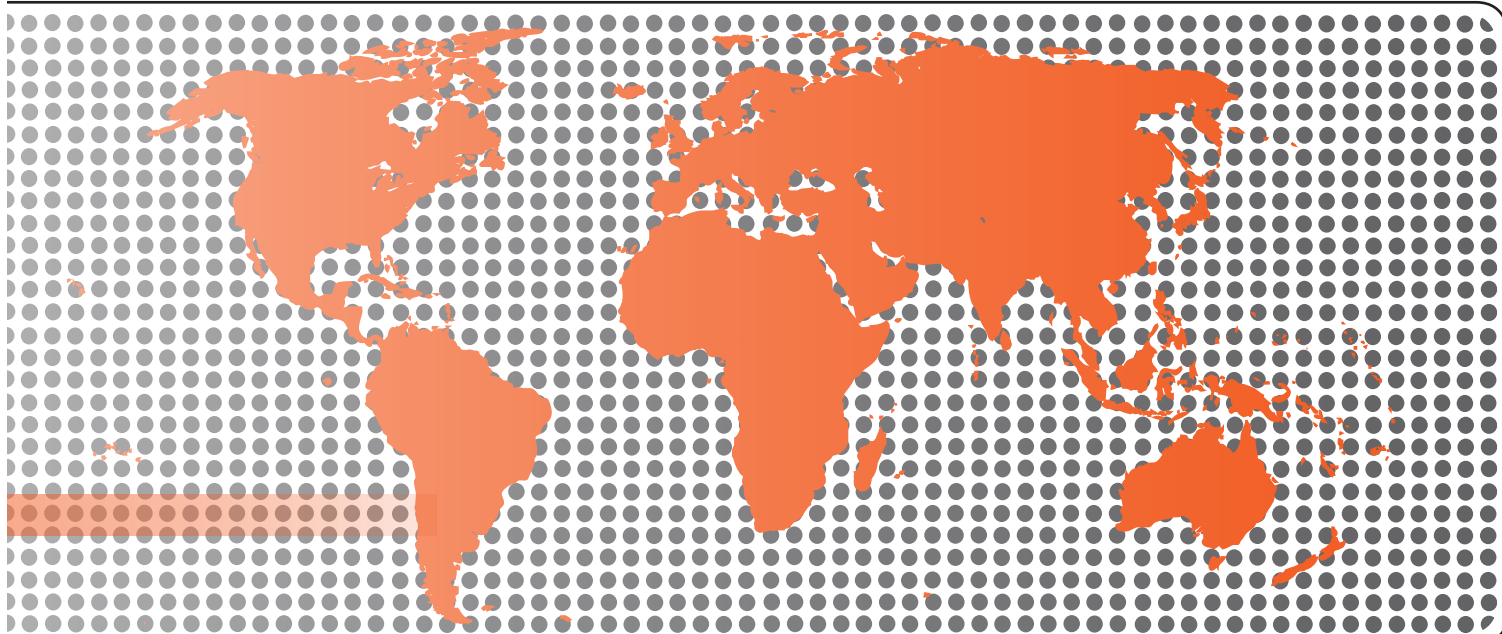


GTSS Global Adult Tobacco Survey (GATS)



Assurance qualité GATS : Directives et documentation



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

**Assurance qualité :
Directives et documentation**

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'échantillonnage
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 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 d'une enquête GATS

Modèle de Fiche d'information
Rapport national : Plan et lignes directrices relatives aux tabulations
Définitions des indicateurs

Publication et diffusion des données 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): Quality Assurance: Guidelines and Documentation, 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
- RTI International (Research Triangle Institute)
- 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 du manuel	1-2
1.3 Assurance qualité de la GATS	1-3
2. Plan du processus d'assurance qualité	2-1
3. Assurance qualité : Période Pré-collecte des données	3-1
3.1 Conception du questionnaire et Processus d'examen	3-1
3.2 Programmation et contrôle de version	3-2
3.3 Finalisation du questionnaire	3-5
3.4 Transfert et gestion des données	3-6
3.5 Conception de l'échantillon	3-6
3.6 Préparation du fichier maître de sélection de l'échantillon	3-7
4. Assurance qualité : Collecte et Gestion des données	4-1
4.1 Collecte des données sur le terrain : Installation et maintenance	4-1
4.2 Gestion des données aux niveaux du PSU(UPE)/régional/de État	4-7
4.3 Gestion des données au Centre national de données (NDC)	4-8
5. Assurance qualité : Période suivant le recueil des données	5-1
5.1 Nettoyage et préparation des données pour les calculs de pondération de l'échantillon.....	5-2
5.2 Mesures de la qualité : Échantillonnage, erreur d'échantillonnage, et pondération	5-6
5.3 Mesures de qualité : Couverture, non-réponse, et autres erreurs n'ayant pas rapport à l'échantillonnage	5-12
5.4 Examen officiel de la qualité statistique	5-17
5.5 Création du fichier des données analytiques	5-20
6. Données et exigences de documentation	6-1
7. Bibliographie	7-1
Annexe A : Glossaire	A-1
Annexe B : Période avant le recueil des données	B-1
B.1 Adaptation du Questionnaire du GATS et processus d'examen	B-1
B.2 Processus de programmation du questionnaire du GATS.....	B-2
B.3 Processus de chargement (création de la structure) du programme GSS sur l'ordinateur de poche.....	B-2
B.4 Variables standard pour le fichier maître de sélection de l'échantillon et le fichier de cas de l'iPAQ	B-4

Annexe C : Collecte et gestion des données	C-1
C.1 Fiche d'information des variables d'indicateurs du GATS.....	C-1
Annexe D : Période Post-Collecte de données	D-1
D.1 Nettoyage et préparation des données pour les calculs de pondération de l'échantillon utilisant une méthode de fusion manuelle	D-1
D.2 Codes de décision finale et calculs des taux de réponse	D-4
D.3 Modèle de correction de l'étalonnage de la pondération poste-stratification entre les cellules d'ajustement.....	D-8
D.4 Effet de la variable de pondération de l'échantillon sur la précision des estimations de l'enquête	D-10
D.5 Effet global de la conception sur la précision des estimations de l'enquête et sur l'homogénéité intragroupe dans les PSUs (UPEs) des estimations clés correspondantes de l'enquête.....	D-11
D.6 Marge d'erreur pour les estimations clés de l'enquête.....	D-13
D.7 Estimations des erreurs d'échantillonnage	D-16
D.8 Taux de couverture de la base des ménages.....	D-23

Documents annexes

Numéro	Page
Document 2-1. Organigramme de l'assurance qualité	2-1

1. Introduction

Le tabagisme est la principale cause 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 8 millions par an d'ici 2030. Si la tendance actuelle continue, 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 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 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 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 recommandé 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, non institutionnalisés, 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.

La GATS utilise une méthodologie d'échantillonnage en grappes à 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, les unités primaires d'échantillonnage sont

L'entretien de la 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 Mathers, C.D., and Loncar D. Projections of Global Mortality and Burden of Disease from 2002 to 2030. PLoS Medicine, 2006, 3(11):e442.

divisées en segments, et les segments en ménages. Ensuite, un échantillon aléatoire de ménages est sélectionné pour participer à la GATS.

L'entretien de la GATS se compose de deux parties : le *Questionnaire Ménage* et le *questionnaire individuel*. Le *questionnaire Ménage* (présélection du foyer) et le *questionnaire individuel* (entretien individuel) seront effectués en utilisant un dispositif électronique de collecte 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 foyer sélectionné répond aux exigences d'éligibilité du GATS, et de faire une liste, ou un tableau de tous les membres éligibles 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, et les connaissances, les attitudes et les perceptions du tabac.

1.2 Utilisation du manuel

Ce manuel fournit des recommandations sur l'évaluation et l'assurance de la qualité des données recueillies au cours du GATS. C'est un complément des manuels existants du GATS, par conséquent il ne comprendra pas les descriptions des procédures documentées ailleurs. Au lieu de cela, il décrit les éléments du processus d'assurance qualité qui doivent être réalisés au cours du GATS. Le respect de ces directives d'assurance qualité est extrêmement important pour le succès de cette enquête. Le manuel est destiné à être utilisé par tous les participants à la mise en œuvre de la GATS et au processus d'assurance qualité, y compris le personnel informatique et les statisticiens responsables de l'échantillonnage et de la pondération. Ce manuel contient les définitions des termes en italique dans le Glossaire des termes (**Annexe A**).

Dans ce document, le processus d'assurance qualité du GATS est divisé en trois étapes chronologiques : la période pré collecte de données, la collecte et la gestion des données, et la période post- le collecte des données. Les activités d'assurance qualité varient suivant les étapes du processus d'assurance qualité, et des directives sont fournies pour chaque étape. Le **Chapitre 2** commence par fournir un plan du processus décrivant la portée globale des activités d'assurance qualité.

Chapitre 3 : la période Pré- Collecte des données comprend des sections décrivant les activités d'assurance qualité qui doivent être réalisées au cours de la conception et de l'examen du questionnaire spécifique au pays, la programmation du questionnaire sur le PDA de poche [portatif], utilisé pendant la collecte de données, la conception de l'échantillon, et la création du fichier maître de sélection de l'échantillon.

Chapitre 4 : Collecte et gestion des données comportent trois sections. La première section traite des activités d'assurance qualité effectuées sur le terrain, y compris l'installation /configuration et la maintenance du système de gestion des cas et du questionnaire. Elle comprend également des sous-sections relatives à la gestion de données et définit les normes pour les rapports de routine au cours des travaux sur le terrain. La deuxième section décrit la gestion des données et l'assurance qualité pratiquées après les entretiens et le téléchargement des données au niveau de l'État ou régional, selon le cas.

La troisième section fait référence à l'agrégation et aux questions de contrôle de la qualité au niveau national. Cette section décrit également la préparation du fichier de données brutes pour le nettoyage.

Chapitre 5 : la période post-collecte des données décrit les activités d'assurance qualité à effectuer une fois que le travail sur le terrain est terminé et que le fichier de données brutes a été préparé pour le nettoyage. La première section décrit les procédures de nettoyage et de préparation des données pour les calculs de pondération de l'échantillon. La deuxième section décrit l'évaluation de la qualité de l'échantillonnage, l'erreur d'échantillonnage, et la pondération de l'échantillon tandis que la troisième section porte sur l'évaluation de la qualité de la couverture de la base de sondage, la non-réponse, et d'autres erreurs non liées à l'échantillonnage. Enfin, la quatrième section décrit la création du fichier final des données analytiques.

Ces chapitres ont pour but de fournir des directives d'assurance qualité et des recommandations de bonnes pratiques. Les méthodologies et/ou la documentation détaillée pour chaque section des chapitres sont fournies dans les annexes, selon le cas.

1.3 Assurance qualité du GATS

L'assurance qualité est un processus consistant en des activités systématiques visant à assurer, évaluer et confirmer la qualité des données recueillies lors d'une enquête (Biemer et Lyberg, 2003). Dans le passé, la qualité des données était synonyme d'exactitude des données, mais le terme a évolué vers un concept plus général. Les données de haute qualité réfèrent simplement à des données qui conviennent à l'usage prévu. Cela comprend non seulement l'exactitude, mais aussi des attributs tels que la ponctualité, l'accessibilité et la comparabilité, faisant de la qualité un concept multidimensionnel. Un ensemble de données est exact dans la mesure où il est exempt d'erreurs. Il respecte les délais s'il est disponible au moment où il est nécessaire. L'accessibilité est déterminée par la relative facilité ou difficulté à l'utiliser. Des données de qualité sont comparables si elles sont les mêmes d'une unité à une autre, que cette comparaison d'unité soit entre des individus, des enquêteurs, des UPEs (PSUs) ou des frontières internationales. La normalisation accroît la comparabilité des données d'enquête, par exemple, l'utilisation de questions normalisées garantit que différents enquêteurs poseront tous les mêmes questions de la même manière. De même, la normalisation des processus d'assurance qualité assure une résolution uniforme des erreurs de données.

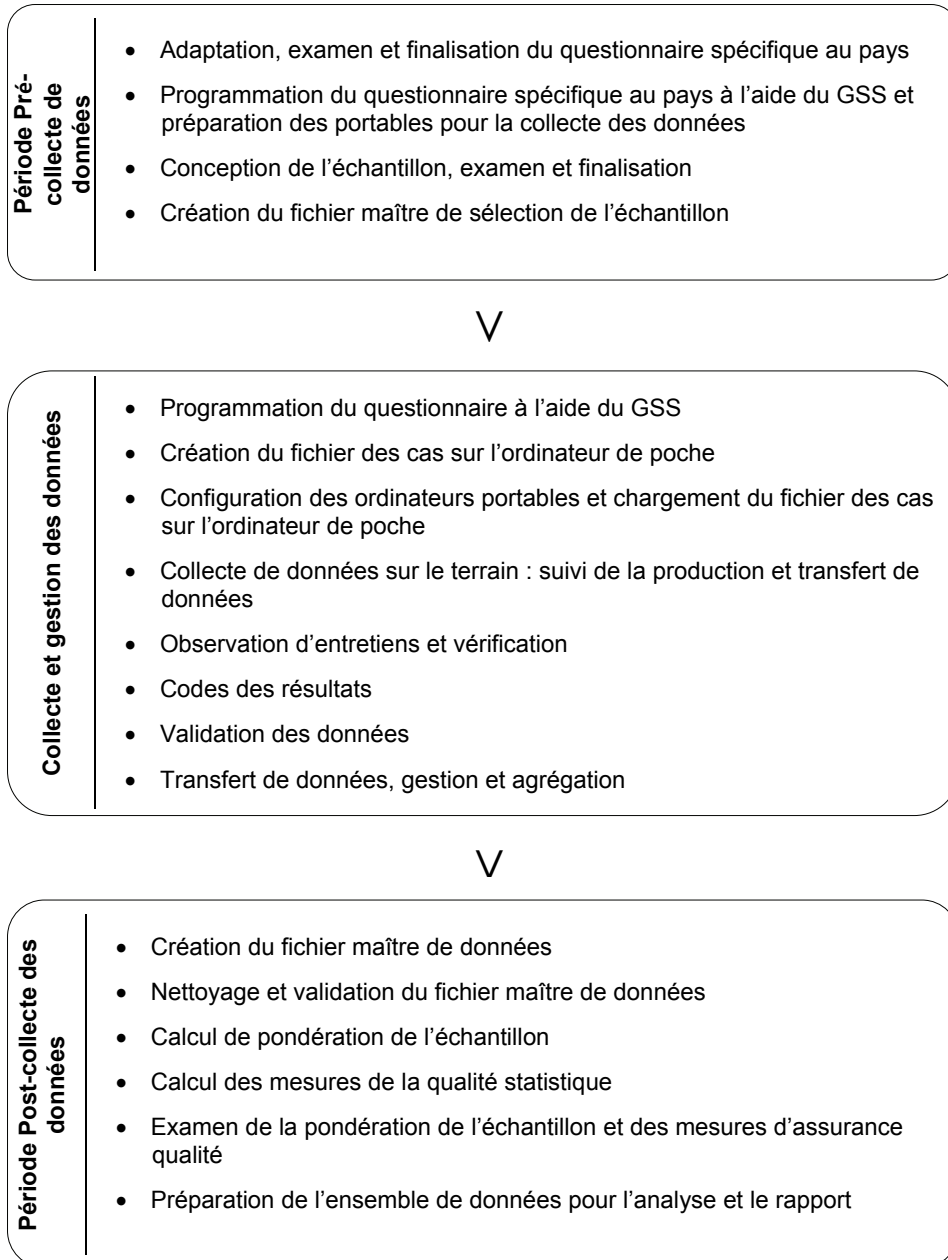
Les multiples dimensions de la qualité se limitent souvent les unes les autres, par conséquent l'optimisation de la qualité est une question de compromis et d'équilibre entre les différentes dimensions. Par exemple, des données de haute qualité doivent être exactes, mais doivent également respecter les délais. S'il faut beaucoup de temps pour assurer un degré d'exactitude très élevé, les données peuvent devenir inapplicables.

Les lignes directrices décrites dans ce document représentent des procédures normalisées pour l'assurance qualité des données recueillies par la mise en œuvre du GATS. Les pays sont encouragés à intégrer des activités supplémentaires d'assurance qualité, en plus de celles décrites dans le document.

2. Plan du processus d'assurance qualité

Le document ci-dessous présente un diagramme des activités d'assurance qualité en matière de calendrier général et la succession des étapes ne se chevauchant pas. Ces étapes sont organisées en trois étapes : période pré-collecte des données, Collecte des données, et période post-collecte des données.

Document 2-1. Organigramme de l'assurance qualité



3. Assurance qualité : Période Pré-Collecte des données

La phase précédant la collecte des données comprend une série de tâches qui doivent être accomplies afin de préparer la collecte de données de l'enquête complète. Ces tâches comprennent la conception du questionnaire, la programmation, la réalisation du prétest, la conception de l'échantillon, et la préparation du fichier maître de sélection de l'échantillon. Ce chapitre décrit le mécanisme d'assurance qualité et les directives standard qui doivent être suivies dans ces différents domaines. Ce chapitre aborde spécifiquement :

- 3.1 Conception du questionnaire et Processus d'examen
 - 3.1.1 Directives d'adaptation du questionnaire
- 3.2 Programmation et contrôle de version
 - 3.2.1 Processus de programmation
 - 3.2.2 Vérification du questionnaire
 - 3.2.3 Chargement de l'IPAQ (Développement)
- 3.3 Finalisation du Questionnaire
 - 3.3.1 Administrer du pré-test
 - 3.3.2 Finalisation du questionnaire
- 3.4 Transfert et gestion des données
- 3.5 Conception de l'échantillon
- 3.6 Préparation du fichier maître de sélection de l'échantillon

3.1 Conception du questionnaire et Processus d'examen

Le GATS maintient un processus normalisé suivant lequel les pays participants conçoivent leurs propres questionnaires. Le Comité d'examen du questionnaire (QRC) [CEQ] du GATS — un groupe composé d'experts dans la lutte antitabac et la conception du questionnaire, provenant de pays développés et en développement — examine et approuve tous les questionnaires du GATS. Le QRC travaille en étroite collaboration avec chaque pays pour adapter le questionnaire du GATS à la situation de chaque pays tout en maintenant les questions principales normalisées de la GATS (voir le **Questionnaire principal du GATS comprenant des questions facultatives** pour plus de détails) pour assurer la comparabilité entre les pays. Les détails spécifiques de l'adaptation du questionnaire et du processus d'examen figurent dans l'**annexe B.1**.

Le comité d'examen du questionnaire (QRC/CEQ) du GATS examine et approuve le questionnaire afin d'assurer la qualité et la comparabilité entre les pays.

3.1.1 Directives d'adaptation du questionnaire

Le QRC [CEQ] recommande au pays de suivre des directives précises en matière d'adaptation du **questionnaire principal du GATS** afin de faciliter le processus d'examen et assurer les normes de qualité :

- Mettez en surbrillance toutes les adaptations au pays apportées au **Questionnaire Principal du GATS** (pour faciliter la référence).

Cela comprend les listes de questions, catégories de réponse, questions facultatives, et questions ajoutées par le pays.

