveau individuel doit être utilisée pour la correction : zone urbaine/rurale et l'âge (15 à 24, 25 à 34, 35 à 44, 45 à 54, plus de 55 ans), le sexe (masculin, féminin) et le statut tabagique actuel (fumeur et non-fumeur) consignés dans la liste. Dans le cas où aucune estimation régionale n'est nécessaire, les classes de pondération pour les composantes au niveau individuel peuvent être formées par une combinaison de variables zone urbaine/rurale, sexe, âge et statut tabagique.

RECOMMANDATION :
vous devez disposer de formulaires de cellules de pondération de classe pour la composante au niveau individuel de l'ajustement pour la non-réponse par sexe, l'âge et le statut tabagique actuel enregistrés sur la liste, ainsi que suivant la région, si des estimations régionales de qualité sont requises.

Une fois que les classes de pondération ont été formées pour chaque composante, les deux contributions multiplicatives à la correction de la classe de pondération pour la GATS seront calculées comme suit dans le cas de \mathcal{R} . Les numérateurs et les dénominateurs dans les exemples ci-dessous sont les sommes des pondérations de base. Pour la composante du foyer, si l'ensemble des $H_{\alpha} = 56$ foyers sélectionnés dans la α -th PSU où se trouve le foyer de \mathcal{R} est admissible pour l'étude et la somme des pondérations de base parmi ces foyers est 170 013,3, et $r_{\alpha}^{(HH)} = 50$ d'entre eux acceptent de participer à l'étude (total pondéré parmi eux = 156 261,9), la composante du foyer de la correction pondérée pour la non-réponse sera alors calculée à partir de l'expérience de réponse pondérée de l'échantillon des foyers sélectionnés dans la PSU ou le segment où se trouve le foyer de \mathcal{R} . Pour cet exemple, supposons que nous ayons constaté que e *au niveau du foyer* a dépassé 90 % nous permettant ainsi d'utiliser la formule 20 ci-dessus. Le taux de réponse pondéré au niveau du foyer de la classe de pondération pour \mathcal{R} (résidant dans l' i -ème foyer) est

$$RR_h^{(pondéré,HH)} = \hat{\pi}_{hi}^{(wca,HH)} = \frac{\sum_{i=1}^{r_{\alpha}^{(HH)}} B_{hi}}{\sum_{i=1}^{H_{\alpha}} B_{hi}} = \frac{156\,261,9}{170\,013,3} = 0,9191 \quad (26)$$

La composante d'ajustement des ménages de la classe de pondération pour \mathcal{R} (résidant dans l' i -ème ménage) est alors calculée comme suit

$$A_{hi}^{(nr,HH)} = \frac{1}{[\hat{\pi}_{hi}^{(wca,HH)}]} = \frac{1}{\sum_{i=1}^{r_{\infty}} B_{hi} / \sum_{i=1}^{H_{\infty}} B_{hi}} = \frac{1}{0,9191} = 1,088 \quad (27)$$

où les B_{hi} sont les pondérations de base variables pour les ménages dans la h -ème classe de pondération.

Si moins de 90 % des ménages fournissant des informations d'admissibilité ont été jugés admissibles à participer à la GATS dans tout le pays, la composante inconnue de l'admissibilité HH sera multipliée par (e) donnant des taux de réponse légèrement plus élevés et des corrections HH inférieures en conséquence.

La composante d'ajustement des ménages (1,088) est inférieure à 3,00, par conséquent, il n'est pas nécessaire de limiter la valeur de la correction.

Le taux de réponse pondéré individuel doit être calculé en utilisant (23) ci-dessus, si la proportion (e) des personnes sélectionnées admissibles à remplir le questionnaire de la GATS est de 0,90 ou plus. Si (e) est inférieur à 0,90, la formule (25) doit être utilisée pour calculer le taux de réponse au niveau individuel. Comme avec la composante de correction des ménages, la composante de correction au niveau individuel pour \mathcal{R} est un divisé par le taux de réponse pondéré pour la classe de pondération de \mathcal{R} . Supposons maintenant que \mathcal{R} est une femme, vivant en milieu urbain, fumeuse dans la région X, et qu'elle fait partie des $r_{\mathcal{R}} = 680$ répondants (nombre pondéré = 119 009 025) parmi les $n_{\mathcal{R}} = 771$ résidents des foyers (valeur pondérée = 134 935 233) dans la classe de pondération au niveau individuel composée de tous les résidents des foyers sélectionnés de la région X qui sont des fumeurs actuels, vivant en milieu urbain, de sexe féminin. Supposons, en outre, que le taux de réponse pondéré pour la classe de pondération de \mathcal{R} est

$$RR_h^{(pondéré,individuel)} = \hat{\pi}_{hj}^{(wca,individuel)} = \sum_{j=1}^{680} B_{hj} / \sum_{j=1}^{771} B_{hj} = \frac{119\,009\,025}{134\,935\,233} = 0,8820 \quad (28)$$

La composante d'ajustement au niveau individuel pour \mathcal{R} est donc calculée comme suit

$$A_{hj}^{(nr,individuel)} = \frac{1}{[\hat{\pi}_{hj}^{(wca,individuel)}]} = \frac{1}{\sum_{j=1}^{680} B_{hj} / \sum_{j=1}^{771} B_{hj}} = \frac{1}{0,8820} = 1,1338 \quad (29)$$

La composante d'ajustement au niveau individuel est également inférieure à 3,00, par conséquent, aucune réduction n'est nécessaire pour plafonner la valeur. La correction finale pour la non-réponse de \mathcal{R}

$$A_j^{(nr)} = \frac{1}{\pi_j^{(wca)}} = \frac{1}{[\hat{\pi}_i^{(wca,HH)}][\hat{\pi}_j^{(wca,individuel)}]} = [A_i^{(nr,HH)}][A_j^{(nr,individuel)}] = [1,0880].[1,1338] = 1,2336 \quad (30)$$

Suivant la valeur de la pondération de base et la correction finale de non-réponse pour \mathcal{R} , la pondération de l'échantillon corrigée pour la non-réponse pour \mathcal{R} est calculée comme suit

$$W_j^{(nr)} = [B_j].[A_j^{(nr)}] = [12501,3081].[1,2336] = 15421,61 \quad (31)$$

3.3 Ajustement de l'étalonnage post-stratification

Nécessaire pour effectuer cette étape de calcul :

- Le dénombrement de la fréquence de la population, à partir de chiffres du recensement effectué dans les 5 ans de l'entretien ou de toute autre source de données fiables de la population actuelle, des personnes âgées de 15 ans ou plus conjointement en fonction des variables catégorielles liées au tabagisme et au déséquilibre restant de l'échantillon (par exemple, la couverture). Les variables d'étalonnage possibles incluent le sexe, l'éducation, l'âge, le milieu urbain/rural et la région, si le dénombrement de la population régionale est disponible et la taille des échantillons des répondants est assez grande au niveau régional (8 000 recommandés) pour produire des estimations régionales de qualité.
- Des questions formulées de manière comparable et des catégories de réponse aux questions sur le sexe du répondant et les questions d'éducation dans le questionnaire de la GATS.
- Les pondérations ajustées pour la non-réponse ($W_j^{(nr)}$) pour tous les répondants de l'échantillon

Calculer les corrections post-stratification avec l'exemple :

Alors que les deux types de correction pour la non-réponse que nous venons de décrire sont efficaces pour compenser le déséquilibre de l'échantillon dû à des variables utilisées pour définir des classes de pondération et prévoir les propensions à répondre, il peut y avoir d'autres caractéristiques importantes de l'échantillon pour lesquelles aucune correction n'a été apportée. Par exemple, il peut y avoir des caractéristiques supplémentaires de l'échantillon choisi pour lesquelles des taux de réponse différentiels se sont produits. Il peut également y avoir des caractéristiques de la population avec des taux de couverture différentiels de la base, et il peut y avoir des variations de la taille de l'échantillon sélectionné sur des caractéristiques autres que celles en fonction desquelles le processus de sélection de l'échantillon a été stratifié. Une solution fréquente pour faire face à ce déséquilibre qui subsiste est de continuer à étalonner l'échantillon, mais, cette fois, en fonction de la population à partir de laquelle l'échantillon a été sélectionné.