- Barrez pour indiquer les questions principales que le pays veut supprimer (pour faciliter la référence).
- Conservez la numérotation et l'ordre des questions principales et facultatives pour maintenir la cohérence qui rend possible la comparaison entre pays. (Ce n'est pas toujours possible ou optimal, mais cela doit être suivi autant que possible).
- Pour les questions nouvellement ajoutées spécifiques au pays, utilisez le lettrage double selon la section (par exemple, AA10, BB17, EE4). Cela ne perturbera pas la numérotation actuelle en ajoutant les nouvelles questions du pays.
- Changez seulement les instructions d'enchaînement si nécessaire pour accommoder les questions facultatives ajoutées ou les questions ajoutées par le pays.
- Ne révisez pas les questions fondamentales normalisées (sauf pour les catégories spécifiques au pays) afin de maintenir la cohérence pour permettre la comparaison entre pays. (Une justification doit être fournie au QRC pour toutes les exceptions).
- Limitez le nombre de questions supplémentaires pour que la longueur du questionnaire reste raisonnable.
- Limitez la complexité des questions supplémentaires afin de faciliter la programmation de chaque questionnaire du pays
- Fournissez une justification pour les adaptations. L'inclusion de justifications facilite le processus d'examen et d'approbation des questionnaires de chaque pays par le QRC.

3.2 Programmation et contrôle de version

Le GATS utilise un mode de collecte de données par entretien personnel assisté par ordinateur (CAPI) à l'aide d'un PDA de poche comme instrument électronique. Les appareils de poche sont équipés du logiciel General Survey System (GSS)¹. Il y a toujours des caractéristiques uniques à chaque pays comme chaque pays adapte ses questionnaires, a différentes langues avec des polices particulières, et utilise peut-être du matériel différent (par exemple, différents modèles d'appareils portables peuvent être utilisés).

Compte tenu de la complexité, un certain nombre d'étapes communes et de bonnes pratiques doivent être suivies.

Une séquence particulière d'étapes (décrit dans l'**annexe B.2**) nécessite un apport de coopération énergique de la part du pays d'accueil bien à l'avance (de préférence 6 à 8 semaines avant le début de la formation pour un pré-test). Ces étapes exigent que le personnel du pays d'accueil, tel que le personnel informatique et les programmeurs, le personnel de l'enquête, et les responsables de l'enquête soient

Des processus standard et des vérifications doivent être suivis pour veiller à ce que les questionnaires du GATS soient programmés de façon rigoureuse et efficace.

¹ Le logiciel General Survey System (GSS) est conçu pour fonctionner sur une plate-forme Windows Mobile, et a été testé et utilisé sur les ordinateurs de poche iPAQ®, de type PDA, de Hewlett-Packard (HP). Faire fonctionner le GSS sur un autre type de PDA de poche exigerait des modifications du logiciel. (L'utilisation de « iPAQ » est uniquement à des fins d'identification et n'implique pas l'appropriation des organisations participant au GATS).

identifiés et disponibles pour travailler, soit pendant l'atelier d'orientation ou peu après. Ces personnes seront formées en personne ou par le biais de webinars pour démarrer les travaux préparatoires, et elles sont tenues de consacrer un nombre important d'heures de travail pour organiser le prétest du GATS et la réalisation de l'enquête complète.

Les paragraphes ci-dessous présentent des questions spécifiques portant sur le questionnaire et le processus de préparation de l'enquête sur lesquelles l'accent doit être mis pour accroître la qualité.

3.2.1 Processus de programmation

Les processus suivants doivent être suivis afin de maintenir la qualité et l'efficacité lors des préparations du GATS :

- Le personnel de l'enquête/de terrain/les points focaux de mise en œuvre de l'enquête, non pas le personnel informatique, sont chargés de vérifier l'exactitude du langage spécifique au pays (traductions) en fonction du questionnaire du pays. Le personnel informatique travaillera avec le personnel d'enquête/de terrain/les points focaux pour la mise en œuvre pour insérer le texte en utilisant le concepteur de questionnaire GSS IDE.
- Le point focal de l'agence de mise en application doit approuver le *Questionnaire Ménage /QM*) et le *Questionnaire Individuel (QI)* de l'iPAQ (par écrit) avant la formation du personnel de terrain (avec le numéro de version).
- Modifications du texte : une modification du texte doit se refléter à travers toutes les langues.
- La modification de la logique de programmation (plages, contrôles de validité, etc.) d'un questionnaire approuvé nécessite un nouvel examen par le QRC.
- Changer le contenu ou la formulation d'une question ou des réponses nécessite un nouvel examen par le QRC.
- Contrôle de version des fichiers MDB : Le pays d'accueil est propriétaire des fichiers MDB et le personnel informatique/de l'enquête du pays est responsable du maintien d'une piste de vérification des modifications apportées aux questionnaires et des spécifications du programme GSS.
- Après le bilan du pré-test et le démarrage de la mise en œuvre de l'enquête complète, le pays doit envoyer au Centre de coordination des données (DCC) du GATS, le CMSDB, les fichiers SDF Survey0 et Survey1 et les fichiers MDB. Chaque fois qu'il y a une modification des fichiers MDB il faut envoyer des fichiers MDB mis à jour au DCC.

3.2.2 Vérification du questionnaire

Plusieurs étapes d'assurance qualité permettent de vérifier le questionnaire et doivent être suivies et documentées :

- Les programmeurs doivent mettre à jour la version du questionnaire à chaque fois qu'il est modifié et qu'un nouveau fichier SDF est créé. En outre, cela doit être confirmé avant de commencer à examiner le questionnaire. Les archives des anciens fichiers MDB doivent être conservées par le logiciel GSS IDE à titre de documentation et de sauvegarde.
- Pour chaque langue, l'ensemble des spécifications de programmation du GSS est examiné et chaque question sur l'ordinateur de poche doit se conformer exactement au texte de la

question, des réponses et ainsi qu'aux caractères spéciaux du document, **GATS Questionnaire Programming Specifications (Spécifications de programmation du questionnaire du GATS)**. Si des problèmes sont détectés, ils sont clairement documentés dans les spécifications, après une itération, les problèmes sont résolus, puis la vérification complète d'itérations est effectuée. Ce processus itératif continue jusqu'à ce que chaque question soit approuvée sur l'ordinateur de poche. Il n'y a absolument aucune exception à ce processus. Cela doit être fait avec un expert de l'enquête en langue locale. Cela ne doit pas être relégué à des programmeurs et certainement pas à du personnel ne possédant pas l'expertise linguistique requise. Le point focal ou le représentant du pays doit signer la version définitive après avoir revu les modifications et/ou les mises à jour effectuées. Les spécifications doivent être examinées avant la signature pour vérifier que chaque question a été validée par comparaison avec le questionnaire (voir le **Guide de formation à la gestion des données du GATS, Module 2.1** pour plus de détails).

- Les enchaînements et le texte anglais sont toujours rigoureusement testés avant la visite du site. Le logiciel du questionnaire ne doit pas être modifié pendant la collecte des données sur le terrain, sauf si cela est absolument crucial.
- Pour chaque langue, le compte de symboles *, (,), [,], {, et } doit être le même. Sinon, cela doit être revérifié et corrigé.

3.2.3 Chargement de l'iPAQ (Programmation)

Le chargement du programme GSS sur l'iPAQ doit se conformer à des procédures strictes d'assurance qualité :

- Avant le chargement, tous les ordinateurs de poche doivent être réinitialisés et toutes les cartes Secure Digital (SD) sont formatées pour assurer l'ajout d'un nouveau chargement.
- Chaque iPAQ doit contenir la version sur la carte SD de l'iPAQ.
- Des étiquettes sont imprimées avec des cases à cocher correspondant aux étapes de contrôle de la qualité (pour l'installation le GATS sur les ordinateurs de poche) qui sont apposées sur les ordinateurs de poche. La personne qui effectue le contrôle de la qualité du chargement de l'ordinateur de poche doit confirmer que chaque étape est accomplie conformément aux règles d'assurance qualité.

L'**annexe B.3** fournit des détails supplémentaires, étape par étape, sur le processus de chargement des ordinateurs de poche (se référer au **Module 2.2** du **Guide de formation à la gestion des données de la GATS** pour plus de détails).

3.3 Finalisation du questionnaire

3.3.1 Administration du pré-test

L'administration du pré-test dans le pays, avant la mise en œuvre de l'enquête complète, a six objectifs. Atteindre les objectifs suivants contribuera à assurer la haute qualité de la collecte des données lors de l'enquête complète GATS :

1. Fournir au personnel du pays une assistance à la programmation, au test, et à la maintenance d'une enquête sur ordinateur de poche [appareils portatifs], suivant le besoin ;
2. Examiner le questionnaire du prétest du GATS, y compris la formulation des questions et des réponses, les instructions de sondage [la façon de poser les questions], la programmation du logiciel branchement conditionnel (skip logic) et la longueur et l'heure de l'interview ;
3. Aider le personnel du pays à utiliser le matériel [informatique] de poche et le logiciel, au besoin ;
4. Observer la mise en œuvre du pré-test dans une zone urbaine et une zone rurale, si cela est réalisable et applicable ;
5. Tester le plan de collecte, de transfert, de gestion de données, pendant le prétest qui sera adapté pour la mise en œuvre de l'enquête complète ;
6. Présenter aux partenaires de la GATS un compte rendu sur les points forts et des points faibles, et des recommandations sur les préparatifs et l'administration du pré-test.

Le pré-test du GATS fournit une occasion de tester le questionnaire et les procédures de collecte et de gestion des données afin d'assurer que l'enquête complète sera exécutée en suivant des méthodes de haute qualité.

3.3.2 Finalisation du Questionnaire

Afin de finaliser le questionnaire du GATS pour la mise en œuvre de l'enquête complète, les étapes suivantes doivent être suivies :

- **Compte rendu du pré-test.** Tous les membres du personnel (Enquêteur de terrain, Superviseurs de terrain, personnel du projet) doivent fournir leurs commentaires sur la façon dont le questionnaire s'est déroulé dans son ensemble, ainsi que sur des questions spécifiques de l'enquête qui doivent faire l'objet de révisions.
- **Examen des données du pré-test.** L'analyse des données du pré-test peut fournir des éléments précieux pour effectuer des révisions. Quelques problèmes spécifiques à explorer :
 - Non-réponse : Identifier les questions qui ont des taux plus élevés que prévu de réponses « ne sais pas » ou « refusé de répondre ».
 - Fréquences : Examiner les distributions de fréquences des questions clés pour déterminer si les données semblent appropriées. Si ça sort de l'ordinaire, cela pourrait signaler un problème de conception du questionnaire ou un autre problème (par exemple un problème de formation).

- Plages admissibles : Examiner la pertinence des plages autorisées programmées dans les ordinateurs de poche et adapter pour la mise en œuvre de l'enquête complète au besoin.
- Catégories de réponses : Déterminer si les adaptations du questionnaire spécifiques au pays (par exemple, la modification des listes de questions et les catégories de réponse) étaient appropriées. Par exemple, mesurer les fréquences de marques de cigarettes afin de modifier la liste des marques pour le questionnaire de l'enquête complète.
- **Révision pour la mise en œuvre de l'enquête complète.** La version révisée post prétest doit être soumise à l'examen du QRC, y compris les justifications (basées sur les résultats du prétest) des révisions. Les directives d'assurance qualité décrites dans les **sections 4.1 et 4.2** (au besoin) doivent être suivies pour préparer le questionnaire pour la mise en œuvre de l'enquête complète (y compris la programmation révisée, etc.).

3.4 Transfert et Gestion des données

Le pré-test est une simulation de la mise en œuvre de l'enquête complète pour le processus de gestion des données. Sur la base de l'expérience du pré-test et des recommandations, le transfert de données et le processus de gestion seront finalisés pour la mise en œuvre de l'enquête complète. Par conséquent, chaque étape du processus de gestion des données doit être menée conformément au plan de gestion de données en prenant en compte l'enquête complète. Par exemple :

- Le mécanisme de commentaires dans la communication des rapports doit être mis à l'essai durant le pré-test.
- La méthode de transmission de données telle que le site FTP ou GSM doit être testée au cours du pré-test et les éventuels ajustements nécessaires doivent être faits pour l'enquête complète suivant les leçons tirées du prétest.

3.5 Conception de l'échantillon

Le GATS maintient un processus normalisé suivant lequel les pays participants élaborent leur propre échantillonnage. Le Comité d'examen de l'échantillonnage (SRC) du GATS — un groupe composé d'experts en méthodologie d'échantillonnage provenant de pays développés et de pays en développement — examine et finalise la conception de l'échantillon des pays participants. Les points focaux du SRC et du CDC collaborent étroitement avec chaque pays pour adapter la conception de l'échantillon du GATS, tout en maintenant les normes du GATS afin d'assurer la comparabilité pour l'analyse entre les pays. Les détails spécifiques du plan d'échantillonnage du GATS figurent dans le **Manuel de conception de l'échantillon du GATS**.

Dans le cadre du processus d'échantillonnage du GATS, la cartographie et le listage des adresses doivent être faits afin de créer un cadre approprié d'échantillonnage probabiliste par zone. Pour les procédures complètes et des recommandations sur l'assurance qualité en matière de cartographie et listage des adresses, consultez le **Manuel de cartographie et listage des adresses du GATS**.

3.6 Préparation du fichier maître de sélection de l'échantillon

Une fois que l'échantillon du GATS a été tiré, un fichier maître de sélection de l'échantillon doit être préparé. Le fichier maître de sélection de l'échantillon est un ensemble de données qui contient un numéro de cas (Case ID) pour chaque foyer dans la liste d'échantillonnage ainsi que les informations nécessaires pour calculer la pondération de l'échantillon et analyser les données complexes de l'enquête.

Le fichier maître de sélection de l'échantillon comprend les identifiants de l'échantillon, et les identifiants géographiques. Des identifiants tels que la strate, l'unité primaire d'échantillonnage (PSU) [UPE], unité secondaire d'échantillonnage (SSU/segment) [USE], le type de lieu de résidence (urbain/rural), et l'attribution au foyer du sexe présélectionné, si la randomisation en fonction des sexes est utilisée. Identification géographique telle que l'adresse, y compris les codes pour le pays, la région, la province/l'État, le comté/district et le village/la ville.

Le fichier maître de sélection de l'échantillon comprend un numéro de cas [CaseID] unique, les identifiants de l'échantillon, et les identifiants géographiques du ménage.

Reportez-vous à l'**annexe B.4** pour plus d'informations sur les éléments du fichier maître de sélection de l'échantillon.

Une fois que le fichier maître de sélection de l'échantillon est finalisé, il doit être entré en utilisant l'écran GSS IDE > edit case file (modifier le fichier de cas dans le menu Edition), les identifiants d'échantillon peuvent être supprimés du fichier afin de créer un fichier de cas iPAQ (sous-ensemble du fichier maître de sélection de l'échantillon) qui sera chargé sur les appareils de poche pour la collecte des données, selon le cas (à l'aide du GSS IDE > CMS Grid Designer). Le chapitre suivant fournit des détails supplémentaires sur ce processus.

4. Assurance qualité : Collecte et Gestion des données

Ce chapitre décrit les processus de qualité des données et les procédures de collecte et de gestion de données du GATS. La collecte et la gestion des données englobent l'installation du logiciel GSS du GATS, l'utilisation du logiciel pendant la collecte de données sur le terrain, le transfert des données collectées, l'agrégation et la création de fichiers CSV de réponses transposées et d'un fichier maître de base de données agrégées. Cela comprend également la surveillance, la rédaction et la soumission des rapports et le support technique lors de la collecte des données sur le terrain.

Pour un pays donné, la collecte de données, l'agrégation et l'assurance qualité peuvent être effectués à trois niveaux, 1) sur le terrain, 2) au niveau de la PSU [UPE]/de l'État/régional et 3) au Centre national de données du pays (NDC) [CND]. Ce chapitre décrit les processus de qualité des données et les procédures applicables à chacun des trois niveaux, abordant en particulier :

- 4.1 Collecte des données sur le terrain : Installation et maintenance
 - 4.1.1 Fichier de cas IPAQ et gestion des cas
 - 4.1.2 Appareil électronique de recueil de données IPAQ et logiciel de questionnaire d'enquête
 - 4.1.3 Procédures d'assurance qualité de recueil de données sur le terrain
- 4.2 Gestion des données aux niveaux PSU/régional/État
 - 4.2.1 Agrégation et transfert de données
 - 4.2.2 Contrôle de la qualité et validation
- 4.3 Gestion des données au niveau du NDC [CND] du pays
 - 4.3.1 Agrégation
 - 4.3.2 Contrôle de qualité des données recueillies
 - 4.3.3 Communication et assistance technique
 - 4.3.4 Préparation des fichiers de données brutes

4.1 Collecte des données sur le terrain : Installation et maintenance

Le protocole standard du GATS recommande que certaines procédures systématiques soient établies et suivies par le personnel informatique et de terrain afin de maintenir des données exactes et complètes pendant toute la période de collecte des données. Ces procédures sont décrites ci-dessous.

4.1.1 Fichier de cas IPAQ et gestion des cas

1. Il est recommandé que le dossier iPAQ soit préparé et maintenu au CND (NDC) à l'aide du GSS IDE > CMS Grid Designer, le cas échéant suivant le plan national de mise en œuvre de la gestion des données. (Notez que le fichier de cas de l'iPAQ est un sous-ensemble du fichier maître de sélection de l'échantillon).
2. Pendant la construction du fichier de cas et avant le passage à l'iPAQ ('iPAQ mastering') :
 - Tous les Enquêteurs de terrain (FI) [ET] doivent être identifiés et un iPAQ doit lui être assigné.

Les numéros d'identification de cas (CaseID) assignés aux ET (FI) doit être uniques et spécifiques à chaque ET(FI). Au moment du chargement de l'iPAQ, le même numéro d'identification de cas ne doit pas être attribué à plus d'un ET (IF). Cela permettra d'éviter les doublons, d'attribuer une charge de cas raisonnable au ET (IF), et que les cas soient facilement identifiables.

- Tous les FIIDs (numéros d'identification des ETs) doivent être conformes à la convention du GATS consistant en un identifiant numérique à 6 chiffres.
 - Tous les identifiants de cas doivent être conformes à la convention du GATS consistant en un identifiant numérique à 6 chiffres.
3. Le fichier de cas doit comprendre une assignation pour chaque iPAQ/FI qui collectera les données.

Assurez-vous que tous les ETs (FIs) enregistrés sur le tableur FIID ont une assignation dans le fichier des cas.

Le fichier doit être examiné avant le chargement sur l'iPAQ. Lorsqu'il est chargé sur l'iPAQ, le nombre de cas prévu doit apparaître dans le CMS, sinon :

Vérifiez l'identifiant du ET (FI) assigné dans l'iPAQ et vérifiez l'identifiant du ET (FI) figurant sur le fichier des cas. Il est à noter que pour chaque CaseID, il y aura deux dossiers créés qui seront chargés sur le CMS (HH = 00 et IQ = 01).