Deville et Sarndal (1992) ont été les premiers à utiliser le terme « étalonnage » conjointement avec la pondération d'échantillon, mais des approches qui, de fait, contraignent le comportement des pondérations existent depuis plus de 60 ans. En principe, l'objectif de la correction de la pondération de l'étalonnage est de rendre les sommes pondérées des données de l'échantillon conformes avec les dénombrements correspondants dans la population cible. La post-stratification et le ratissage étaient les premières formes importantes de l'étalonnage de la pondération, et il a été démontré qu'ils peuvent être des applications spéciales du cadre d'étalonnage généralisé discuté par Deville et Sarndal (1992). Tous deux sont encore utilisés fréquemment aujourd'hui.

Le rôle de l'étalonnage dépend de la ou des corrections, le cas échéant, effectuées, et l'ordre dans lequel elles sont faites. Par exemple, lorsque l'ordre des trois uniques corrections est non-réponse et ensuite étalonnage/ajustement de l'étalonnage ajuste les déséquilibres de l'échantillon qui n'ont pas été spécifiquement traités par la non-réponse. D'autre part, s'il s'avère que seule une correction d'étalonnage est pratiquée, elle devient la seule intégration de toutes les sources de déséquilibre de l'échantillon.

L'ensemble final des pondérations peut être étalonné pour la répartition de la population sur la base des données démographiques d'une source externe statistiquement supérieure (par exemple, le recensement le plus récent ou les résultats d'une autre enquête nationale contemporaine, dont les estimations de la taille de la population, sont de qualité égale ou supérieure). Les projections démographiques fiables et généralement acceptées peuvent également être utilisées comme l'objet de l'étalonnage. Dans le cas où le recensement le plus récent a eu lieu cinq ans ou plus avant la date du recueil des données de la GATS, d'autres sources de données d'étalonnage doivent être envisagées. Les pays ne disposant pas de sources existantes, ou périmées, de données d'étalonnage peuvent ne pas être en mesure de réaliser cette étape de calibration. Les points focaux nationaux et les statisticiens de la SRU doivent être consultés si cela est le cas. Cette étape consiste essentiellement à corriger l'échantillon pondéré (basé sur $W_j^{(nr)}$ à partir de l'échantillon de la GATS) en fonction de la répartition de la population d'un ensemble de variables d'étalonnage catégorielles de l'une des deux façons suivantes : (1) par le biais de la *post-stratification* (ou de la pondération des cellules) en fonction de la répartition de la population conjointe ou croisée de ces variables, ou (2) par le biais de *ratissage* (ou d'ajustement proportionnel itératif) de telle sorte que les marges des répartitions de la population conjointe de ces variables correspondent à celles de la population. Même si la variation des pondérations corrigées finales sera probablement quelque peu inférieure avec le ratissage, la taille prévue relativement grande des échantillons de la GATS se prêtera plus facilement à la post-stratification, qui étalonne plus précisément l'échantillon par rapport à la population, et qui donc est la méthode recommandée pour l'étalonnage de la GATS. Des instructions détaillées pour l'utilisation de la post-stratification sont présentées ci-dessous. Les statisticiens du pays qui souhaiteraient utiliser les procédures de ratissage doivent prendre contact avec les points focaux nationaux et le comité d'examen de l'échantillon (SRC) avant de poursuivre.

Comme avec d'autres corrections, l'étalonnage est plus efficace lorsque les variables utilisées pour définir les répartitions de contrôle sont étroitement corrélées avec les variables clés de l'étude. Bien que le meilleur ensemble de facteurs prédictifs varie souvent suivant les variables d'étude dans les enquêtes de santé, le sexe et l'éducation sont généralement de bons facteurs prédictifs du comportement tabagique et sont donc un bon choix pour les échantillons de la GATS. L'âge, la résidence en milieu rural ou urbain, et la région, si les échantillons sont de tailles suffisantes pour des estimations régionales, sont également des variables potentielles d'étalonnage. La résidence en milieu rural ou urbain doit remplacer l'éducation dans la définition des cellules d'ajustement de l'étalonnage à chaque fois que des comparaisons du milieu urbain par rapport au milieu rural sont considérées plus importantes que les avantages de l'étalonnage par rapport à l'éducation comme facteur prédictif de l'usage du tabac. En définitive, la dernière pondération d'analyse (W_j) pour le *j-ème* enregistrement des données du répondant est obtenue à partir de l'équation (1), $W_j = B_j A_j^{(nr)} A_j^{(cov)} A_j^{(cal)}$, où $A_j^{(cal)}$ est

RECOMMANDATION :

Vous devez avoir à votre disposition un formulaire de cellules d'ajustement pour la post-stratification en fonction du sexe, de l'âge et de l'éducation indiqués par le répondant ainsi qu'en fonction de la région, si des estimations régionales de qualité sont nécessaires. La résidence en milieu rural ou urbain doit remplacer l'éducation dans la définition des cellules d'ajustement de l'étalonnage à chaque fois que des comparaisons du milieu urbain par rapport au milieu rural sont considérées plus importantes que les avantages de l'étalonnage par rapport à l'éducation comme facteur prédictif de l'usage du tabac.

calculé suivant une stratégie d'étalonnage. Notez que $A_j^{(cov)} = I$ puisqu'aucune correction n'est recommandée spécifiquement pour la couverture de la base.

Une post-stratification pour étalonner les pondérations finales de la GATS doit être appliquée de la manière suivante. Tout d'abord, les cellules d'ajustement doivent être définies suivant la classification croisée de quelques variables d'étalonnage catégoriques (ou classées), généralement connues pour être corrélées aux mesures clés de l'usage du tabac qui seront présentées à partir des échantillons de la GATS. Comme indiqué précédemment, les variables prédictives que nous recommandons (au minimum) pour l'étalonnage de l'échantillon de la GATS sont le sexe du répondant (masculin ou féminin) et quatre catégories du niveau d'éducation officiel du répondant. Les catégories d'éducation doivent être définies de façon à ce que le pourcentage de répartition marginale entre les catégories soit aussi proche de l'uniforme que possible (soit environ 25 % de la population dans chaque groupe basé sur le recensement ou d'autres sources externes statistiquement supérieures).

Si l'on revient à notre exemple de calcul des pondérations pour le répondant fictif de la GATS, \mathcal{R} , sera assigné à la cellule d'ajustement de l'étalonnage comprenant les répondants de la région X de sexe féminin et ayant le même niveau d'éducation que \mathcal{R} . Si le dénombrement de la population des individus présentant ces caractéristiques, basé sur le dernier recensement, s'avère être $N_h = 2,724,182$, et que la somme pondérée de l'échantillon avec ces caractéristiques est, $\sum_{j=1}^{r_h} W_{hj}^{(nc)} = 2,919,669.3442$, la correction post-stratification ($A_j^{(cal)}$) est alors calculée pour \mathcal{R} et pour tous les autres répondants de la GATS dans la cellule d'ajustement de l'étalonnage de \mathcal{R} comme

$$A_j^{(cal)} = \frac{N_h}{\sum_{j=1}^{r_h} W_{hj}^{(nc)}} = \frac{2\,724\,182}{2\,919\,669,3442} = 0,9330, \quad (32)$$

où N_h est le dénombrement de la population selon la source externe de données de dénombrement de

la population, de façon à ce que $N = \sum_{h=1}^H N_h$ soit la taille de la population totale, comme indiquée par la

source d'étalonnage externe, et r_h est la taille de l'échantillon des répondants de la cellule d'ajustement.

Notez que $A_j^{(cal)}$ est le même pour tous les répondants dans la cellule d'ajustement « h ». Lors de la vérification de l'ensemble des corrections ($A_j^{(cal)}$) de toutes les cellules d'ajustement à la fin de cette étape, il faut observer que la plupart sont soit légèrement supérieures ou inférieures à 1, avec les segments de la population qui étaient encore sous-représentés dans l'échantillon après la correction pour la non-réponse présentant des valeurs supérieures à un, et ceux qui étaient surreprésentés des valeurs inférieures à un.