4. Chaque CMS iPAQ ne doit contenir que les cas attribués au ET (FI) et sur lesquels il est prévu qu'il travaille. Généralement, un maximum de 300 CaseIDs (numéros d'identification de cas) doit être assigné par ET (FI) et chargé sur l'iPAQ. Si un nombre de cas supérieur au nombre recommandé est chargé, la performance, par exemple le chargement du CMS et la capacité d'identifier facilement un cas au cours du travail sur le terrain, peut s'en trouver affectée.
5. Avant de charger l'iPAQ, effectuez un examen de la fréquence des champs du fichier des cas lorsque vous êtes encore sur l'ordinateur du NDC. Pour les exigences du travail sur le terrain, consultez le guide sur le fichier CMS des cas, ***GATS Programmer's Guide to General Survey System (Guide du programmeur sur le système général de l'enquête GATS)*** pour les exigences du travail sur le terrain. Pour le GATS :
- Le champ CASEID (numéro de cas) doit être vérifié pour les doublons et doit être conforme à la norme du GATS consistant en un identifiant numérique à 6 chiffres.
 - Vérifiez que le champ PROJECTNAME (Nom du Projet) indique le nom du pays pour assurer qu'un champ d'identifiant unique au pays existe dans le fichier des cas de l'iPAQ (en utilisant la fonction de recherche de mise à jour du projectname).
 - Vérifiez que la valeur de la variable CreateDate est la date de création du fichier.
 - Comparez le fichier maître de sélection de l'échantillon avec le fichier des cas de l'iPAQ pour vous assurer de son exhaustivité et de son exactitude. (Notez que le fichier de cas de l'iPAQ est un sous-ensemble du fichier maître de sélection de l'échantillon).

Le **Guide du programmeur du système général de l'enquête** contient des informations détaillées sur la création du fichier des cas de l'iPAQ, les descriptions des champs et le tableur d'inventaire de l'iPAQ du FI.

4.1.2 Appareil électronique de collecte de données IPAQ et Logiciel du Questionnaire de l'enquête

1. **Paramètres de l'Enquêteur de terrain** : Chaque iPAQ doit être étiqueté avec un identifiant FIID unique conforme à la convention du GATS avec la même série de chiffres pour l'identifiant du ET, FIID et un affichage dans le menu CMS Admin. Cette information doit être enregistrée au CND (NDC) dans un tableur avec le nom de l'ET (FI) et le numéro de série de l'iPAQ.

Faites une recherche croisée des paramètres de FIID : vérifiez que l'assignation du tableur, le CMS Admin Name de l'iPAQ, et le champ Admin Name and ID et l'étiquette de l'iPAQ correspondent tous avant la collecte de données.

2. **Versions du logiciel : CMS, Questionnaires Ménage et Individuel** : Les l'iPAQs doivent être chargés avec les versions finales approuvées du questionnaire.

Pour chaque iPAQ, vérifiez que la version prévue du logiciel est chargée et fonctionnelle. Amenez le CMS et les Questionnaires Ménage (HH) et les Questionnaires Individuels (IQ), en lisant les informations sur chaque version sur l'écran de démarrage ou utilisez l'option de notification de la version dans le menu de l'ordinateur de poche.

Dans la mesure du possible, le passage aux iPAQs [le chargement des iPAQs] doit se faire au même moment pour assurer la cohérence dans la version, la maîtrise et le processus QC.

3. **Cas chargés** : Au début de la collecte de données/du travail sur le terrain, seuls les cas figurant dans le fichier des cas de l'iPAQ doivent apparaître dans le CMS. Les cas de formation doivent être effacés avant que le travail de terrain ne commence pour éviter la possibilité de travailler sur un cas de formation et que celui-ci ne soit exporté. Les analystes doivent être prêts à filtrer les cas de formation à partir des données réelles si un FI enfreint cette règle.

Lorsque le fichier des cas est disponible (souvent au moment de la prise de contrôle de l'iPAQ), ouvrez le CMS et examinez la liste complète des CasIDs, effectuez une référence croisée avec la gamme des CasIDs du fichier des cas de l'iPAQ. Si des cas de formation apparaissent dans le CMS, utilisez la fonctionnalité Admin « Erase Training Cases » (effacer les cas de formation). Si des cas destinés à un autre ET (FI) apparaissent dans la grille du CMS et ne sont pas des cas de formation, vérifiez les paramètres de l'identifiant du ET (FI) sur l'iPAQ dans le menu ADMIN du CMS pour vous assurer que l'identifiant du ET (FI) correct a été entré. Si les paramètres de l'identifiant du ET (FI) sont corrects et correspondent à ce qui est prévu, effacez les anciens cas en amenant un nouveau CMSDB.sdf (qui sera vide) dans l'iPAQ et chargez les cas à nouveau.

Le **GATS Programmer's Guide to General Survey System a (GATS Guide du programmeur sur le système général de l'enquête)** contient des instructions détaillées sur les paramètres de l'iPAQ, la création et l'installation des fichiers connexes de l'iPAQ.

4.1.3 Procédures d'assurance qualité de la collecte de données sur le terrain

Cette section donne un aperçu des procédures d'assurance qualité recommandées qui peuvent être appliquées pendant la collecte des données sur le terrain. Le **Manuel de l'Enquêteur de terrain du GATS** et le **Manuel du Superviseur de terrain du GATS** doivent être consultés pour de plus amples informations.

Rôle du Superviseur de Terrain

Les superviseurs de terrain (FSs) [STs] sont le lien essentiel entre les ETs (FIs) et le personnel de gestion de l'enquête GATS. Ils sont chargés de surveiller la production et la performance, et de communiquer toutes les questions de terrain qui peuvent avoir un impact sur l'achèvement du GATS dans les délais impartis. Ils sont responsables de veiller à ce que les données de leur équipe soient collectées conformément au protocole de collecte de données du GATS, et soient de la plus haute qualité. Pour aider les Enquêteurs de terrain à comprendre la valeur que l'équipe du GATS attache à la qualité des données, les FSs doivent faire du contrôle de qualité une partie intégrale de leurs activités hebdomadaires.

La première mesure du FS(ST) pour assurer la qualité des données est de transmettre les attentes de qualité à leurs FIs. Les superviseurs sont tenus de fournir des commentaires continus à l'attention des FI concernant les problèmes de qualité, à la fois positifs et négatifs, et ils souligneront l'importance que l'équipe du GATS attache à la *qualité*.

Surveillance de la production

La production comprend toutes les activités nécessaires pour réussir à atteindre les objectifs de taux de réponse du GATS. Ces objectifs comprennent des activités d'initiation pour chaque foyer attribué, la prise de contact, la localisation, les efforts de conversion des refus, et l'achèvement de ces activités avec succès, en conformité avec toutes les spécifications du GATS. Le FS, les analystes du pays doivent surveiller régulièrement le temps passé à réaliser un entretien. En outre, le suivi de la production lors de la collecte des données est essentiel pour assurer l'achèvement du GATS dans les délais et dans les limites du budget alloué.

Les types de rapports suivants peuvent être générés à l'aide du GSS IDE > menu, « Reporting », et doivent être utilisés par l'analyste du pays pour surveiller les différents aspects de la production et, idéalement, devraient être communiqués aux Enquêteurs de Terrain (FSs) régulièrement. Quand et où ces rapports seront disponibles dépendront du modèle de gestion de données en place dans chaque pays.

- *Statut de cas HH et statut de cas IQ* : ces rapports présentent les codes de résultat les plus récents pour les cas des foyers et individuels par enquêteur. Ce rapport peut être consulté à l'échelon national, au niveau du FS, ou par un FI donné.
- *Rapport d'exceptions* : ce rapport aidera le personnel du GATS à surveiller le nombre de visites effectuées plusieurs fois pour tous les cas, qui ne sont pas terminées, et permettra également de comprendre les raisons d'une visite supplémentaire.
- *Rapport des répartitions des fréquences des IQ* : ce rapport aidera le personnel du GATS à surveiller les réponses au questionnaire par enquêteur ou PSU, en vue d'évaluer la qualité des entretiens d'enquête en étudiant les mesures de qualité telles que :

- les problèmes potentiels avec le routage de l'entretien ou des données « manquantes » ;
- les questions qui ont des taux plus élevés que prévu de réponses « ne sais pas » ou « refusé » ;
- les questions avec des taux plus élevés que prévu de réponses « autres », plutôt que l'un des choix de réponses pré codées ;
- les réponses « autre, précisez » qui ne sont pas claires et ne peuvent pas être facilement codées, ou qui répètent l'une des options pré codées.

Surveillance du transfert de données

La transmission régulière des données des Enquêteurs de Terrain (FIs) au Centre National de Données (NDC) du pays est vitale pour la gestion du GATS, et permet de détecter plus tôt d'éventuels problèmes et de les régler. Les Superviseurs de Terrain (FSs) sont chargés de faire en sorte que chaque Enquêteur/Enquêteuse de terrain transmette les données suivant un calendrier régulier. Si la transmission se fait par cartes SD, le FS doit s'assurer que les FI exportent les données vers la carte SD *quotidiennement*. Dans la situation où la gestion des données se fait à l'aide d'une transmission par lignes téléphoniques fixes ou Internet sans fil, le journal de rapport de transmission peut être utilisé pour s'assurer que les FI transmettent aussi souvent qu'ils le doivent. Si le nombre de fichiers transférés est égal à zéro, cela peut indiquer que le FI ait des difficultés pour transmettre. Si le rapport du journal de transmission indique qu'un enquêteur de terrain n'a pas transmis de données récemment, le FS doit contacter le FI et lui demander de transmettre les données immédiatement, puis lui rappeler l'importance de la transmission régulière et ponctuelle. Les FIs doivent remplir un formulaire de transmission de matériel pour accompagner chaque carte SD envoyée à leur FS pour établir une chaîne de traçabilité. De même, les FSs doivent tenir un journal de tous les fichiers d'exportation recueillis auprès de leurs FIs.

Observations

Observer les enquêteurs de terrain est une partie importante des efforts de contrôle de qualité du Superviseur de Terrain (FS) et peut constituer un moyen d'aider les enquêteurs de terrain à s'améliorer. La participation au travail des enquêteurs de terrain et l'observation de ceux-ci permettront également d'éviter toute tentation de leur part de prendre des raccourcis dans l'administration du questionnaire ou de ne pas suivre d'une autre façon le protocole d'entretien du GATS. Au minimum, les FS sont tenus d'observer chacun de leurs enquêteurs de terrain pendant les premiers jours de la période sur le terrain, puis moins fréquemment, en fonction des directives des agences de mise en application. Sur ces observations, le FS accompagnera l'Enquêteur de terrain dans les ménages échantillonnés pour vérifier qu'il enregistre le résultat de la visite correctement dans le système de gestion des cas du GSS, et qu'il mène la pré-sélection du foyer ainsi que l'entretien ultérieur correctement. Ces premières visites permettront de s'assurer que les erreurs d'administration du questionnaire ou d'utilisation des codes de résultat sont identifiées rapidement, et de déterminer si l'enquêteur de terrain doit recevoir une formation supplémentaire ou des conseils au besoin. En outre, les FS peuvent décider d'accompagner les enquêteurs de terrain lors de visites particulièrement difficiles (par exemple, dans des foyers qui ont refusé précédemment).

Vérification

Un moyen de vérifier la qualité des données recueillies par les enquêteurs est de faire de brefs entretiens de vérification avec les ménages qui ont déjà été présélectionnés et interrogés. Un bref entretien de vérification permettra aux FSs de confirmer que l'Enquêteur de terrain a bien fait ce qui suit :

- Identifié et présélectionné le bon ménage. Occasionnellement, un Enquêteur de terrain peut sélectionner un foyer différent de celui qui a été échantillonné. Si cela arrive, le FS doit demander à l'enquêteur de terrain d'aller à l'adresse correcte du ménage et effectuer la présélection et l'entretien avec ce ménage.
- Enregistré l'âge, le sexe et le statut tabagique corrects des membres du foyer. Si l'âge n'est pas correctement enregistré (par exemple, le FI répertorie les résidents âgés de moins de 15 ans, en plus des résidents âgés de 15 ans ou plus), cette erreur aura un impact sur l'identification des personnes admissibles et pourrait entraîner la sélection par le FI d'un membre du foyer non admissible à l'entretien.
- Administré le *Questionnaire Individuel* au membre du ménage sélectionné.

Le nombre exact des entretiens de vérification sera déterminé en collaboration avec le personnel de gestion de l'agence de mise en application, mais il est généralement convenu que les FS mèneront des entretiens de vérification sélectionnés au hasard sur environ 10 % de chaque mission de l'enquêteur de terrain. Les entretiens de vérification peuvent consister en (1) une vérification brève qui soumet au répondant sélectionné un petit nombre de questions afin de vérifier que le répondant a récemment terminé une enquête sur des sujets liés au tabagisme, et d'évaluer la performance de l'entretien ou (2) si un pays le souhaite, une nouvelle administration complète du *questionnaire des ménages* et, si possible, du *questionnaire individuel*. Étant donné que les répondants peuvent percevoir le processus de nouvel entretien comme un fardeau, vous n'avez pas besoin d'administrer l'ensemble du *questionnaire individuel*. Par ailleurs, les entretiens de vérification peuvent se faire avec un papier et un crayon ou avec un ordinateur de poche chargé avec les foyers à vérifier. L'ordinateur de poche du FI ne doit jamais être utilisé pour mener les entretiens supplémentaires, un ordinateur de poche séparé doit être utilisé.

Plus tard, le FS comparera les réponses obtenues à partir du nouvel entretien avec celles enregistrées par le FI. Si des différences existent et qu'il semble que l'enquêteur de terrain a mené un entretien avec la mauvaise personne dans le ménage, le FS doit renvoyer l'enquêteur de terrain dans le ménage pour interroger l'individu correct, et les réponses doivent ensuite être comparées à nouveau. Dans tous les cas, qu'une erreur ait été découverte au cours du processus de vérification ou non, les informations provenant de l'entretien de vérification doivent être transmises au NDC du pays avec les autres documents.

Attribuer des codes de résultat finaux de non-entretien

Les FS doivent approuver le code de résultat final de non-entretien avant qu'un questionnaire ne soit finalisé. (Le *Manuel du Superviseur de terrain du GATS* fournit des informations supplémentaires sur les essais visant à terminer les cas de non-entretiens en attente). Si un cas ne doit être soumis à aucune autre mesure, le FI doit saisir le code de résultat final approprié de non-entretien dans le système de gestion de cas (voir le *Manuel de l'Enquêteur de terrain du GATS* pour une liste et une description des codes de résultat de terrain).

Lorsque le recueil des données est terminé dans une zone donnée (par ex., PSU, région, État, etc.), tous les cas exploités doivent avoir reçu un code de résultat final. Les *questionnaires des foyers* et *individuels* terminés recevront automatiquement un code de résultat final complété par le programme. Pour les cas des *questionnaires Ménage* et *Individuel* qui n'ont pas abouti à

Lorsque la collecte des données est terminée, tous les cas exploités doivent avoir reçu un code de résultat final.

un entretien complet (par exemple, les refus, personne au domicile), des codes de résultat finaux de non-entretien doivent être attribués par les FIs avec l'approbation de leurs superviseurs.

4.2 Gestion des données aux niveaux du PSU/régional/de l'État

Cette section décrit les procédures de recueil et de gestion des données des pays gérant les données au niveau du terrain. Cela comprend notamment les pays avec un accès à Internet. Le personnel du PSU ou régional peut avoir accès à un ordinateur de bureau/ordinateur portable qui peut être utilisé pour recueillir des données à partir des ordinateurs de poche des FI, permettant l'envoi des fichiers de données exportés au NDC directement. Dans certains cas, les pays peuvent disposer d'un personnel informatique régional qui agrège les fichiers de données exportés recueillis à partir des ordinateurs de poche des FI, puis envoie les données agrégées au NDC.

4.2.1 Agrégation et transfert de données

Lors de la collecte de données sur le terrain, une méthode de collecte de fichiers, de communication et de documentation des fichiers doit être établie.

L'agrégation ne peut être réalisée que sur un ordinateur de bureau/ordinateur portable. Le FS ne doit pas se voir confier la responsabilité de l'agrégation des données à moins qu'il y soit bien formé. Cette tâche doit normalement être effectuée par le personnel informatique qui est formé à l'utilisation du logiciel GSS IDE.

L'examen des données des questionnaires ne peut être effectué que sur un ordinateur de bureau/ordinateur portable. Le système GSS de l'ordinateur de poche ne dispose pas d'un mécanisme d'examen des données des questionnaires. L'ouverture d'un questionnaire pour examiner simplement les données n'est pas recommandée, car cela peut invalider des champs et/ou des données précédemment saisis (faux dans le tableau des réponses). L'examen des données doit être effectué sur un ordinateur de bureau/ordinateur portable après l'agrégation.

Tous les fichiers sdf exportés à partir de l'ordinateur de poche du FI doivent être sauvegardés et archivés dans un dossier C:\GATS\Archive, après leur agrégation.

À la fin de la collecte des données, une étape de réconciliation doit avoir lieu pour s'assurer que tous les fichiers exportés au niveau du ET (FI) ont été pris en compte et agrégés au niveau du CND (NDC).

Pour transférer électroniquement des données, un protocole de transfert sécurisé de fichiers (FTP) est recommandé.

Si une messagerie électronique est utilisée pour transférer des données, la vérification de la réception peut être particulièrement critique et peut ne pas constituer un moyen sécurisé de transfert.

Une fois que le transfert de données est terminé, le destinataire doit être notifié. Le destinataire doit confirmer la réception et enregistrer l'information suivant une méthode de préétablie.

Si l'agrégation a lieu à ce niveau, confirmez que les fichiers ont été inclus dans le processus d'agrégation.

Vérifiez que chaque FI a exporté le ou les fichiers de données pour la période prévue.

Si l'agrégation est effectuée à ce niveau, plusieurs points doivent être vérifiés :

- L'exactitude et l'exhaustivité des données d'agrégation : Tous les fichiers sdf ont-ils été pris en compte et inclus dans le processus ?
- Il y a au moins un sdf pour chaque ET (FI).
- Examinez le rapport de statut d'agrégation (Status.htm) pour confirmer le compte des statuts d'entretien par le ET (FI).

Pour plus d'efficacité, agrégez les fichiers sdf les plus récents d'un Enquêteur de terrain donné au fichier maître actuel contenant les données précédemment agrégées.

4.2.2 Contrôle de qualité et validation

Un examen préliminaire et une validation des données doivent être effectués au niveau de la PSU/régional/de l'État. L'analyste doit examiner les événements ainsi que les questionnaires remplis pour identifier les problèmes potentiels, tendances ou données incohérentes. L'examen des données produira aussi des fréquences, par question, pour revue. Les informations sur la façon d'examiner les données au format CSV et le questionnaire transposé (tableau des réponses) figurent dans le **Chapitre 5**, GSS Developer's Tool Set, du **Guide du programmeur sur le système général de l'enquête GATS** et dans l'annexe d'Affichage des données brutes du **Plan de mise en œuvre de la gestion des données du GATS**.

4.3 Gestion des données au Centre national de données [CND] (NDC)

Cette section décrit les procédures de collecte et de gestion des données au niveau du NDC : rassemblement/agrégation des fichiers, aspects de contrôle de qualité, y compris des rapports et la surveillance, la communication et l'assistance technique, et la création de fichiers de données brutes. Il est à noter que toutes les procédures énumérées sous la gestion de données aux niveaux de la PSU/régional/de l'État doivent également s'appliquer à la gestion des données du NDC.

4.3.1 Agrégation

Comme les données font partie de chaque point de l'agrégation des données, les fichiers doivent être examinés, traités et soumis à des activités de contrôle de qualité. Un processus de documentation et d'archivage des fichiers de données doit avoir lieu.