En multipliant la pondération existante de \mathcal{R} par la correction pour la cellule de correction de \mathcal{R} ($A_j^{(cal)}$) la pondération finale corrigée de \mathcal{R} (qui est ajoutée à l'enregistrement des données de \mathcal{R} sur le fichier d'analyse) sera

$$W_j = W_j^{(nr)} \cdot A_j^{(cal)} = [1542,61] \cdot [0,9339] = 14402,24 \quad (33)$$

et la répartition de fréquence relative pondérée de l'échantillon utilisant cette pondération finale correspond désormais à la répartition de la population correspondante à l'égard des variables d'étalonnage, c'est-à-dire

$$\sum_{j=1}^{r_h} W_{hj} / \sum_{h=1}^H \sum_{j=1}^{r_h} W_{hj} = N_h / N . \quad (34)$$

Cette égalité peut être utilisée comme un contrôle des calculs de cette correction.

Méthode alternative :

Dans les pays avec de plus petites tailles d'échantillon de répondants et un échantillonnage disproportionné par région plus grand, le ratissage peut être une alternative plus pratique pour la post-stratification. Ratisser pour forcer les sommes de pondération finales corrigées à correspondre aux totaux catégoriques marginaux pour un ensemble de variables d'étalonnage (mais pas pour les cellules de classement croisé) a d'abord été proposé par Deming et Stephan (1940). En commençant par le même type de configuration multidirectionnelle de variables d'étalonnage catégoriques comme post-stratification, chaque itération du ratissage implique d'amener séparément les sommes de catégorie pour chaque variable aux totaux catégoriques correspondants de la population (voir Kalton et Flores-Cervantes, 2003 pour un exemple de calcul pour ce cas où il y a deux variables d'étalonnage). Les pays souhaitant utiliser le ratissage plutôt que la post-stratification pour l'étalonnage des pondérations finales de la GATS doivent discuter de cette alternative avec le point central de chaque pays et le comité d'examen de l'échantillon (SRC).

4. Assurer la qualité des pondérations de l'échantillon du GATS

Afin d'assurer que les pondérations de l'échantillon pour la GATS sont produites de manière uniforme et exacte dans chaque pays, nous offrons plusieurs suggestions sur les façons d'obtenir cette assurance en concluant ce manuel.

Documentation minutieuse de la sélection de l'échantillon et des probabilités associées

Une bonne documentation du processus de sélection suivi dans chaque étape de l'échantillonnage du processus de sélection de l'échantillon du GATS est cruciale. Il est donc extrêmement important que le personnel technique dans chaque pays de la GATS s'engage de manière ferme à tenir cette documentation. Il y a plusieurs façons de s'assurer que cette documentation sera complète et pleinement utile au calcul des pondérations. Premièrement, une fois que l'algorithme de sélection pour chaque étape est fixé, le personnel doit consulter un texte d'échantillonnage reconnu pour trouver la formule de calcul correspondant à la stratégie de sélection utilisée. Certaines stratégies de sélection couramment utilisées pour les échantillons d'enquête sont l'échantillonnage aléatoire simple, systématique, un au hasard, PPS systématique, PPS sans remplacement et PPS avec remplacement. Les textes qui peuvent servir de références sont Hansen, Hurwitz et Madow (1953), Kish (1965), Raj (1968), Cochran (1977), Sukhatme, et al., (1984) et Lohr (1999). Deuxièmement, les moyens de calculer ces probabilités de sélection doivent être en place pour faire des calculs corrects pendant le déroulement de la sélection. En pratique, cela signifie que la sélection informatisée est préférable à une sélection manuelle et que le codage informatique pour faire la sélection de l'échantillon doit également inclure des mesures pour calculer et stocker les probabilités de sélection au même moment que les membres individuels de l'échantillon sont choisis. Cela signifie également que si certaines étapes d'échantillonnage doivent être effectuées manuellement, les formulaires utilisés pour la sélection de l'échantillon doivent prévoir un moyen d'assurer que toutes les informations nécessaires pour calculer la probabilité de sélection pour cette étape peuvent être produites à ce moment-là, ou plus tard si cela est préférable. Troisièmement, un membre du personnel technique du pays doit être investi de la responsabilité globale de s'assurer que les probabilités de sélection sont calculées correctement, dans les délais, et correctement stockées pour être utilisées lorsque vient le temps de produire les pondérations. Quatrièmement, les procédures et les formules utilisées pour calculer toutes les composantes de la pondération de l'échantillonnage doivent être planifiées et documentées, avec suffisamment de détails à l'avance du processus de calcul des pondérations pour qu'un autre membre compétent de l'équipe technique ne travaillant pas sur la GATS puisse effectuer correctement le processus de calcul des pondérations. Cette série de spécifications de planification spécifique du pays pour le calcul des pondérations de la GATS doit être modifiée à l'issue du processus de calcul des pondérations pour servir de documentation finale sur les pondérations de la GATS. Et, enfin, la documentation sur le processus de pondération doit être, en général, mise à la disposition des partenaires de GATS et du comité d'échantillonnage de la GATS.