Pendant ce processus, le logiciel d'agrégation trie et traite les fichiers dans le dossier sélectionné (généralement C:\GATS\DATA) par ordre alphabétique indépendamment de la date. Il est essentiel que les fichiers soient agrégés dans l'ordre de leur exportation. Lors de l'agrégation de tous les fichiers exportés, la convention du nom de fichier d'exportation sdf FI#_YYYY_MM_DD.sdf entraîne le tri alphabétique automatique et le traitement correct des données. L'exportation des fichiers est cumulative, par conséquent le fichier le plus récent contient toutes les données, y compris les mises à jour ou les modifications qui se trouvaient dans les fichiers précédents.

Le processus d'agrégation ne permet pas les doublons de numéros de cas.

Processus d'agrégation

- Au début de la collecte de données, un fichier maître d'agrégation existe dans le dossier C:\GATS\DATA appelé master.sdfzero. Ceci est nécessaire au début la collecte de données afin que l'agrégation puisse commencer avec un fichier maître ne contenant aucune donnée. Le fichier master.sdfzero doit être copié, non pas renommé, en master.sdf lorsque la réinitialisation est nécessaire.
- Un fichier Node.id doit résider dans le dossier C:\GATS\
 - Chaque ordinateur exécutant l'agrégation doit avoir un nom de nœud distinct. Avant que l'agrégation ne soit exécutée pour la première fois sur l'ordinateur, le fichier nœud doit être modifié pour contenir un nom unique et significatif.
 - Une fois que les données ont été agrégées, le fichier « Node.ID » ne doit pas être modifié.
 - Le fichier Node.id doit être unique pour chaque point d'agrégation.
- Agrégez les fichiers sdf collectés les plus récemment à ceux précédemment agrégés à des fins d'efficacité aussi longtemps que les fichiers sont reçus et traités dans l'ordre (par date d'exportation). Bien qu'il soit acceptable d'agréger vers un fichier maître vide à chaque agrégation, il convient de noter que cela prendra plus longtemps, car tous les fichiers exportés sdf auront besoin d'être traités.
- Enregistrez tous les fichiers exportés reçus. Archivez-les une fois agrégés dans le dossier C:\GATS\Archive. Ils n'ont pas besoin d'être agrégés à nouveau.

Il est important que les fichiers exportés de l'Enquêteur de terrain soient agrégés dans l'ordre de leur date d'exportation.

Pendant et après l'agrégation, plusieurs points doivent être vérifiés :

- Tous les fichiers reçus sont des noms de fichier datés lorsqu'ils sont exportés, assurant ainsi que le traitement n'est pas effectué lorsque les fichiers ne sont pas dans l'ordre. Si des fichiers sont reçus dans le désordre (c.-à-d., des fichiers de données antérieurs sont reçus plus tard) les données les plus récentes pour un numéro de cas (CaseID) particulier seront remplacées par des données antérieures, peut-être incomplètes. Afin de vous assurer lors de l'agrégation à un fichier sdf maître non vide que vous avez traité les données les plus récentes, consultez les fichiers archivés sdf. (Agrégez toujours les fichiers dans l'ordre des dates ou tout simplement utilisez les fichiers les plus récents).
- Exactitude et exhaustivité de l'agrégation : Tous les fichiers sdf prévus ont-ils été pris en compte et inclus dans le processus d'agrégation ?
- Le fichier sdf le plus récent pour chaque FI a-t-il été agrégé ?
- Examen des données : Une étape de vérification et de validation doit être effectuée. L'analyste doit examiner les données d'événements ainsi que les questionnaires remplis pour identifier les problèmes potentiels, tendances ou données incohérentes. Il générera aussi des

fréquences, par question, pour examen. Des informations sur Raw Data Viewer figurent dans le **Plan de mise en œuvre de la gestion des données du GATS**.

Examinez le rapport de synthèse d'agrégation et le rapport de statut d'agrégation générés par l'agrégation pour confirmer les comptes. Le rapport de statut est produit après le processus d'agrégation.

Le rapport de synthèse d'agrégation des données fournit des informations sur les résultats de l'agrégation. Chaque ligne présente les résultats d'une unité d'agrégation (généralement un enquêteur de terrain) et un tableau de base de données spécifique. Le tableau est noté dans la colonne AggTable, le nombre de lignes dans le tableau d'entrée est présenté dans la colonne AggRowsTotal, et le nombre de lignes insérées à partir de ce tableau est présenté dans la dernière colonne AggRowsInserted.

Le rapport Household Screening Status (rapport de présélection des ménages) est un récapitulatif du statut de tous les cas dans le fichier de données agrégées. Pour chaque unité d'agrégation (généralement un enquêteur de terrain), le nombre total de cas est rapporté pour cette unité dans la colonne Cases (cas). Puis le total est réparti en différentes catégories de statut. Les colonnes de statut (Completed Ints, Unworked, Pending, et Final) dépendent du code du résultat le plus récent du cas.

- En utilisant les données agrégées ou tout autre fichier sdf, un fichier plat de données formatées ou un fichier CSV peut être créé à l'aide de la fonctionnalité d'agrégation des données (Data Aggregation) dans les outils des développeurs (Developers Tools). L'ensemble d'outils du développeur du GSS utilise le sdf2csv et transpose les applications.
- Il est recommandé d'examiner les données d'événement (contenues dans les fichiers DU et DuEvt) pour leur contenu et la compréhension des opérations sur le terrain.
- Le fichier DU ne contient qu'un enregistrement par numéro de cas (HH et IQ ont des numéros de cas distincts -00 et -01 respectivement) et le code de résultat le plus récent pour ce numéro de cas. Le fichier DuEvt contient un historique des événements, pour un numéro de cas donné, saisi par le ou les FI, et peut contenir plus d'un enregistrement par numéro de cas.

Générer des fichiers de données pour examen

- Les données du questionnaire (tableau des réponses) sont stockées dans un tableau correspondant aux deux formulaires, HH et IQ. Il y a une ligne par question (QID), par conséquent ces données doivent être restructurées avant l'analyse. Les données des réponses doivent être transposées pour être facilement examinées et utilisées avec un logiciel statistique. Cela créera un enregistrement basé sur le CaseID à la fois pour les données du HH et de l'IQ. L'ensemble d'outils du développeur du GSS transpose les données des réponses sur la base des fichiers MDB de l'enquête spécifiés avec des fichiers séparés au niveau du numéro de cas, des fichiers plats, pour le HH et l'IQ.
- Pour créer un enregistrement de numéro de cas contenant à la fois les données de réponses au questionnaire de l'enquête et les données d'événements, l'analyste du système doit fusionner les fichiers de réponses transposées pour le questionnaire HH et IQ avec le DuEvt par numéro de cas. Le fichier DuEvt doit être examiné pour les codes de résultat temporaires et les éventuelles données incomplètes. À la fin de la collecte des données, chaque numéro de cas exploité doit avoir reçu un code de résultat final.

- Les données d'événements du GATS sont stockées dans le tableau DUEVT, avec une ligne pour chaque résultat pour chaque numéro de cas. Les données du GATS sur les unités d'habitation (DU, soit dwelling unit) sont stockées dans le tableau DU avec une ligne par numéro de cas. Les données de DU contiennent également une copie du code de résultat le plus récent pour chaque cas. Les données d'événements et de DU ne nécessitent pas de transposition et les fichiers pour l'analyse de ces données peuvent être créés en utilisant les utilitaires de conversion CSV (Export Data to CSV) dans le menu d'agrégation de données des outils du développeur du GATS.

Le tableau ci-dessous résume ces sources de données provenant du système iPAQ.

Tableau des sources de données	Tableau	Lignes	Transposition requise ?	Outil pour obtenir les données
Données du Questionnaire	Réponses	Une ligne par question pour chaque CasID et formulaire	Oui	Menu d'agrégation et de transposition de données
Données d'événement	DUEVT	Une ligne par résultat pour chaque CasID et formulaire	Non	Menu d'agrégation des données et d'exportation au format CSV
Unité d'habitation	DU	Une ligne par unité d'habitation	Non	Menu d'agrégation des données et d'exportation au format CSV
Modifications ou changements d'adresse	Address Log	Une ligne par modification d'adresse	Non	Menu d'agrégation des données et d'exportation au format CSV
Remarques du FI	Remarques	Une ligne par type de remarque (soit au niveau du CasID ou par question (QID))	Non	Menu d'agrégation des données et d'exportation au format CSV

4.3.2 Contrôle qualité des données recueillies

Au NDC, des procédures de contrôle de la qualité doivent être en place pour examiner, surveiller, rapporter et sauvegarder régulièrement les données comme elles arrivent et sont traitées. Le NDC doit établir un calendrier de traitement hebdomadaire au minimum pour les rapports et le suivi, ainsi qu'un calendrier qui comprend également un processus de sauvegarde de routine.

Le protocole standard du GATS consiste en une liste de variables indicatrices du GATS et des rapports suggérés au niveau national. La liste des variables indicatrices clés figure dans l'**annexe C.1** tandis que les détails et informations complémentaires se trouvent dans les **Définitions des indicateurs du GATS**. Les pays peuvent ajouter/adapter les

Il est recommandé que les variables indicatrices soient examinées au niveau des données brutes pendant la collecte des données.

variables indicatrices au questionnaire final approuvé au besoin.

Rapports développés nationaux : Il est recommandé au pays de développer des rapports de statut des événements au niveau des cas. Le suivi de l'enquête (par exemple, progrès sur le terrain, recueil, transmission des données) est une activité permanente clé, coordonnée.

- Pendant la période de recueil de données, le NDC doit fournir un rapport hebdomadaire de taux de réponse généré à l'aide de l'utilitaire GSS IDE > Aggregation > Reporting. Les rapports doivent indiquer le nombre d'entrevues complétées au niveau du pays/de la région/de l'État et le taux de réponse global. Des exemples de rapports de travail sur le terrain et des statuts de taux de réponse se trouvent dans le ***GATS Programmer's Guide to General Survey System (Guide du programmeur du système général de l'enquête du GATS)***.
- Le NDC est encouragé à utiliser les rapports produits par l'IDE du GSS et au besoin de rédiger des rapports avec Access^{1,2}, SPSS^{1,3}, SAS^{1,4}, STATA^{1,5} ou d'autres logiciels qui permettront l'examen des données pour les incohérences, les anomalies et l'inachèvement. Les données agrégées transposées doivent être utilisées pour générer les rapports.

Les incohérences dans les données peuvent être identifiées en effectuant des contrôles de validation de toutes les variables dans l'ensemble et par rapport aux HH et IQ. Veuillez-vous référer au **Chapitre 5** sur l'assurance qualité suivant le recueil des données pour plus d'informations sur les contrôles de validation.

Procédures de sauvegarde : Les données doivent être sauvegardées régulièrement, par exemple une fois par semaine, en fonction de la fréquence de l'agrégation.

- Les fichiers sdf individuels (au niveau du FI) doivent être archivés dans le dossier C:\GATS\Archive.
- Les données archivées doivent également être sauvegardées au niveau d'un point secondaire, par exemple un réseau ou un lecteur flash et une copie doit être conservée dans deux lieux physiques distincts.
- Lorsque le recueil des données est terminé, la réconciliation des données reçues doit être effectuée par rapport au fichier maître de sélection de l'échantillon ainsi que par rapport aux fichiers exportés à partir du terrain.

4.3.3 Communication et Support technique

La communication du NDC lors de la collecte de données doit comporter les rapports hebdomadaires des statuts envoyés au Centre de coordination des données du GATS (DCC) et le signalement de tout problème technique. Tous les problèmes techniques associés à l'iPAQ et aux logiciels, qu'ils soient résolus ou non doivent être signalés par courrier électronique dès que possible quand ils se produisent. Cela permettra de faciliter un support technique rapide, la résolution des problèmes et le partage d'informations.

¹ Les marques de produits cités le sont uniquement à titre informatif et n'ont pas fait l'objet de recommandations de la part du Département de la santé et des services humanitaires des États-Unis.

² Microsoft Office Access© (Microsoft Corporation, Redmond, Washington).

³ SPSS© (SPSS Inc., Chicago, Illinois).

⁴ SAS© (SAS Institute Inc., Cary, North Carolina).

⁵ STATA© (Stata Corp., College Station, Texas).

4.3.4 Préparation des fichiers de données brutes

Fichiers d'entrée : Pour la surveillance et l'établissement de rapports au niveau du pays, l'analyste doit lire le fichier de questionnaire transposé (réponses) ou les fichiers CSV qui ont été créés après l'agrégation (DU et/ou DuEvt). Le **Guide du programmeur sur le système général de l'enquête GATS** inclut des informations concernant la structure des tableaux de l'iPAQ et des informations de terrain.

Seules les données des réponses nécessitent une transposition.

Considérations sur le format des données et la structure des fichiers TXT et CSV :

Format :

Réponses

- Les données de l'enquête transposées à partir des tableaux des réponses sont créées à l'aide de l'utilitaire de transposition dans le menu d'agrégation des données des outils des développeurs. Les données de réponse sont écrites dans un fichier txt ou CSV (fichier plat). Le fichier txt ou CSV peut être lu avec un logiciel de statistiques tel que SAS ou SPSS en utilisant la liste des variables générées avec l'ensemble d'outils du développeur du GSS. La liste des variables produite par l'utilitaire de transposition peut être utilisée pour créer des états d'entrée pour un progiciel de statistiques donné. À la fin de la collecte des données ou pendant les intervalles intermédiaires, l'utilitaire Master File Merge peut être utilisé pour combiner toutes les bases de données (Response, Events, Master Sample Selection dataset, Notes, AddressLog) en un seul fichier délimité par des virgules et créer des programmes d'entrée SAS, SPSS, et STATA capables de lire le fichier nouvellement fusionné.
- Le processus de transposition ne sélectionne que les données des réponses valides pour un CaseID (numéro de cas donné), puis trie par QID, date. Pour un numéro de cas donné, l'élément de donnée le plus récent pour chaque QID est inscrit dans le fichier de données transposées.

DU, DuEvt, Notes, Address log

- Ces tableaux sdf peuvent être écrits en format CSV (valeurs séparées par des virgules) en utilisant soit l'exportation des données au format CSV (Export Data to CSV) ou l'application afficher les données (View Data) dans le menu d'agrégation des données des outils du développeur du GSS. Les pays qui n'utilisent pas l'ensemble des outils du développeur du GSS peuvent utiliser Raw Data Viewer qui est disponible en tant qu'utilitaire autonome dans le dossier C:\GATS\bin.
- Les fichiers texte CSV créés sont délimités par des virgules avec des champs encadrés de doubles guillemets. Ces fichiers délimités sont des tableaux de textes ASCII organisés, avec des colonnes séparées par des virgules et des lignes séparées par des retours. La première ligne du fichier contient les noms des colonnes. Les fichiers texte délimités par des virgules peuvent être ouverts par plusieurs types d'applications, notamment Excel. La plupart des outils d'analyse statistique peuvent également les lire si l'analyste spécifie le délimiteur.

Structure :

- Pour chaque formulaire d'enquête, les données agrégées de l'iPAQ contiennent des enregistrements du niveau de la question à partir du tableau des réponses. Le fichier des données agrégées avant la transposition contient les enregistrements HH et IQ, chacun figurant séparément, par numéro de cas. Le formulaire est identifiable par les deux derniers chiffres du champ du numéro de cas (00 = HH, 01 = IQ). Le programme de transposition dans l'ensemble des outils du développeur du GSS utilise le contenu des fichiers MDB Survey0 et Survey1 pour transposer les données. Par conséquent, les données HH et IQ sont écrites de façon à séparer les fichiers de données transposés.
- Conformément à la lecture de tout fichier texte, il est nécessaire d'identifier le format des variables pour chaque champ dans l'enregistrement. L'analyste doit être prêt à lire les données du fichier texte en utilisant une entrée ou à obtenir les énoncés des données selon le logiciel utilisé, sinon toutes les variables seront lues comme des « chaînes » ou des données texte. L'ensemble d'outils du développeur du GSS peut générer une liste de variables à partir du menu Data Aggregation, Transpose. Il existe plusieurs types de données du GSS dans un fichier transposé et les analystes doivent prévoir de traiter les types numériques et ATA en tant que chiffres et tous les autres (CaseID, Tstamp, text, list) comme du texte.
- Certaines langues contiennent des données texte en format Unicode. Pour ces langages, il est nécessaire d'utiliser l'option Unicode (par exemple avec SPSS, SAS ou STATA) lors de la lecture et du traitement des données.

Les fichiers utilisés pour l'analyse ne doivent contenir que des données valides. Utilisez l'ensemble des outils du développeur du GSS, cette règle s'applique aux données des réponses au moment de l'agrégation.

5. Assurance qualité : Période Post Collecte des données

La phase suivant la collecte des données comprend une série de tâches qui doivent être accomplies afin de préparer un fichier de données analytiques pour effectuer des analyses de données, et désigne une étape ayant lieu lorsque toutes les données de l'enquête ont été recueillies et agrégées. Cela comprend la préparation des données pour les calculs de pondération de l'échantillon, l'évaluation de la qualité de l'échantillonnage, les erreurs d'échantillonnage et de pondération, la mesure de la qualité de la couverture de la base de sondage, le niveau de non-réponse, et d'autres erreurs non dues à l'échantillonnage. Cela inclut également la préparation d'un fichier de données analytiques. Ce chapitre décrit les directives d'assurance qualité et les procédures applicables et recommandées pour chacune de ces activités, portant spécifiquement sur :

5.1 Nettoyage et préparation des données pour les calculs de pondération de l'échantillon

- 5.1.1 Création du fichier maître de base de données
- 5.1.2 Suppression des variables d'identification confidentielles
- 5.1.3 Nettoyage et validation du fichier de données fusionnées
- 5.1.4 Création des codes de décision finale

5.2 Mesures de la qualité : Échantillonnage, erreur d'échantillonnage, et pondération

- 5.2.1 Modèle de correction de l'échantillonnage de la pondération poste-stratification entre les cellules d'ajustement
- 5.2.2 Effet multiplicatif de la variable de pondération de l'échantillon sur la précision des estimations de l'enquête
- 5.2.3 Effet global de la conception sur la précision des estimations de l'enquête et sur l'homogénéité intragroupe dans les PSUs des estimations clés correspondantes de l'enquête
- 5.2.4 Marge d'erreur pour les estimations clés de l'enquête

5.3 Mesures de la qualité : Couverture, non-réponse, et autres erreurs non dues à l'échantillonnage

- 5.3.1 Taux de couverture de la base des foyers
- 5.3.2 Modèles de taux de réduction des répondants
- 5.3.3 Modèles de taux de réponse des ménages par strates d'échantillonnage de la première phase
- 5.3.4 Modèles des taux de réponse des personnes parmi les variables utilisées pour les ajustements en fonction des non-réponses
- 5.3.5 Modèles des taux de refus des personnes parmi les variables utilisées pour les ajustements en fonction des non-réponses
- 5.3.6 Taux de réponse aux questions pour les variables indicatrices de la Fiche d'Information

5.4 Examen officiel de la qualité statistique

5.5 Création du fichier des données analytiques

5.1 Nettoyage et préparation des données pour les calculs de pondération de l'échantillon

Cette section fournit des directives sur la fusion des fichiers de données, la validation des variables et des enchaînements, ainsi que la création des codes de décision en utilisant un logiciel statistique¹.

5.1.1 Création du fichier maître de base de données

L'option Master File Merge combine les données de l'enquête dans un seul fichier délimité par des virgules et crée des entrées de programmes SAS, SPSS et STATA capables de lire le fichier maître de base de données nouvellement fusionnées. Il peut être créé en utilisant l'utilitaire GSS IDE > Aggregation > Master File Merge.

Les fichiers utilisés dans Master File Merge comprennent le fichier échantillon maître de sélection de l'échantillon, le fichier SDF créé en agrégeant les fichiers sdf de tous les enquêteurs, et les deux bases de données des questionnaires (GATS_Survey0.mdb et GATS_Survey1.mdb). Le fichier maître de sélection de l'échantillon est la base de données MS Access utilisée pour créer le fichier de cas pour l'ordinateur de poche (reportez-vous à l'**annexe B.4** pour plus d'informations sur le fichier maître de sélection de l'échantillon). Des champs supplémentaires peuvent être ajoutés à cette base de données, mais les noms des champs, les types et tailles ne doivent pas être mis à jour. La description des champs utilisés doit être mise à jour afin que les étiquettes descriptives produites par SAS et SPSS ou STATA soient correctes.

Lorsque vous utilisez l'utilitaire de fusion des fichiers maître, assurez-vous d'utiliser l'identification des emplacements de fichiers mis à jour du fichier maître de sélection de l'échantillon, le fichier agrégé, et les bases de données du questionnaire de l'enquête. Pour changer, mettez à jour l'emplacement du fichier ou le nom dans la zone de texte ou sélectionnez le bouton « ... » et le système vous permettra de sélectionner un fichier existant. Le fichier maître de base de données sera un fichier au format Unicode (UTF-8) délimité par des virgules. L'utilitaire génère également le code d'entrée pour la lecture des données avec SAS, SPSS et STATA.

Pour afficher le fichier CSV, un éditeur de texte Unicode tel que WordPad est nécessaire. Les fichiers SAS, SPSS et STATA sont également Unicode pour s'assurer que les étiquettes et les formats sont affichés correctement. Les programmes SAS, SPSS et STATA peuvent avoir besoin d'être mis à jour avant de les lancer pour s'assurer que chaque option spécifique d'un pays est définie. Le processus pourra créer un ensemble de données maître. Ce fichier maître de base de données créé doit être utilisé pour enlever toutes les variables d'identification confidentielles et doit être communiqué au point focal informatique du pays.

(Pour créer manuellement l'ensemble de données maître, veuillez-vous référer à la procédure décrite dans l'**annexe D.1**)

5.1.2 Suppression des variables d'identification confidentielles

À partir de l'ensemble de données maître créé, supprimez les champs identifiants confidentiels de la section sélection de l'échantillon maître. Les variables qui doivent rester dans les champs de la nouvelle

¹Le centre de coordination des données (DCC) de la GATS fournit une assistance technique pour l'utilisation des logiciels statistiques suivants : SAS, SPSS et STATA.

section sélection de l'échantillon maître sont les suivantes : NUMÉRO DE CAS, TYPE, SEXE (si applicable), STRATE, PSU, SEGMENT, et MENANAGES.

5.1.3 Nettoyage et validation du fichier de données fusionnées

Vérifiez que les variables ont des valeurs appropriées, que les enchaînements fonctionnent correctement, et vérifiez les champs vides. Alors que beaucoup de ces contrôles de qualité des données sont exécutés lors de l'administration de l'entretien par le logiciel de l'ordinateur de poche, il est important de vérifier à nouveau au cas où des erreurs n'auraient pas été détectées dans la programmation du logiciel de l'ordinateur de poche.

1. Vérifiez que les enchaînements de données ont fonctionné comme il est spécifié dans le questionnaire. Les enchaînements originaux des données tirés du questionnaire approuvé final du pays doivent être vérifiés pour toutes les variables de base, et il est recommandé de vérifier également les modèles pour toutes les autres variables.
 - Un répondant qui n'a pas répondu à une question en raison d'un enchaînement aura un champ vide correspondant à la question sautée.
 - Par exemple, si un répondant a refusé de répondre à la question A02a (le mois de naissance du répondant) ou A02b (l'année de naissance du répondant), alors le répondant doit répondre à la question A03 (l'âge du répondant), mais si le répondant a fourni une réponse à A02a et A02B, alors A03 doit être passée.
 - Par exemple, la personne interrogée répond à la question B01 qu'il fume « quotidiennement » (une valeur de 1) alors elle doit passer à la question B04 et ne doit pas répondre aux questions B02 et B03.
 - Dans SAS, un champ vide pour une variable numérique est égale à '.' (Pour les variables de caractère, c'est ' '). Dans SAS, cet enchaînement peut être vérifié en utilisant le code ci-dessous. Tous les enregistrements qui ne suivent pas l'enchaînement seront transférés dans un ensemble de données SAS.

```
IF B01 = 1 & (B02 NE. OR B03 NE.) THEN FLAG=1;
ELSE IF B01 = 2 & (B02 = . OR B03 = .) THEN FLAG=1;
ELSE IF B01 IN (3,9) & (B02 NE. OR B03 = .) THEN FLAG=1;
ELSE FLAG=0;
IF FLAG=1 THEN OUTPUT; *Une valeur de 1 pour FLAG indique que l'enchaînement
n'a pas fonctionné et une valeur de 0 indique l'enchaînement a fonctionné ;
```
2. Vérifiez que seuls les champs vides sont ceux qui résultent d'un enchaînement (par exemple, ils ne sont pas applicables pour le répondant et doivent être vides). Tous les autres champs vides ne sont pas valides.
 - Les champs qui sont vides en raison des enchaînements devront recevoir une nouvelle valeur. Cela permettra de les différencier des champs vides qui ne sont pas attribuables aux enchaînements (champs vides invalides). Puisque les valeurs des variables d'origine ne doivent pas être remplacées, créez un nouvel ensemble de données pour compléter cette étape.

- Dans SAS utiliser le code '.S' pour une valeur manquante, lorsque la question n'était pas applicable (valablement sauté). Le code SAS pour l'enchaînement utilisé ci-dessus pourrait être modifié comme suit :

```
IF B01 = 1 & (B02 = . OR B03 = .) THEN B02 = .S & B03 = .S;
ELSE IF B01 IN (3,9) & B02 = . THEN B02 = .S;
```

- Lorsque toutes les valeurs d'enchaînement ont été remplacées, il ne devrait pas rester de champs vides. Si d'autres champs vides existent, transférez ces enregistrements dans le fichier d'erreur.
3. Vérifiez chaque variable afin de vous assurer qu'il n'y a pas de valeurs non valides. Utilisez le dictionnaire des données du pays (livre des codes) pour voir les valeurs valides pour chaque variable dans l'ensemble des données.
- Vérifiez que les valeurs pour les réponses à chaque question sont valides.
 - Par exemple : Les seules réponses valides pour B01 sont : 1 (« quotidiennement »), 2 (« moins d'une fois par jour »), 3 (« pas du tout »), 9 (refusé), ou le code de valeur manquante spécifique au logiciel utilisé (par exemple, .S en utilisant SAS).
 - Par exemple : Certaines variables ont des plages qui leur sont associées, de ce fait une valeur non incluse dans la plage ne doit pas exister. Donc, pour la variable B04, la valeur la plus basse possible doit être 0 et la valeur la plus élevée possible, doit être 98. Il y a aussi une option pour la valeur de 99 si la réponse était : « Ne sais pas » ou « Refusé ». Le code de valeur manquante spécifique au logiciel utilisé (par exemple, .S en utilisant SAS) serait également une valeur valide.
 - *Avec SAS, les enregistrements avec des réponses invalides pour B01 peuvent être sélectionnés et transférés dans un fichier d'erreurs en utilisant le code suivant :*

```
SI B01 NOT IN (1,2,3,9) PUIS OUTPUT;
```
 - *Suivant la préférence de l'analyste, des fréquences simples peuvent être exécutées sur chacune des variables pour voir si les valeurs ne sont pas dans la plage. Tous les enregistrements avec des valeurs invalides peuvent alors être identifiés et entrés dans le fichier d'erreurs.*
 - Vérifiez l'âge du répondant
 - Avant de vérifier que le répondant a le bon âge, l'âge doit d'abord être calculé et inclus dans l'ensemble de données. Afin de faire le calcul, utilisez la date de l'entretien de l'IQ (IQ_EVENTDATE) et le mois (A02a) et l'année (A02b) de naissance des répondants, avec le jour censé être le 15 du mois. Si les répondants répondent « Ne sais pas » ou « Refusé » à la question A02a ou A02b, alors l'enquêteur passe à la question A03 où les répondants sont invités à donner leur âge (toutefois, le répondant ne doit répondre à A03 que si A02a ou A02b ne sont pas consignées). Si le répondant ne donne pas le mois ou l'année de naissance, utilisez A03 pour l'âge du répondant. Si le mois et l'année sont donnés, calculez l'âge en soustrayant la date de naissance à la date de l'entretien. Vérifiez que l'âge du répondant est valide (la plage valide va de 15 à 125).

- Dans SAS l'âge pourrait être calculé, puis vérifié en utilisant le code suivant :

```
IF A02A <=12 AND A02B NE 7777 AND A02B NE 9999 THEN DO;
  BIRTHDATE = MDY(A02A, 15, A02B);
  AGE = FLOOR(YRDIF(BIRTHDATE, IQ_EVENTDATE, 'ACTUAL'));
  END;
  ELSE AGE = A03;
```

4. Autres contrôles de validation des données :

- Assurez-vous que si aucune des variables suivantes n'est sautée, leurs valeurs sont inférieures à l'âge du répondant : B04, B05, B08, B09, B11, B12, B13, C04, C05, C08, C09, C11, C12, et C13 (selon l'unité de la variable, ces valeurs de variables peuvent dans un premier temps avoir besoin d'être converties en années).

5.1.4 Attribution des codes de décision finale

Le fichier des données fusionnées comprendra les codes de résultat finaux qui ont été attribués par les FI à tous les cas exploités de l'échantillon pendant la collecte des données (voir le **Manuel de l'enquêteur de terrain de la GATS** pour la description des codes de résultat de terrain). Chaque cas d'entretien d'un ménage et chaque cas d'entretien individuel doivent avoir un code de résultat final (voir la **section 4.1.3** pour l'attribution des codes finaux de résultat de non-entretien lors de la collecte des données).

Des codes de décision finale doivent être attribués à chaque cas. Les codes de décision finale sont utilisés pour calculer et déclarer les taux de réponse et les mesures d'assurance qualité.

En utilisant les codes de résultat finaux, les codes de décision finale doivent être attribués. Les codes de décision finale sont ensuite utilisés pour calculer et rapporter les taux de réponse et les mesures d'assurance qualité. Reportez-vous à l'**annexe D.2** afin d'attribuer les codes de décision finale à partir des codes de résultat finaux, pour les calculs des taux de réponse et un modèle de rapport de ces résultats issus de la collecte des données.

Voici des directives importantes qui doivent être suivies pour ce processus :

1. Chaque cas doit recevoir un code de décision finale pour le HH basé sur le code final du résultat du HH sur le terrain. Si l'entretien HH est rempli et une personne a été sélectionnée pour l'entretien IQ (code de décision finale 1), un code de décision finale doit aussi être attribué à l'IQ. Si le code décision finale HH n'est pas 1 (par exemple, personne n'a été sélectionné, refus, incomplet), alors il n'y aura pas de code de décision finale pour l'IQ puisque l'entretien IQ n'a jamais eu lieu.
2. L'IQ comporte des questions clés qui déterminent la légitimité de l'entretien. Si la réponse enregistrée pour une de ces questions clés est « Ne sais pas » ou « Refusé », l'entretien n'est pas considéré comme légitime. Ainsi, pour tous les cas IQ avec un code de résultat final de 400 (*Questionnaire individuel Rempli*), les règles ci-dessous doivent être suivies afin de déterminer quel code de décision doit être attribué.
 - Si B01, B02, ou B03 (questions clés de prévalence du tabagisme) ont été enregistrées comme « Ne sais pas » ou « Refusé », un code de décision finale de 16 (répondant sélectionné incompetent) doit être attribué. Cela ne devrait pas se produire souvent. (Cette règle peut également être appliquée pour C01, C02, et C03 (questions clés de prévalence

du tabac sans fumée) uniquement si l'utilisation du tabac sans fumée est fréquente dans le pays).

- Pour tous les autres cas d'IQ avec un code de résultat final de 400 (*questionnaire individuel* terminé), un code de décision finale de 11 (*questionnaire individuel* terminé) doit être attribué.

Cela se produira pour la plupart des cas.

3. Pour tous les cas IQ avec un code de résultat final de 402 (*questionnaire individuel* complété partiellement), les règles ci-dessous doivent être suivies afin de déterminer le code de décision à attribuer.
 - Si l'IQ a été complété au moins jusqu'à la question E01 ET aucune des questions clés (décrites dans le paragraphe 2 ci-dessus) n'a reçu de réponses « Ne sais pas » ou « Refusé », un code de décision finale de 11 (*questionnaire individuel* terminé) doit être attribué.
 - Si l'IQ a été interrompu avant la question E01, un code de décision finale de 12 (incomplet) doit être attribué. (Les cas avec un code de décision finale de 12 (incomplet) sont considérés comme des non-répondants pour la GATS et les données ne sont pas incluses dans l'analyse).
 - Si l'IQ a été complété au moins jusqu'à la question E01, mais une des questions clés (décrites dans le paragraphe 2 ci-dessus) a reçu une réponse « Ne sais pas » ou « Refusé », un code de décision finale de 16 (répondant sélectionné incompetent) doit être attribué. (Comme il est décrit dans le paragraphe 2 ci-dessus, ces répondants sont considérés comme étant incompetents s'ils ne peuvent pas fournir de réponses légitimes aux questions clés et, par conséquent, leurs données d'entretien ne sont pas incluses dans l'analyse).
4. Seuls les cas avec un code de décision de 11 doivent être inclus dans l'ensemble final des données analytiques. (Donc, il est important d'attribuer les codes de décision correctement).
5. Utilisez les onglets croisés pour vérifier tous les codes de résultat finaux par rapport aux codes de décision pour déterminer une éventuelle classification erronée.
Si les deux codes ne correspondent pas comme ils le devraient, cela indique un problème avec le code du logiciel utilisé pour créer les codes de décision.

5.2 Mesures de qualité : Échantillonnage, erreur d'échantillonnage, et pondération

Cette section décrit les calculs pour évaluer directement la qualité des estimations à partir des échantillons du GATS, et pour indiquer les effets de grappes d'échantillonnage et de pondération inégale sur ces estimations. Des directives pour vérifier l'exactitude de la pondération calculée sont également incluses.

5.2.1 Modèle de correction de l'étalonnage de la pondération poste-stratification entre les cellules d'ajustement

La dernière étape de production des pondérations de l'échantillon implique l'étalonnage de la pondération du dénombrement de la population par des corrélats connus de mesures de résultats d'étude clé,

appelées variables d'étalonnage (par exemple, le sexe, l'éducation, l'âge, environnement urbain/rural et la région, comme le suggère le **Manuel de pondération de l'échantillon du GATS**).

Le contexte général et les instructions pour le calcul des corrections poste-stratification sont décrits ci-dessous dans cette section. Pour plus de détails sur ces sujets, référez-vous à l'**annexe D.3**.

1. **Contexte** — L'étalonnage compense le déséquilibre qui n'est pas modifié par la correction en fonction de la non-réponse. Des corrections distinctes sont appliquées à tous les membres de cellules d'ajustement formées de manière stratégique. On souhaite augmenter l'importance de ces sous-groupes de population qui demeurent sous-représentés et diminuer l'importance de ces sous-groupes de population qui sont surreprésentés. Plus les valeurs de correction d'étalonnage s'écartent de 1,00 (augmentent ou diminuent), plus fort sera l'impact potentiel du déséquilibre de l'échantillon (au-delà de celui compensé par une correction à la non-réponse) sur le biais des estimations de l'enquête.
2. **Effectuer des corrections post stratification** — En termes de procédure, l'étalonnage sous la forme de post stratification consiste à former des « cellules d'ajustement » par la classification croisée des mesures de corrélats. La « correction post stratification » (PSA) dans chacune de ces cellules d'ajustement est < 1 si le sous-échantillon dans cette catégorie était surreprésenté dans l'échantillon après comptabilisation de la sélection de l'échantillon et des non-réponses, et le PSA > 1 dans les catégories où l'échantillon était sous-représenté.
3. **Rapports de correction post stratification** — Créez un tableau qui répertorie l'ensemble des cellules d'ajustement, en indiquant pour chacune la manière dont les cellules sont définies par les variables catégorielles utilisées pour l'étalonnage. Pour chaque cellule, enregistrez la valeur calculée de l'ajustement post stratification et observez sa taille par rapport à 1,00. Elle est optimale si tous les PSA sont proches de 1, avec certains un peu supérieurs à 1, et le reste un peu inférieur à 1,00.

5.2.2 Effet multiplicatif de la variable de pondération de l'échantillon sur la précision des estimations de l'enquête

Le **Manuel de conception de l'échantillon du GATS** recommande une conception où les probabilités de sélection (et donc la pondération de l'échantillon) varient quelque peu en raison de l'utilisation de mesures estimées des tailles des grappes, des corrections à la pondération de l'échantillon, et de l'attribution égale des tailles d'échantillon entre les régions lorsque les estimations régionales répondant aux normes de précision du GATS sont nécessaires. Le **Manuel de conception de l'échantillon de la GATS** décrit comment ces pondérations doivent être calculées. Une fois que les données du questionnaire ont été nettoyées, et la pondération finale de l'échantillon a été attribuée, les données de l'échantillon du GATS sont prêtes pour l'analyse et cette tâche.

Des informations générales et des instructions de calcul à cet effet sont présentées ci-dessous dans cette section. Pour plus de détails sur ces sujets, référez-vous à l'**annexe D.4**.

1. **Contexte** — La variation de la pondération de l'échantillon peut augmenter le nombre d'erreurs d'échantillonnage dans les estimations de l'enquête et par

Plus les pondérations sont variables, plus la valeur de M_{eff_w} est grande.

conséquent entraîner des écarts plus grands et des erreurs moyennes plus importantes pour ces estimations. Plus précisément, l'augmentation multiplicative de la variance des estimations d'enquête, appelé $Meff_w$, dépend du niveau de variabilité de la pondération de l'ensemble des observations de l'échantillon qui sont utilisées pour produire l'estimation.

2. **Estimation de $Meff_w$** — La moyenne simple et la variance de la pondération pour l'ensemble des observations de l'échantillon utilisées pour produire des estimations d'enquête sont nécessaires pour les calculs $Meff_w$.

- La valeur de $Meff_w$ pour les estimations tirées de cet ensemble d'observations de l'échantillon est mesurée en calculant d'abord le rapport de la variance sur le carré de la moyenne, puis en ajoutant un à ce rapport. La formule spécifique de calcul est l'équation (1) de l'annexe D.4.