Contrôler l'effet statistique négatif des pondérations variables

Kish (1965) a montré qu'une variation de la pondération de l'échantillon peut augmenter la variance des estimations de l'enquête par un facteur à effet multiplicatif (*Meff*). Par exemple, si $\bar{w}=12394.0091$ est la moyenne de l'échantillon des pondérations finales corrigées, W_j , et $s_w^2=(5463.3760)^2$ est la variance de ces mêmes pondérations des membres de l'échantillon de la GATS dont le répondant fictif, \mathcal{R} , fait partie, alors

$$Meff_w = 1 + \frac{s_w^2}{\bar{w}^2} = 1 + \left(\frac{5463.3760}{12394.0091} \right)^2 = 1.1943 \quad (35)$$

Certains des facteurs contribuant à cette variation sont les caractéristiques de la conception de l'échantillon visant à réduire la variance des estimations (par exemple, répartition de l'échantillon parmi les strates), tandis que d'autres sont le résultat de stratégies pour contrôler les biais de ces estimations (par exemple, les corrections pour non-réponse et l'étalonnage de l'échantillon). Contrôler cette variation relève d'un fragile équilibre entre la réduction de l'effet négatif de $Meff_w$ sur la précision des estimations tout en minimisant les changements des avantages de la réduction du biais de la correction de la pondération.

L'approche recommandée pour traiter les effets de pondération variable pour les échantillons de la GATS sera de surveiller $Meff_w$ pour la pondération finale corrigée (W_j), et de consulter

le personnel central de la GATS si $Meff_w > 2.00$. Cette consultation peut conduire à l'utilisation de diverses stratégies existantes (par exemple, la troncature des pondérations) pour restreindre la taille des corrections de pondération ou à l'utilisation de différentes stratégies de réduction des

pondérations. Par exemple, Kalton et Flores-Cervantes (2003) décrivent le processus de réduction des cellules d'ajustement pour limiter la taille des corrections produite par la dernière série de cellules. Cette approche préventive peut être utilisée pour contrôler la taille des corrections pour la non-réponse, mais elle peut également être utilisée pour restreindre la taille des corrections d'étalonnage comme juste milieu entre la post-stratification et le ratissage (voir Deville et Sarndal, 1992). Potter (1988, 1990, 1993), d'autre part, examine plusieurs méthodes de réduction qui créent de multiples ensembles de pondérations réduites basées sur divers niveaux de réduction, puis choisissent l'ensemble de pondérations réduites ayant l'erreur quadratique moyenne estimée la plus basse pour les estimations clés de l'enquête. Un inconvénient est que la pondération réduite produisant l'erreur quadratique moyenne la plus basse pour une estimation de l'enquête peut ne pas produire le même résultat pour une autre estimation.