3. **Rapport de $Meff_w$** — Puisque $Meff_w$ s'applique à toutes les estimations tirées de l'ensemble spécifique des observations de l'échantillon qui définit un domaine de rapport pour les estimations d'enquête (par exemple, $Meff_w$ couvrant toute la population, la population rurale, l'éducation, l'âge, le sexe, etc.),

- doit être calculé et rapporté pour tous les sous-groupes principaux de population pour lesquels des estimations du GATS seront calculées et rapportées. Cela peut être fait dans un tableau avec une liste de rapports des sous-groupes et les valeurs associées de

$Meff_w$

4. **Interprétation de $Meff_w$** — une interprétation correcte de la valeur de $Meff_w$ calculée pour un domaine de rapports est la suivante :

- « La variation de la pondération de l'échantillon a augmenté la variance de toutes les estimations (du domaine de rapports) par un facteur de (). »
- Si $Meff_w < 2.00$, les efforts pour réduire la variation de pondération sont inutiles.

**Il est préférable d'avoir
un $Meff_w < 2.00$.**

$Meff_w$

5. **Exemple** — Supposons que pour les estimations en milieu rural $Meff_w = 1.62$

« La variation de la pondération de l'échantillon a augmenté la variance de toutes les estimations des répondants ruraux par un facteur de 1,62. »

- $Meff_w = 1.62$ est considéré comme suffisamment faible pour ces estimations.

5.2.3 Effet global de la conception sur la précision des estimations de l'enquête et sur l'homogénéité intragroupe dans les Unités Primaires d'Echantillonnage, PSUs des estimations clés correspondantes de l'enquête

Le **Manuel de conception de l'échantillon du GATS** recommande une conception où les grappes des répondants à l'enquête sont sélectionnées et le **Manuel de pondération de l'échantillon du GATS** décrit comment la pondération de l'échantillon doit être calculée. Une fois que les données du questionnaire ont été nettoyées, et la pondération finale de l'échantillon a été attribuée, les données de l'échantillon du GATS sont prêtes à être examinées avant que les résultats de l'analyse ne soient rapportés.

Le contexte général et les instructions pour le calcul de l'effet global de la conception sur une estimation de l'enquête et l'homogénéité au sein des PSU sont présentés ci-dessous dans cette section. Pour plus de détails sur ces sujets, référez-vous à l'**annexe D.5**.

1. **Contexte** — La variance d'une estimation de l'enquête ($\hat{\theta}$) d'une conception d'échantillon complexe, divisée par la variance d'une estimation comparable basée sur un échantillon aléatoire simple de même taille est appelée l'« effet de la conception » global de $\hat{\theta}$, ou $Deff(\hat{\theta})$. Il y a deux composantes multiplicatives de $Deff(\hat{\theta})$: l'effet de conception dû à l'échantillonnage en grappes, $Deff_{cs}(\hat{\theta})$ et l'effet multiplicateur de la variable de pondération de l'échantillon $Meff_w$ (voir la **section 5.2.2**). Par ailleurs $Deff_{cs}(\hat{\theta})$ est directement liée au taux d'homogénéité au sein des PSUs, ρ_p , et à la taille moyenne des grappes de l'échantillon.
2. **Estimation de $Deff(\hat{\theta})$ et R^b** — Une estimation de $Deff(\hat{\theta})$ peut être obtenue directement avec des ensembles de logiciels d'analyse d'enquête (par exemple, SUDAAN), ou elle peut être calculée à partir de l'estimation et de sa variance lorsque l'estimation est un rapport ou un taux. Puisque $Deff(\hat{\theta})$ dépend de $Deff_{cs}(\hat{\theta})$, R^b , et de la taille moyenne des grappes de l'échantillon, on peut obtenir une estimation de R^b en la résolvant et en utilisant les estimations fournies de $Deff(\hat{\theta})$ et ces autres mesures.
 - Des formules détaillées pour estimer $Deff(\hat{\theta})$, quand le logiciel d'analyse de l'enquête ne les fournit pas, et pour estimer R^b en toutes circonstances, se trouvent dans les équations (1) et (2) à l'annexe D.5. Les estimations obtenues sont notées $d_{eff}(\hat{\theta})$ et \hat{R}^b , respectivement. Nous utilisons $d_{eff_{cs}}(\hat{\theta})$ pour désigner l'estimation de $Deff_{cs}(\hat{\theta})$.
 - Puisqu'il y aura de nombreuses valeurs de $d_{eff}(\hat{\theta})$ et \hat{R}^b , résumez-les en rapportant les valeurs médianes, minimales et maximales.
3. **Rapports de $Deff(\hat{\theta})$ et de R^b** — Les estimations de $Deff(\hat{\theta})$ et de R^b doivent être rapportées pour toutes les mesures clés de l'étude (voir l'**annexe D.5** pour les définitions).

4. **Interprétation de $Deff(\hat{\theta})$ et de R^b** — Une interprétation correcte des valeurs estimées de $deff(\hat{\theta})$ et de R^b , est la suivante :

- « La variance de l'estimation de l'enquête (de la caractéristique de la population, θ), étant donné la conception de l'échantillon de la GATS, est ($deff(\hat{\theta})$) fois supérieure à la variance produite si un échantillonnage aléatoire simple avait produit le même nombre de répondants. »
- Il est préférable que l'effet de la conception global pour une estimation, ou $Deff(\hat{\theta})$, soit inférieur à 2,50.
- La somme de la variation totale de la mesure de l'étude (associée à θ) qui n'a pas été comptabilisée dans les PSU est $\hat{\rho}_\theta$.
- Est rarement $\hat{\rho}_\theta > 0.15$ ou négatif, est plus généralement $\hat{\rho}_\theta < 0.05$ ou même $\hat{\rho}_\theta < 0.01$.

Est rarement $\hat{\rho}_\theta > 0.15$, est plus généralement $\hat{\rho}_\theta < 0.05$ ou même $\hat{\rho}_\theta < 0.01$.

5. **Exemple** — L'analyse pour estimer le taux de prévalence du tabagisme, p , produit les résultats suivants à partir d'un échantillon dont la taille moyenne des grappes de l'échantillon était de 20 : $deff(\hat{\theta}) = 1.92$, $Meff_w = 1.36$.

« La variance de l'estimation de l'enquête du taux de prévalence du tabagisme actuel, étant donné la conception de l'échantillon du GATS, est 1,92 fois supérieure à la variance produite si un échantillonnage aléatoire simple avait produit le même nombre de répondants. »

- La valeur calculée de $\hat{\rho}_\theta = 0.022$ est typique.

5.2.4 Marge d'erreur pour les estimations clés de l'enquête

Il y a plusieurs façons de rendre compte de la précision statistique des estimations de l'enquête. Une marge d'erreur d'une estimation (MOE) est l'une d'entre elles. Le GATS conseille de rapporter la marge d'erreur estimée (M_{OE}) ainsi que les estimations pour les mesures clés de l'enquête. Le **Manuel de conception de l'échantillon du GATS** décrit les trois principales caractéristiques des échantillons du GATS qui vont influencer la qualité statistique des estimations et des résultats tirés des données. Ces caractéristiques sont l'utilisation de la stratification, la sélection de grappes de membres de la population cible, et la sélection de membres de la population avec des probabilités inégales (nécessitant donc l'utilisation de facteurs de pondération dans l'analyse). Divers logiciels d'analyse d'enquête peuvent être utilisés pour tenir compte de ces caractéristiques dans l'analyse des données du GATS.

Des informations générales et des instructions de calcul sont présentées dans cette section. Pour plus de détails sur ces sujets, référez-vous à l'**annexe D.6**.

1. **La marge d'erreur est spécifique à la mesure** — Chaque estimation obtenue à partir des données du GATS a sa propre marge d'erreur. La [ME] (MOE) est par définition la moitié de la largeur attendue d'un intervalle de confiance d'une estimation d'une mesure d'enquête clé. Une autre façon d'interpréter la [ME] (MOE) est d'établir à quel point l'estimation est susceptible d'être proche de la mesure véritable de l'enquête dans la population.

2. **Mesures clés de l'enquête** — Les estimations de changement d'une phase à une autre ainsi que les estimations spécifiques des phases du GATS seront calculées pour divers taux de comportement/d'usage du tabac. Ces estimations sont susceptibles d'être signalées pour la population dans son ensemble, ou en suivant des sous-groupes étudiés, par exemple par sexe, milieu urbain/rural, et catégories d'âge. Des estimations régionales peuvent être signalées lorsque les tailles globales des échantillons régionaux sont au moins de 8 000.

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.
3. **Estimation de la MOE** — Le logiciel d'analyse de l'enquête ne calcule généralement pas d'estimation de la MOE, mais les informations nécessaires pour la calculer sont généralement disponibles. Trois choses sont nécessaires pour calculer et interpréter la MOE : (i) l'estimation de la mesure de l'enquête, (ii) son erreur type estimée (ou variance), et (iii) une mesure spécifiée associée à la confiance statistique souhaitée de la valeur de la MOE [ME].

Le niveau de confiance est généralement basé sur une valeur (Z) de la distribution normale standard. Par exemple, pour un niveau de confiance de 95 %, nous pouvons utiliser $Z = 1,96$.

La MOE correspond à la mesure de la confiance désirée multipliée par l'erreur type de l'estimation. La formule réelle de calcul est l'équation (1) de **l'annexe D.6**.

4. **Rapports MOE** — Il est recommandé de calculer la MOE pour toutes les estimations liées aux mesures clés de l'enquête. Cela inclut les estimations nationales globales de ces mesures ainsi que des estimations de ces mesures pour tous les sous-groupes de rapports importants (par exemple, par sexe, milieu urbain/rural, région et âge).

Une estimation de la MOE [ME] doit être calculée pour toutes les estimations associées à des mesures clés de l'enquête.

- Les estimations de l'enquête clés et leurs valeurs associées de MOE doivent être présentées ensemble. Des informations sur les rapports des estimations des erreurs d'échantillonnage et un exemple de matrices de tableau avec des indicateurs clés du GATS et des valeurs recommandées pour les rapports sont fournis dans **l'annexe D.7**.
5. **Interprétation** — Lorsqu'il est pris avec la valeur d'une estimation de l'enquête, la MOE indique à quel point l'estimation est susceptible d'être proche de la valeur réelle dans la population.

Lorsque vous utilisez $Z = 1,96$ pour calculer la MOE, l'estimation de l'enquête avec sa valeur de la MOE peuvent être interprétées comme suit :

« Nous sommes sûrs à 95 % que l'estimation (VALEUR DE L'ESTIMATION) se trouve à (VALEUR DE SA MOE) de la valeur de la population correspondante. »

6. **Exemple** — Supposons que la valeur rapportée d'une estimation de l'enquête du GATS soit de 43,4 %, avec une erreur type de 1,5611 % qui a été calculée conformément à la conception de l'échantillon de ce pays.

La valeur de MOE pour cette estimation sera de $3.1\% = (1.96)(1.5611\%)$

Interprétation :

« Nous sommes confiants à 95 % que l'estimation, 43,4 %, est à 3,1 % de la valeur de la population correspondante. »

5.3 Mesures de qualité : Couverture, non-réponse, et autres erreurs non dues à l'échantillonnage

5.3.1 Taux de couverture de la base des ménages

Comme il est décrit dans le *Manuel de conception de l'échantillon du GATS*, l'avant-dernière étape de l'échantillonnage du GATS consiste à sélectionner un échantillon de ménage parmi ceux figurant dans les UPEs (PSUs) ou les segments. La qualité de l'échantillon de ménage et, finalement, des estimations du GATS dépendent donc de l'exhaustivité de ces listes des ménages, qui sont préparées de diverses façons.

Le contexte général et des instructions pour le calcul d'un taux de couverture de la base (FCR, soit frame coverage rate) des foyers sont présentés ci-dessous dans cette section. Pour plus de détails sur ces sujets, référez-vous à l'*annexe D.8*.

1. **Contexte** — Un élément clé du processus d'échantillonnage dans une enquête auprès des ménages est le dénombrement des ménages au cours de la phase d'échantillonnage lorsque les adresses des ménages sont choisies. Différentes approches de dénombrement des ménages peuvent être utilisées (par exemple, établissement d'une liste traditionnelle des ménages par des agents recenseurs sur le terrain, listes électorales modifiées). Le résultat final ordinaire est un sous-dénombrement net des ménages et donc une couverture incomplète dans l'échantillon final des ménages. Comme l'effet de biais du sous-dénombrement dépend des facteurs suivants : (i) la différence cumulée entre les mesures clés de l'étude pour les ménages couverts par la base de sondage et ceux non couverts, et (ii) l'ampleur de la non-couverture (ou 100 % moins le taux de couverture des ménages) dans l'échantillon, il est utile d'estimer le taux de couverture de la base de sondage pour les ménages.
 - Un taux de couverture d'une base de sondage (FCR) est défini comme le nombre de ménages pris en compte dans la base de sondage des ménages divisé par le nombre de ménages dans la population.
 - Bien qu'elles soient utiles pour décrire la couverture de la base de sondage de l'échantillon des ménages, souvent, on ne dispose pas de données provenant d'un recensement récent du nombre total de foyers dans la population. Si disponible, le FCR doit être calculé, mais ce n'est pas un élément de données critique pour l'examen officiel de la GATS des mesures de la qualité des statistiques.

2. **Estimation du FCR** — Le numérateur du FCR peut être estimé en utilisant la somme des pré-ajustements de la pondération (de base) sur tous les ménages sélectionnés. Le dénominateur doit être obtenu auprès d'une source à jour et hautement réputée (par exemple, un recensement récent ou une autre enquête récente dont l'échantillon est d'une taille supérieure au GATS).
3. **Rapports du FCR** — Le FCR estimé doit être signalé dans le *Rapport national* des résultats avec la description de la conception de l'échantillon ou avec une évaluation séparée de la qualité des données.
4. **Interprétation du FCR** — Le FCR indique l'ampleur de la couverture de la base de sondage des ménages, mais pas la taille du biais de couverture, puisque cette dernière dépend de la première plus la différence cumulée (dans les mesures clés de l'enquête) entre les ménages couverts et non couverts. Un taux de couverture des ménages supérieur à 95 % serait considéré comme acceptable.
5. **Exemple** — La meilleure estimation extérieure disponible du nombre de ménages dans la population cible d'un pays du GATS est de 36 436 103, alors que la somme pondérée des 13 000 ménages sélectionnés de l'échantillon pour le GATS est de 34 788 543. Le taux de couverture résultant de $0.955 = \frac{34,788,543}{36,436,103}$ serait considéré comme acceptable.

Un taux de couverture des ménages supérieur à 95 % serait considéré comme acceptable.

5.3.2 Modèles de taux de rejet des répondants

Certains entretiens d'enquête ne sont pas entièrement terminés. Le répondant peut trouver la participation à l'entretien pesante et perdre patience avec l'enquêteur, ou le répondant peut avoir besoin de mettre fin à l'entretien plus tôt en raison de restrictions légitimes de son emploi du temps et ne pas être en mesure ou désireux de reprogrammer un moment pour terminer l'entretien. Dans les deux cas, un entretien partiellement terminé est un indicateur du désengagement du répondant, qui peut être vu comme un reflet de la qualité des efforts de l'enquêteur ou des attitudes des répondants à l'égard du sujet de l'enquête et donc finalement de la qualité des données.

- **Source(s) des données** : Le fichier de données contenant les codes de décisions finales des ménages /personnes doit être utilisé pour ces calculs (voir l'**annexe D.2**). Un code de décision de 11 à l'IQ indique que le répondant a rempli au moins les sections A à D de l'entretien individuel du GATS. Un code de décision de 12 à l'IQ indique que le répondant a demandé ou exigé que l'entretien soit interrompu avant que les sections A à D puissent être terminées.
- **Méthode/calcul** : Définissez un « répondant » à l'enquête comme toute personne sélectionnée ayant reçu un code de décision finale à l'IQ de 11 ou 12. Définissez également un taux de rejet (*COR*) comme :

$$COR = \frac{R_c}{R} = \frac{\# \text{ of GATS respondents who only partially completed the interview}}{\# \text{ of GATS respondents}}$$

ou

Disposition code 12

=

Disposition codes (11 + 12)

- **Utilisations** : Les valeurs de COR doivent être calculées directement pour l'échantillon dans une phase finale de l'assurance de la qualité, après la collecte des données. Par ailleurs, les FS doivent continuellement s'efforcer d'identifier les FI inefficaces et de prendre des mesures correctives pendant toute la collecte de données (en utilisant les codes de résultat de terrain correspondant pour le calcul). Cela implique le calcul du COR par enquêteur ou équipe enquêtrice, et l'identification des enquêteurs/équipes d'enquêteurs qui ont des taux exceptionnellement élevés (par exemple, des taux supérieurs à la moyenne plus deux écarts-types des taux parmi tous les enquêteurs/équipes). Les valeurs de COR peuvent également être calculées en fonction de : (1) la semaine de collecte de données pendant laquelle l'entretien a eu lieu, (2) l'âge du répondant, (3) le sexe du répondant et (4) milieu urbain/rural du répondant.
- **Interprétation** : En général, plus la valeur de COR est basse, mieux cela vaut. Une baisse de COR sur l'ensemble de la durée de collecte des données pourrait indiquer que les répondants étaient moins motivés pour fournir des données utiles au GATS, vers la fin de collecte de données. Bien que les taux de rejet soient des mesures utiles du rendement de collecte de données et de la qualité globale de l'enquête, leur calcul pour l'examen officiel du GATS des mesures de qualité statistique n'est pas important.

5.3.3 Modèles de taux de réponse des ménages par strates d'échantillonnage de la première phase

Le taux de réponse des foyers (HRR) est calculé comme le rapport du nombre pondéré de foyers pour lesquels une liste complète de données a été fournie sur le nombre pondéré de foyers de la zone (voir la formule 19, *Manuel de pondération de l'échantillon du GATS*). Les ménages fournissant une liste de données partielle sont traités comme des refus ou des interruptions pour le taux de réponse des ménages, et ne sont pas inclus dans le numérateur du calcul du taux de réponse des ménages. Le *Questionnaire Ménage* est considéré comme complet s'il a un code de décision finale des ménages de 1 (voir l'*annexe D.2*). En utilisant seulement les ménages de la zone de l'enquête, le taux de réponse des ménages est calculé comme suit :

$$\text{Household - Level RR} = \frac{[1]}{[1] + [3] + [4] + [5] + [6] + [9]} \times 100$$

où

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

Notez que le code 2 de décision finale HH est exclu à la fois du numérateur et du dénominateur du taux de réponse des ménages, puisque ces ménages sont considérés comme n'étant pas admissibles. Ce taux de réponse est conforme à RR1 tel que défini par l'Association américaine de recherche sur l'opinion publique (AAPOR) et documenté dans le rapport disponible à l'adresse : http://www.aapor.org/uploads/standarddefs_4.pdf.

Une documentation complète de la façon dont les taux de réponse des ménages sont calculés doit être fournie par les pays. Les taux de réponse des ménages doivent être présentés pour tout le pays par la première étape d'échantillonnage des strates (c.-à-d., milieu urbain/rural, région), si la taille des échantillons est suffisante pour une estimation régionale. Un fractionnement supplémentaire, si nécessaire, doit être effectué en utilisant la classification croisée du milieu urbain/rural et régional.

5.3.4 Modèles des taux de réponse des personnes parmi les variables utilisées pour les ajustements en fonction des non-réponses

Un taux de réponse au niveau d'une personne correspond au rapport du nombre pondéré d'entretiens du GATS partiellement ou totalement terminés sur le nombre pondéré de personnes admissibles échantillonnées choisies dans la liste de tous les membres du foyer. Plus précisément, les codes de décisions finales au niveau des personnes sont les suivants :

- 11 = *Questionnaire Individuel Rempli*
- 12 = Entretien incomplet
- 13 = L'individu sélectionné s'est avéré ultérieurement inéligible pour le GATS
- 14 = Le répondant sélectionné n'est pas chez lui
- 15 = Refus du répondant sélectionné
- 16 = Répondant choisi incompetent
- 17 = Autre non-réponse individuelle.

Une façon de calculer un taux de réponse individuelle est la suivante :

$$\text{Person - Level RR} = \frac{[11]}{[11] + [12] + [14] + [15] + [16] + [17]} \times 100$$

Le taux de réponse individuelle ci-dessus présume que toutes les personnes sélectionnées dont l'admissibilité est inconnue (par exemple, code de décision finale 14) sont admissibles à la GATS. Cela peut conduire à sous-estimer les taux de réponse individuelle, si les enquêteurs choisissent souvent des répondants qui se trouvent être inéligibles pour l'enquête une fois 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 GATS en utilisant les fréquences de codes de décision pondérées comme :

$$e_{\text{person-level}} = \frac{[11] + [12] + [15] + [16] + [17]}{[11] + [12] + [13] + [15] + [16] + [17]}$$

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 individuelle 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 (2004) :

$$\text{Person - Level RR} = \frac{[11]}{[11] + [12] + [e^{\text{person-level}}[14]] + [15] + [16] + [17]} \times 100$$

Ce taux de réponse est conforme à RR3 tel que défini par l'AAPOR et documenté ci-dessus.

Une documentation complète de la façon dont les taux de réponse sont calculés doit être fournie par les pays.

Les taux de réponse au niveau des personnes doivent être présentés pour l'ensemble du pays et par sous-zones régionales lorsque la taille des échantillons est suffisante pour l'estimation régionale. Un découpage supplémentaire à l'intérieur de la région (si nécessaire) doit être effectué pour les variables utilisées pour définir les ajustements pour la non-réponse. Il peut inclure le milieu urbain/rural et la liste des âges déclarés (15 à 24, 25 à 34, 35 à 44, 45 à 54, plus de 55), le sexe (masculin, féminin) et le statut tabagique actuel (fumeur, non-fumeur).

Le *Manuel des Superviseurs du GATS* décrit les responsabilités du FS [ST] pour l'examen et la mise en œuvre de solutions dans le cas de faibles taux de réponse de l'équipe de l'Enquêteur.

5.3.5 Modèles des taux de refus des personnes parmi les variables utilisées pour les ajustements pour la non-réponse

Les taux de refus sont calculés comme le rapport du nombre pondéré de répondants admissibles sélectionnés qui ont refusé de participer au *Questionnaire Individuel* du GATS (code de décision finale du *Questionnaire Individuel* 15) sur le nombre total pondéré de questionnaires individuels du GATS remplis (11) et incomplets (12), plus les non-répondants dont l'éligibilité n'est pas connue (14), ceux qui ont refusé (15), les répondants incompetents (16) et toute autre personne non-répondante (17). En utilisant les codes de décision au niveau individuel, le calcul est le suivant :

$$\text{Person - Level Refusal Rate} = \frac{[15]}{[11] + [12] + [14] + [15] + [16] + [17]} \times 100$$

Ce taux de refus est conforme à REF1 tel que défini par l'Association américaine de recherche sur l'opinion publique (AAPOR).

Les taux de refus individuels doivent être calculés directement pour l'échantillon dans une phase finale de l'assurance de la qualité, après le recueil des données. Par ailleurs, les FSs doivent continuellement s'efforcer d'identifier les FI inefficaces et de prendre des mesures correctives pendant tout le recueil des données. Cela implique le calcul des taux de refus par enquêteur ou équipe enquêtrice, et l'identification des enquêteurs/équipes d'enquêteurs qui ont des taux exceptionnellement élevés (par exemple, des taux supérieurs à la moyenne plus deux écarts-types des taux parmi tous les enquêteurs/équipes). Ces taux de refus peuvent aussi être calculés suivant : (1) la semaine de collecte de données pendant laquelle

l'entretien a eu lieu, (2) l'âge du répondant, (3) le sexe du répondant et (4) l'emplacement du répondant en milieu urbain/rural.

Une documentation complète de la façon dont les taux de refus sont calculés doit être fournie par les pays. Les taux de refus doivent être présentés pour l'ensemble du pays et par sous-zones régionales lorsque la taille des échantillons est suffisante pour l'estimation régionale. Un découpage supplémentaire à l'intérieur de la région (le cas échéant) doit être effectué pour les variables utilisées pour définir les ajustements de non-réponse. Il peut inclure le milieu urbain/rural et la liste des âges déclarés (15 à 24, 25 à 34, 35 à 44, 45 à 54, plus de 55), le sexe (masculin, féminin) et le statut tabagique actuel (fumeur, non-fumeur).

5.3.6 Taux de non-réponse partielle pour les variables indicatrices de la Fiche d'Information Le taux de non-réponse du GATS (par exemple, le pourcentage de tous les répondants qui ne parviennent pas à fournir des données utiles pour un élément donné du questionnaire de l'entretien sur l'ensemble des répondants censés répondre à la question) doit être calculé et documenté après le traitement des données de l'enquête extraites du questionnaire qui sont associées à tous les indicateurs qui seront inclus dans la Fiche d'Information du GATS spécifique à chaque pays (voir **GATS Indicator Definitions (Définitions des Indicateurs du GATS** pour plus d'informations). Une liste de toutes les variables utilisées pour créer les indicateurs de base de la Fiche d'Information figure dans l'**Annexe C.1**. Les taux de non-réponse (NRRI) correspondent au rapport du nombre de répondants qui n'ont pas fourni de réponse dans le champ de validité (**MISSING^x for item x**) [**MANQUANTE^x** pour l'élément x], sur le nombre total de répondants au niveau des unités (I) moins le nombre de répondants avec un enchaînement valide pour l'élément x (**V^x**) :

$$NRRI^x = \frac{MISSING^x}{(I - V^x)} \times 100$$

Le nombre total des non-répondants au niveau des unités de x (**MANQUANTE^x**) sera obtenu à partir d'une fréquence non pondérée des répondants avec des données manquantes pour l'élément x après un nettoyage approprié pour assurer que des enchaînements corrects ont été suivis. Le nombre total de répondants au niveau des unités sera obtenu à partir de la fréquence totale non pondérée des répondants de sexe masculin ou féminin à la question A01 (une variable n'ayant pas de champs vides prévus). Le nombre total de répondants avec un enchaînement valide pour l'élément x peut être obtenu comme étant la fréquence de l'élément x avec une réponse .S (voir la **section 5.1.3**). De nombreux indicateurs de la Fiche d'Information sont des mesures composites obtenues à partir de réponses à plusieurs questions. La non-réponse à tout élément utilisé pour obtenir une mesure composite entraînera une non-réponse de la mesure composite. Les taux de non-réponses inférieurs à 5 % sont considérés comme faibles. À l'instar des mesures de qualité des paragraphes 5.3.2 et 5.3.5 ci-dessus, les FS doivent s'efforcer de cibler continuellement les questions administrées par des enquêteurs ou des équipes d'enquêteurs obtenant un taux excessif de non-réponses, comme un moyen d'améliorer la qualité de la collecte des données alors que les données sont collectées sur le terrain.

5.4 Examen officiel de la qualité statistique

Un protocole spécifique a été développé pour effectuer un examen officiel des mesures de qualité statistique dans chaque pays du GATS, après le traitement des données d'enquête et le calcul de la pondération de l'échantillon, mais avant la diffusion des résultats. Cet examen a pour objectifs de vérifier

que la pondération de l'échantillon a été calculée conformément aux spécifications du **Manuel de pondération du GATS**, d'examiner les diverses mesures indiquant la qualité supérieure des données d'enquête, et d'assurer que les principaux aspects de l'échantillonnage et de l'analyse sont documentés de manière exhaustive. Une assistance technique pour la préparation de cet examen sera disponible.

Les mesures spécifiques de qualité et la documentation qu'il faut pour effectuer ces examens sont décrits ci-après.

Documentation de la pondération de l'échantillon :

Chaque étape de la pondération des données et du processus d'ajustement doit être entièrement décrite dans un document préparé par les statisticiens du pays directement responsable du calcul et de l'ajustement de la pondération statistique de l'échantillon du GATS. Le document doit comprendre un énoncé décrivant en détail le processus de pondération et de correction, y compris des exemples illustrant précisément comment toutes les composantes de la pondération finale ont été calculées pour 5 à 10 répondants de l'échantillon réel. Plus spécifiquement, chacun des éléments suivants doit être préparé pour l'examen par les personnes chargées de l'effectuer¹:

- Résumez la conception de l'échantillon de l'enquête par étape et décrivez comment les probabilités de sélection pour chaque étape et la pondération de base globale ont été calculées.
- Décrivez en détail comment les stratégies d'ajustement pour la non-réponse et d'étalonnage ont été établies, en indiquant précisément : (i) comment les classes de pondération ont été définies pour les corrections en fonction de non-réponses des foyers et au niveau des personnes, (ii) quelles cellules d'ajustement ont été formées pour étalonner la pondération, et (iii) par rapport à quel dénombrement de population la pondération a été étalonnée. Reportez-vous au **Manuel de pondération de l'échantillon du GATS** pour vous guider pour cette description.
- Choisissez 5 à 10 répondants réels pour montrer le processus de calcul des pondérations qui s'est appliqué dans leur cas. Les répondants doivent être sélectionnés dans des strates, PSU et groupes de sexes différents. Si des éléments de pondération basés sur intervalles semi-ouverts ont été utilisés, un répondant d'un de ces groupes géographiques doit être inclus dans l'un des 5 à 10 exemples. À l'issue de cette étape, le calcul de la pondération finale ajustée doit être affiché pour chaque exemple. Plus précisément, pour chacun de ces répondants :
 - Identifiez les composantes de chaque probabilité conditionnelle et inconditionnelle (à partir de chaque étape de l'échantillonnage) qui a été utilisée pour produire la PONDÉRATION DE BASE.
 - Présentez et décrivez complètement chaque composante du calcul d'AJUSTEMENT POUR LA NON-RÉPONSE des ménages et au niveau des personnes. À l'issue de cette étape, le poids de base doit être ajustée pour la non-réponse des ménages et au niveau des personnes.

¹ Notez que puisque de nombreux articles sur cette liste doivent également être inclus dans la documentation pour chaque *Rapport national* de la GATS, les ébauches de texte de cette documentation (mais contenant les versions finales des résultats numériques requis) peuvent être envoyées aux personnes qui effectuent l'examen officiel de la qualité des statistiques pour le pays.

- Présentez et décrivez complètement chaque composante de la méthode D'ÉTALONNAGE D'AJUSTEMENT DU POIDS pour forcer la distribution finale pondérée de l'échantillon à correspondre à une population d'étalonnage choisie.
- Présentez les résultats de calcul suivants basés sur les formules données dans **les sections 5.2 et 5.3** de ce manuel :
 - Ajustement d'étalonnage poste-stratification pour chaque cellule d'ajustement utilisée dans l'étalonnage (**5.2.1**)
 - Effet multiplicatif de la variable de pondération de l'échantillon (*Meff*) par tous les sous-groupes de population pour lesquels les estimations d'enquête seront communiquées dans le *Rapport national* (**5.2.2**)
 - Les mesures individuelles et résumées (c.-à-d., minimum, maximum et médiane) de l'effet du plan de sondage global et la mesure de l'homogénéité au sein d'une UPE (PSU) pour toutes les estimations de l'enquête présentées sur La Fiche d'Information *du Pays* (**5.2.3**)
 - Les mesures individuelles et résumées (c.-à-d., minimum, maximum et médiane) de la Marge d'Erreur (*MOE*), ou d'une autre mesure comparable de la précision, pour toutes les estimations de l'enquête présentées dans La Fiche d'Information *du Pays* (**5.2.4**)
 - Les mesures individuelles et résumées (c.-à-d., minimum, maximum et médiane) des taux de réponse des ménages suivant les cellules d'ajustement utilisées pour produire l'ajustement pour la non-réponse au niveau des ménages (**5.3.3**)
 - Taux de réponse au niveau des personnes parmi les cellules d'ajustement utilisées pour les ajustements pour la non-réponse au niveau des personnes (**5.3.4**)
 - Taux de refus au niveau des personnes parmi les cellules d'ajustement utilisées pour les ajustements pour la non-réponse au niveau des personnes (**5.3.5**)
 - Mesures individuelles et résumées (c.-à-d., minimum, maximum et médiane) des taux de non-réponse pour tous les éléments de l'enquête utilisés pour produire le taux de tabagisme de la Fiche d'Information du Pays (**5.3.6**)

Le **Manuel de pondération du GATS (Chapitre 4)** présente plusieurs méthodes suggérées pour mieux assurer la qualité des pondérations de l'échantillon. Les contrôles suivants seront réalisés par les personnes menant les examens de qualité statistique, et donc les données pour les compléter seront nécessaires :

- La taille moyenne de la pondération de base, divisée par la taille moyenne de la pondération ajustée pour la non-réponse, est approximativement égale au taux final de réponse global au niveau des personnes.
- La répartition pondérée de l'échantillon (en utilisant la pondération finale ajustée) dans toutes les cellules d'ajustement d'étalonnage doit correspondre précisément à la répartition correspondant au dénombrement de la population extérieure par rapport à laquelle les pondérations ont été étalonnées.
- La somme des pondérations finales ajustées *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, alors cette somme de pondérations doit être le dénombrement de la population cible à partir du dernier recensement.

- La plupart des ajustements poste-stratification (par cellule d'ajustement) doivent être légèrement supérieures ou inférieures à 1.
- Le *Meff* est $\leq 2,00$ pour chaque sous-groupe de signalement pour le *Rapport national*.

5.5 Création du fichier des données analytiques

Après que la pondération de l'échantillon et tous les contrôles de qualité ont été achevés, un nouveau fichier doit être créé contenant seulement les cas avec le code de décision finale individuelle de 11 (*Questionnaire Individuel Rempli*). Seuls les cas ayant un code de décision finale individuelle de 11 seront considérés comme « répondants » au *Questionnaire Individuel* du GATS.

Notez qu'il n'est pas nécessaire de créer l'ensemble de données analytique final avant de procéder aux contrôles d'assurance qualité en utilisant la Marge d'Erreur, MOE, l'effet de conception ou les estimations clés de l'enquête calculés par les procédures SAS ou SPSS pour des échantillons complexes, parce que les cas avec des codes d'IQ de décision finale autres que 11 n'auront pas de valeurs pour la variable de pondération (c.-à-d., ils auront des valeurs manquantes) et seront omis automatiquement par le logiciel comme un défaut, sauf indication contraire.

Ce nouveau fichier est appelé le fichier de données analytiques, et doit être utilisé pour mener des analyses de données afin de produire des estimations de l'enquête (c.-à-d., les indicateurs), y compris les estimations qui figureront sur la Fiche d'Information du Pays et le *Rapport du Pays*.

6. Exigences de Données et de documentation

Ce chapitre décrit les fichiers de données et la documentation qui seront envoyés par les pays et archivés au Centre de coordination des données du GATS.

Fichiers de données et documents demandés du pays :

1. Fichiers MDB finaux programmés : CMSDB.mdb, Survey0.mdb et Survey1.mdb.
2. Fichier maître SDF créé par le logiciel d'agrégation à partir de CMSDB.mdb, Survey0.mdb et Survey1.mdb.
3. Fichier de base de données maître, généré en utilisant l'utilitaire de fusion de fichier maître, contenant les fichiers plats transposés pour le tableau des réponses avec les réponses valides pour les *Questionnaires Ménage* et *Individuel* et le fichier maître de sélection de l'échantillon (numéro de cas et identifiants de l'échantillon tel que décrit dans l'annexe B.4).
4. Composantes et documentation sur la pondération de l'échantillon (se référer à la **section 5.4**).
5. Ensemble de données pondéré : les fichiers finaux, plats, nettoyés et pondérés, dictionnaires de données et spécifications des pondérations.

7. Bibliographie

- Biemer, Paul P. and Lars E. Lyberg. 2003. Introduction to Survey Quality. John Wiley and Sons, Inc., Hoboken, NJ.
- Encyclopedia of Survey Research Methods*, (2007). Lavrakas, P.J. (Ed.), Sage Publishing, Inc. Los Angeles, CA.
- Gabler, S., Haeder, S., and Lahiri, P. (1999). "A model based justification of Kish's formula for design effects for weighting and clustering," *Survey Methodology*, 25 (1): 105-106.
- Global Adult Tobacco Survey Collaborative Group. *Global Adult Tobacco Survey (GATS): Core Questionnaire Programming Specifications, Version 2.0*. Atlanta, GA: Centers for Disease Control and Prevention, 2010.
- 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): Data Management Implementation Plan, Version 2.0*. Atlanta, GA: Centers for Disease Control and Prevention, 2010.
- Global Adult Tobacco Survey Collaborative Group. *Enquête mondiale sur le tabagisme chez l'adulte (GATS, Global Adult Tobacco Survey) Data Management Training Guide, Version 1.0*. Atlanta, GA: Centers for Disease Control and Prevention, 2010.
- Global Adult Tobacco Survey Collaborative Group. *Global Adult Tobacco Survey (GATS): Data Management Implementation Plan, Version 2.0*. Atlanta, GA: Centers for Disease Control and Prevention, 2010.
- Global Adult Tobacco Survey Collaborative Group. *Global Adult Tobacco Survey (GATS): Data Management Training Guide, Version 1.0*. 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): Programmer's Guide to General Survey System, Version 2.0*. Atlanta, GA: Centers for Disease Control and Prevention, 2010.
- Global Adult Tobacco Survey Collaborative Group. *Global Adult Tobacco Survey (GATS): Question by Question Specifications, Version 2.0*. Atlanta, GA: Centers for Disease Control and Prevention, 2010.
- 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): Sample Weights Manual, Version 2.0*. Atlanta, GA: Centers for Disease Control and Prevention, 2010.
- Référence interne pour les valeurs manquantes spéciales avec SAS :
<http://support.sas.com/documentation/cdl/en/lrcon/61722/HTML/default/a000992455.htm>
- Kish, L. (1965) *Survey Sampling*, Wiley and Sons, New York.
- Kalton, G. and Flores-Cervantes, I. (2003). Weighting Methods. *Journal of Official Statistics*, 19(2), 81-97.
- 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.

Annexe A : Glossaire

Questionnaire Principal : Les questions de base normalisées utilisées dans le GATS. Les pays participants adaptent le questionnaire principal à leurs besoins spécifiques en suivant des directives précises.

General Survey System (GSS) : Le logiciel d'enquête qui est chargé sur l'iPAQ pour mener l'enquête GATS.

Spécifications de programmation du GSS : En plus du questionnaire papier qu'un pays utilise pour créer leur version adaptée, les spécifications de programmation du GSS sont élaborées pour chaque version adaptée au pays et sont utilisées pour la programmation du questionnaire pour l'ordinateur de poche.

iPAQ : Le nom de l'appareil de poche de type PDA de Hewlett-Packard (HP) utilisé pour le GATS.

Fichier de cas de l'iPAQ : La liste des échantillons et les informations chargées sur les ordinateurs de poche iPAQ pour la collecte de données.

iPAQ Mastering ou Maîtrise [passage à] de l'iPAQ : Le processus de copie de la version du logiciel sur l'iPAQ, généralement effectuée sur tous les iPAQs en même temps avant le début de la formation des Enquêteurs/ Enquêteuses de terrain.