RECOMMANDATION :

Contrôlez l'effet de la pondération variable en surveillant son effet multiplicateur sur la variance des estimations de l'enquête ; consultez le personnel de la GATS si nécessaire

Suivre attentivement les procédures de calcul recommandées et les contrôles de qualité dans ce manuel

Le but de ce manuel est de fournir des recommandations pour élaborer une stratégie de calcul des pondérations de chaque échantillon de la GATS et pour assurer que la stratégie peut être réalisée avec succès de sorte que les analystes des données de la GATS puissent être sûrs que les pondérations qu'ils utilisent contribueront à la production des meilleures estimations possibles de la population sur le tabagisme. Pour vous aider à atteindre cet objectif, nous présentons un ensemble de procédures spécifiques pouvant produire des pondérations de l'échantillon pour les données de la GATS qui répondent aux normes les plus strictes pour la recherche d'enquête et qui, si elles sont suivies de façon uniforme dans tous les pays de la GATS, contribueront à la qualité des comparaisons des résultats entre les pays. En décrivant les étapes de calcul des pondérations, nous avons également suggéré ou fait allusion à des moyens d'assurer la qualité des pondérations de l'échantillon. À savoir :

- a. Demandez à quelqu'un, qui n'est pas impliqué dans le calcul des pondérations de la GATS, mais connaît le contenu du protocole recommandé tel que défini dans ce manuel, de vérifier le travail de ceux qui produisent les pondérations pendant le processus de calcul. Il peut être préférable d'effectuer cette vérification après chaque étape (à savoir, le calcul de la pondération de base, de la correction à la non-réponse et de la correction de l'étalonnage), plutôt qu'une fois à l'issue de tout le processus.
- b. Déterminez si la taille moyenne de la pondération de base (B_j), divisée par la taille moyenne de la pondération ajustée pour la non-réponse ($W_j^{(nr)}$) est à peu près égale au taux global de réponse définitif RR1 tel que décrit dans la **section 3.2** de ce manuel.
- c. Utilisez l'équation (34) pour confirmer que la distribution pondérée utilisant les pondérations finales étalonnées (W_j) correspond (pour chaque cellule d'ajustement d'étalonnage) à la répartition de la population par rapport à laquelle les pondérations ont été étalonnées.
- d. Notez que la somme des pondérations finales corrigées (W_j) couvrant l'ensemble de l'échantillon doit être de la taille totale du dénombrement de la population (âgée de 15 ans ou plus) de la source utilisée pour l'étalonnage. Si cette source était le dernier recensement, cette somme de pondérations doit alors être le dénombrement de la population à partir du dernier recensement.
- e. Confirmez que les corrections post-stratification ($A_j^{(cal)}$) produites par l'ensemble des cellules d'ajustement sont, pour la plupart, légèrement supérieures ou inférieures à 1.

Préparer un document écrit détaillé décrivant le processus de calcul des pondérations qui est effectivement utilisé

Comme avec la plupart des activités d'enquête majeures, planifiées puis mises en œuvre, le plan qui est élaboré (c'est-à-dire essentiellement le contenu de ce manuel si le processus de calcul des pondérations recommandé est suivi) diffère souvent quelque peu de ce qui est réellement fait. Des écarts de spécifications de planification peuvent avoir lieu en produisant la pondération à chacune des trois étapes de calcul des pondérations. Par exemple, si une autre méthode reconnue pour la sélection du répondant de l'enquête au sein du foyer est utilisée, le calcul de la probabilité conditionnelle de sélection du répondant au sein d'un foyer participant à la GATS sera alors calculé conformément à la méthode de remplacement. Pour éviter la perte de détails sur cette importante activité de pré-analyse de la GATS, une documentation écrite rigoureuse du processus de pondération réel suivant le même schéma de base que ce manuel doit être effectuée immédiatement après le calcul des pondérations et doit alors devenir un élément permanent de la documentation pour toutes les activités d'enquête de la GATS. Ce document sera utilisé pour mener une évaluation formelle du processus de pondération et d'étalonnage par des biostatisticiens SRU et, finalement, sera inclus dans le rapport final de chaque pays comme enregistrement permanent des méthodes utilisées au cours de la GATS. Pour consulter plus de détails sur le format et les calculs spécifiques nécessaires, référez-vous au Manuel de **l'assurance qualité du GATS : directives et documentation**. La publication des résultats pondérés à partir des données du pays ne doit pas avoir lieu jusqu'à ce que cette documentation soit examinée et officiellement certifiée comme ayant satisfait aux normes de qualité des pondérations fixées par ce manuel.

RECOMMANDATION :

Suivez les étapes recommandées pour la pondération et effectuez les contrôles suggérés au cours du calcul

Consulter le point focal du GATS lorsque des questions ou des problèmes surviennent pendant le processus de pondération

Enfin, le point focal du GATS est prêt à soutenir le processus de calcul de la pondération de l'échantillon dans chaque pays de la GATS. Ce soutien comprendra la clarification des modalités procédurales énoncées dans ce manuel ainsi que l'assistance au dépannage pour tout processus imprévu dans la réalisation de cette importante série de tâches d'enquête.

RECOMMANDATION :

Préparez un document écrit qui donne des détails sur la manière dont les pondérations ont été calculées pour chaque pays du GATS

5. Bibliographie

- American Association for Public Opinion Research. 2009. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 6th edition. AAPOR.
- Berry, C. C., Flatt, S. W., and Pierce, J. P. (1996). Correcting Unit Nonresponse via Response Modeling and Raking in the California Tobacco Survey. *Journal of Official Statistics*, 12(4), 349-363.
- Cochran, W.G. (1977). *Sampling Techniques*, 3rd Ed., New York: Wiley and Sons.
- Deming, W. E. and Stephan, F. F. (1940). On the Least Squares Adjustment of a Sample Frequency Table When the Expected Marginal Totals are Known. *Annals of Mathematical Statistics*, 11(4), 427-444.
- Deville, J. C. and Sarndal, C. E. (1992). Calibration Estimators in Survey Sampling. *Journal of American Statistics Association*, 87, 376-382.
- 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.
- 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): Quality Assurance: Guidelines and Documentation, Version 2.0*. Atlanta, GA: Centers for Disease Control and Prevention, 2010.
- Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953). *Sample Survey Methods and Theory*, Vols. I and II, 1953.
- Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663-685.
- Iannacchione, V. G., Milne, J. G., and Folsom, R. E. (1991). Response Probability Weight Adjustments Using Logistic Regression. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 637-642.
- Kalsbeek, W. D., Morris, C., and Vaughn, B. (2001). Effects of Nonresponse on the Mean Squared Error of Estimates from a Longitudinal Study. *Proceedings of the Section on Survey Research Methods, American Statistical Association*
- Kalton, G. (1983). *Compensating for Missing Survey Data*. Ann Arbor, MI: University of Michigan.
- Kalton, G. and Flores-Cervantes, I. (2003). Weighting Methods. *Journal of Official Statistics*, 19(2), 81-97.
- Kish, L. (1965). *Survey Sampling*, Second Printing, New York: Wiley and Sons.
- Lepkowski, J. M., Kalton, G., and Kasprzyk, D. (1989). Weighting Adjustments for Partial Nonresponse in the 1984 SIPP Panel. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 296-301.
- Lessler, J. T. and Kalsbeek, W. D. (1992). *Nonsampling Error in Surveys*. New York: Wiley and Sons.
- Lohr, S. (1999). *Sampling: Design and Analysis*, Pacific Grove: Duxbury Press.
- Potter, F. J. (1988). Survey of Procedures to Control Extreme Sampling Weights. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 453-458.
- Potter, F. J. (1990). A study of procedures to identify and trim extreme sampling weights. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 225-230.
- Potter, F. J. (1993). The Effect of Weight Trimming on Nonlinear Survey Estimates. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 758-763.
- Raj, D. (1968). *Sampling Theory*, New York: McGraw-Hill.
- Särndal, C.-E., Swensson, B., and Wretman, J. H. a. (1992). *Model assisted survey sampling*. New York: Springer-Verlag.
- Sukhatme, P. V. et al, (1984). *Sampling Theory of Surveys with Applications*, Ames: Iowa State University Press.



GLOBAL TOBACCO SURVEILLANCE SYSTEM (GTSS)