Fichier maître de sélection de l'échantillon : L'ensemble de données qui contient un numéro de cas pour chaque foyer dans la liste d'échantillonnage ainsi que les informations nécessaires pour calculer la pondération de l'échantillon et analyser les données complexes de l'enquête.

Fichiers MDB : La programmation du logiciel d'enquête du GSS se fait en utilisant des fichiers MDB (fichiers Microsoft Access).

Comité d'examen de l'échantillonnage [CEE] (SRC) : Un groupe composé d'experts internationaux en matière d'échantillonnage. Le SRC examine et approuve les conceptions d'échantillon, la pondération de l'échantillon, et les mesures d'assurance qualité des données des pays participants.

Fichiers SDF : Les fichiers du programme d'enquête du GSS qui sont chargés sur l'iPAQ. Cela inclut *survey0.sdf* (*questionnaire des foyers*), *survey1.sdf* (*questionnaire individuel*), et *cmsdb.sdf* (système de gestion de cas).

Assurance qualité : Un processus consistant en des activités systématiques visant à assurer, évaluer et confirmer la qualité des données recueillies lors d'une enquête.

Comité d'examen des questionnaires [CEQ] (QRC) : Un groupe composé d'experts internationaux dans la lutte antitabac et la conception de questionnaires. Le QRC examine et approuve tous les questionnaires de la GATS adaptés aux pays pour assurer la qualité, la normalisation et la comparabilité.

Prise de contrôle de l'iPAQ : Le processus de copie de la version du logiciel sur l'iPAQ, généralement effectuée sur tous les iPAQ en même temps avant le début de la formation des enquêteurs de terrain.

Annexe B : Période Pré-Collecte des données

B.1 Adaptation du questionnaire du GATS et processus d'examen

Adaptation du questionnaire : Les pays du GATS travailleront avec le CDC pour développer un questionnaire du GATS adapté spécifique du pays et qui comprendra les éléments suivants :

- **Adapter le questionnaire principal du GATS.** Les pays devront adapter le questionnaire principal en fonction de leur propre situation. Cela inclut modifier les listes de questions (par exemple, les types de tabac) et les choix de réponse (par exemple, la liste des marques de cigarettes) qui sont appropriés pour le pays.
- **Ajouter des questions facultatives.** Le GATS dispose d'une liste de suggestions de questions facultatives pour l'enquête que les pays peuvent choisir d'ajouter au questionnaire.
- **Ajouter des questions supplémentaires préparées par le pays.** Les pays peuvent inclure leurs propres questions de l'enquête qui ne sont pas incluses dans le questionnaire principal et des questions facultatives.
- **Supprimer des questions non pertinentes.** Les pays peuvent retirer des questions principales qui ne sont pas adaptées à leur situation (par exemple, pas de prévalence du tabac sans fumée). Une justification appropriée doit être fournie.

L'examen du questionnaire du GATS consiste en un processus en 10 étapes :

1. **Adaptation du projet de questionnaire spécifique au pays.** Les pays travailleront avec le CDC afin de créer un projet de questionnaire spécifique au pays à soumettre au QRC.
2. **Mise en forme.** Le CDC examineront le projet de questionnaire présenté et le formateront suivant le besoin (par exemple, modification des enchaînements d'instructions, surlignage pour l'examen du QRC).
3. **Examen spécialisé par le Comité d'examen des questionnaires.** Les membres du QRC examineront le questionnaire et fourniront des commentaires.
4. **Compilation des commentaires des experts.** Le CDC fait la compilation les commentaires du QRC et les envoie au pays concerné pour être revue.
5. **Communication (résolution).** Le QRC et le pays collaborent pour résoudre les questions en suspens (par courriel et/ou conférences téléphoniques).
6. **Finalisation du questionnaire adapté.** Une fois que tous les problèmes sont résolus, le questionnaire adapté au pays est finalisé et l'approbation officielle est donnée par le QRC.
7. **Avis d'acceptation par le pays.** Le pays accorde son acceptation définitive au questionnaire approuvé par le QRC.
8. **Traduction.** Les pays devront traduire le questionnaire approuvé du GATS dans la ou les langues appropriées et fournir également une rétro traduction du questionnaire en anglais. (Remarque : Cela peut se produire en même temps que l'étape 1 ; dans ce cas les pays traduisent d'abord le questionnaire principal et l'adaptent dans leur propre langue, puis le retraduisent [font la traduction inversée] en anglais pour le soumettre au QRC).
9. **Vérification des versions finales de la traduction et de la contre-traduction.** Le QRC examinera les versions traduites et rétro traduites afin de vérifier l'exactitude de la traduction (c.-

à-d. le sens des questions n'a pas changé). (Remarque : Cela fait partie du processus d'examen du QRC si la traduction est faite en même temps que l'étape 1 et qu'une version de traduction inversée est initialement soumise pour examen au QRC).

10. **Programmation.** Des spécifications de programmation pour le questionnaire adapté au pays sont créées et l'enquête spécifique à chaque pays est programmée sur l'ordinateur de poche.

B.2 Processus de programmation du questionnaire de la GATS

Les principales étapes et les recommandations pour les échéances de la programmation des questionnaires du GATS sont les suivantes :

1. Proposition du questionnaire et du pré-test du GATS adaptés au pays approuvés 6 à 8 semaines avant la formation du pré-test.
2. La programmation du questionnaire en anglais commence 6 semaines avant la formation du pré-test.
3. Le matériel iPAQ est livré et opérationnel 4 semaines avant la formation du pré-test.
4. Les questions de localisation sur l'iPAQ commencent 4 semaines avant la formation du pré-test.
5. Les traductions des menus et des messages du système GSS sont terminées 1 semaine avant la formation du prétest.
6. La traduction des textes du questionnaire (*questionnaires Ménage* et *l'individuel*) dans toutes les langues des pays d'accueil est faite et incluse dans l'iPAQ 2 semaines avant la formation du pré-test.
7. Le pays d'accueil approuve officiellement les *Questionnaires Ménage* et *Individuel* (HQ et IQ) 1 semaine avant la formation du prétest.
8. Le personnel informatique du pays d'accueil prend le contrôle et la propriété des questionnaires à ce stade.
9. Le contrôle de version des questionnaires commence avec la version approuvée ci-dessus et un contrôle strict se poursuit après cela.

B.3 Processus de chargement (création de la structure) du programme GSS sur l'ordinateur de poche

Cette section décrit les mesures de qualité qui doivent être mises en œuvre aux différents stades du processus afin de s'assurer que chaque ordinateur de poche allant sur le terrain a le logiciel et les spécifications matérielles identiques appropriés.

1) Créer la version maîtresse (pour carte SD) : La version maîtresse (pour carte SD) doit être finalisée sur un seul ordinateur, exempt de tout virus et doté d'un logiciel antivirus adapté. Après la création d'une carte SD maîtresse, qui devrait être chargée sur deux ordinateurs de poche différents et une itération complète de test doit être effectuée sur les deux ordinateurs de poche, répondant à chaque question du questionnaire. Si des erreurs sont détectées sur un des ordinateurs de poche, la programmation—comme il est mentionné dans la **section 3.3.2**—doit être corrigée, et ce processus doit être répété

jusqu'à ce que l'exactitude soit de 100 %. Ce processus doit être réalisé en présence du coordinateur/point focal et avec le directeur informatique.

2) Créer des copies pour tous les ordinateurs de poche : Compte tenu du grand nombre d'ordinateurs de poche qui ont besoin d'être préparés, il est important que chaque étape soit suivie en séquence, réalisée en équipe, et correctement dirigée. Toutes les cartes SD doivent être disposées dans deux boîtes distinctes, l'une contenant des cartes SD vierges formatées. Les versions doivent être copiées une par une, et lorsque la copie est terminée doit être placée dans la deuxième boîte.

3) Coller des autocollants de vérification CQ (modèle) : Tous les ordinateurs de poche doivent avoir un autocollant avec l'identifiant du FI, numéro de série, et une liste vierge CQ, placé sur l'arrière de l'ordinateur de poche. Voyez l'exemple ci-dessous :

701 010		SN A10	
	<u>CQ1</u>		<u>CQ2</u>
<input type="checkbox"/> Réinitialisation du matériel	<input type="checkbox"/> Date	<input type="checkbox"/> Date	
<input type="checkbox"/> Systèmes chargés	<input type="checkbox"/> IDFI	<input type="checkbox"/> IDFI	
<input type="checkbox"/> Configuration du clavier	HH____	<input type="checkbox"/> DEL Trng	
<input type="checkbox"/> Stylet	IQ____	<input type="checkbox"/> Alimentation____	
	<input type="checkbox"/> Alimentation____	<input type="checkbox"/> Stylet	
Ini____	<input type="checkbox"/> Stylet	Cas____	
	CQ1____	CQ2____	

4) Préparer les ordinateurs de poche : Après que la version finale a été copiée sur toutes les cartes SD, tous les ordinateurs de poche doivent être configurés en utilisant le processus approprié étape par étape (voir le processus sur la préparation des ordinateurs de poche décrite dans le **Guide du programmeur du GATS de General Survey System**). Le processus de séparation des ordinateurs de poche dans des boîtes différentes (avant et après configuration) doit être de nouveau utilisé.

5) Effectuer l'étape 1 du contrôle de qualité (CQ1) : Les ordinateurs de poche doivent être répartis de manière égale parmi le personnel de contrôle de qualité qui devrait provenir de l'équipe ayant effectué la configuration des ordinateurs de poche ou supervisé le processus. Chaque étape de CQ1 doit être réalisée sur tous les ordinateurs de poche. CQ1 doit être un processus de vérification de la qualité rigoureux et les éléments suivants doivent être contrôlés pour l'exactitude sur chaque ordinateur de poche, puis marqués sur l'autocollant avec la liste :

- 1) Date et heure
- 2) IDFI (vérifiez autocollant avec ID HH)
- 3) Version du questionnaire des foyers
- 4) Version du questionnaire individuel
- 5) Alimentation – autonomie de la batterie
- 6) Stylet

D'autres contrôles peuvent être ajoutés en fonction des paramètres du pays comme le nombre de cas, ordinateurs de poche pour les enquêteurs masculins/féminins, etc. Pour tout problème identifié, l'ordinateur de poche doit être mis dans une autre boîte, être réinitialisé de nouveau et l'ensemble du processus doit être répété sur ces ordinateurs de poche à partir de l'étape 1.

6) Effectuer l'étape 2 du contrôle de la qualité (CQ2) : CQ2 est effectuée une fois que tous les ordinateurs de poche sont approuvés par CQ1 et que leur étiquette a été signée. Les ordinateurs de poche doivent être réaffectés de sorte que la personne qui effectue l'étape 1 du contrôle de la qualité soit différente de la personne chargée de l'étape 2 de contrôle de la qualité. Un processus identique à celui de CQ1 ci-dessus doit être effectué durant CQ2 ; l'étape supplémentaire de CQ2 consiste à vérifier si tous les autocollants sont marqués et cochés. Il est important de réaliser CQ2, car il a été observé qu'une faible proportion d'erreurs est détectée lors de cette étape. Tout ordinateur de poche présentant des problèmes identifiés lors de CQ2 doit être de nouveau soumis à l'itération en commençant par une réinitialisation à l'étape 1.

7) Signature de la feuille d'achèvement par le directeur informatique et le point focal du pays : Lorsque chaque ordinateur de poche est passé par CQ1 et CQ2, un bref rapport récapitulatif indiquant le nombre d'ordinateurs de poche, les numéros des versions du HQ et IQ, la date, heure et les membres de la production de l'ordinateur de poche et de l'équipe de contrôle de qualité doit être rédigé et approuvé par la signature du directeur informatique et du point focal du GATS. L'approbation doit stipuler que les ordinateurs de poche sont finalisés pour la formation et le travail de terrain de l'enquête complète, aucun changement supplémentaire de programmation ne sera nécessaire, et ils ne seront pas utilisés à d'autres fins jusqu'à ce que le GATS soit terminée dans le pays.

B.4 Variables standard pour le fichier maître de sélection de l'échantillon et le fichier des cas de l'iPAQ

Le système GSS du GATS (pour iPAQ) contient des tableaux normalisés pour stocker les données relatives à la sélection de l'échantillon. L'ensemble actuel des adresses et les variables d'informations sur l'échantillon dans le système du GSS cependant, peuvent ne pas être facilement généralisés à travers plusieurs pays.

1. Saisissez toutes les listes de cas et les missions dans le fichier GSS IDE > Case file tools > Edit case. Immédiatement après que l'échantillon du GATS a été établi, entrez toutes les données pour les variables décrites dans le **tableau B-1** ci-dessous pour chaque foyer de l'échantillon (celui-ci sera désigné comme le fichier maître de sélection de l'échantillon). Envoyez ce fichier immédiatement au DCC suivant la Politique de diffusion des données (assurez-vous qu'aucune information d'identification n'est contenue dans le fichier avant de l'envoyer, référez-vous à la Politique de diffusion des données pour plus d'informations) et d'assurance de la qualité et l'assistance technique nécessaire.
2. Le concepteur CMS Grid utilise ce fichier pour créer le fichier des cas de l'iPAQ avec GSS. Référez-vous au **tableau B-2** ci-dessous pour les variables actuelles du fichier des cas de l'iPAQ de la GATS.

Table B-1 : Informations contenues dans le fichier maître de sélection de l'échantillon pour le DCC

Nom de la variable	Description
Case ID (Numéro de cas)	Identification du cas, utilisé pour identifier de manière unique chaque foyer/répondant. Dans la plupart des enquêtes, il est créé en enchaînant le numéro de grappe ou d'échantillon, le numéro de foyer et le numéro du questionnaire.
C_CODE	Code du pays, utilisé pour identifier l'enquête au cours de laquelle les données ont été recueillies dans le pays en question. Le code est basé sur un code standard international. Code de pays à 2-caractères figurant sur le site web de l'Organisation internationale de normalisation, à l'adresse http://www.iso.org/iso/country_codes/iso_3166_code_lists/english_country_names_and_code_elements.htm .
R_CODE	Code de région, utilisé pour identifier l'enquête au cours de laquelle les données ont été recueillies dans la région en question.
P_CODE	Code de province et/ou d'État, utilisé pour identifier l'enquête au cours de laquelle les données ont été recueillies dans la province ou l'État en question.
D_CODE	Code du comté/district, utilisé pour identifier l'enquête au cours de laquelle les données ont été recueillies dans le comté/district en question.
V_CODE	Code de village/ville, utilisé pour identifier l'enquête au cours de laquelle les données ont été recueillies dans le village ou la ville en question.
TYPE	Lieu de résidence de fait est le type de lieu de résidence dans lequel le ménage /répondant est interrogé. Utilisez « Urban » pour les zones urbaines et « Rural » pour les zones rurales.
SEX	Sexe du ménage/répondant interrogé. Utilisez « Male » pour masculin et « Female » pour féminin.
STRATUM	Les strates de l'échantillon définissent les appariements ou groupements d'unités primaires d'échantillonnage utilisés pour le calcul des erreurs d'échantillonnage lors de l'utilisation de l'analyse des méthodes complexes de données d'enquête telles que la méthode des développements en série de Taylor.
DOMAIN	Le domaine de l'échantillon définit les unités géographiques de base au sein desquelles l'échantillon a été conçu. Par exemple, si l'échantillon a été conçu pour être auto pondéré à l'intérieur de la région, cette variable définirait la région, si l'échantillon a été conçu pour être auto pondéré dans des zones urbaines majeures, d'autres zones urbaines et des zones rurales, cette variable définirait les zones urbaines majeures, d'autres zones urbaines et rurales. Si l'échantillon est auto pondéré à l'échelle nationale, le code de cette variable est 0.
CLUSTER	Le numéro de grappe est le numéro d'identification du point d'échantillon utilisé lors du travail sur le terrain, mais peut être une variable numérotée séquentiellement pour des échantillons avec une structure plus complexe. Cette variable peut être un composite de plusieurs variables du questionnaire. Cette variable est généralement la même que la dernière unité de zone.
PSU	L'unité d'échantillonnage primaire est un chiffre attribué à des points de l'échantillon pour identifier les unités primaires d'échantillonnage à utiliser dans le calcul des erreurs d'échantillonnage. Cette variable est généralement la même que le numéro de grappe et/ou la dernière unité de zone, mais peut varier si la conception de l'échantillon nécessite un processus de sélection en plusieurs étapes.

Nom de la variable	Description
SEGMENT	Le segment est l'unité d'échantillonnage secondaire affectée à des points de l'échantillon pour identifier les unités d'échantillonnage secondaires à utiliser dans le calcul des erreurs d'échantillonnage. Cette variable est généralement la même que le numéro de PSU et/ou la dernière unité de zone, mais peut varier si la conception de l'échantillon nécessite un processus de sélection en plusieurs étapes.
HOUSEHOLD	Le numéro d'unité d'habitation ou le numéro de ménage est le numéro d'identification du ménage dans lequel le répondant a été interrogé, dans le point d'échantillonnage. Dans certains cas, cette variable peut être la combinaison du numéro d'habitation et du numéro de ménage au sein des habitations.

En plus de ces champs, les champs suivants qui font également partie du fichier maître de sélection de l'échantillon peuvent être sélectionnés pour le fichier des cas à l'aide de GSS IDE > Case file tools > CMS grid designer.

Table B-2 : Exigences des champs casefile.txt du GATS

Variable	Type	Description du champ
FIID*	6 chiffres	ID de l'enquêteur de terrain attribué à chaque CaselD. (Généralement commence par 999 000 et numérote séquentiellement les enregistrements. Exemple : 999 900, 999 001, 999 002 ... etc.)
CaseID*	6 chiffres	Numéro de cas séquentiel unique (Généralement commence par 100 001 et numérote séquentiellement les enregistrements. Exemple : 100 001 12 000)
NumCaselD	6 chiffres	Séquentiel, commence de préférence par 1...n
FormNum*	Texte	
ProjectName	Texte	20 <insérer le nom du pays>
SCRNum*	Nombre	0
CreateDate*	Date	Date created
STREET_NO*	Texte	10 caractères maximum
STREET*	Texte	55 caractères maximum
APT_NUM	Texte	10 caractères maximum
BOX_NUM	Texte	10 caractères maximum
RURAL_ROUTE* (RR)	Texte	10 caractères maximum
TRACT_BG* (Champ intitulé « Other 1 » dans l'écran de modification de l'adresse)	Texte	25 caractères maximum
BLOCK* (Champ Other 2)	Texte	25 caractères maximum
HC * (Champ Other 3)	Texte	25 caractères maximum
C * (Champ Other 4)	Texte	25 caractères maximum
TYPE*	Texte	MASCULIN ou FEMININ ou sexe dans la langue du pays
CITY*	Texte	25 caractères maximum
STATE*	Texte	25 caractères maximum
ZIP*	Texte	25 caractères maximum
COUNTY*	Texte	25 caractères maximum
EvtCode*	Texte	0 (toujours 0)
PERIOD (indicateur urbain/rural)	Nombre	1 (1 = urbain 2 = rural)

YEAR	Texte	1
WAVE	Nombre	1
ROCBASE	Nombre	1
TRAININGFLAG	Texte	N
FSID	Nombre	Numéro d'identification du superviseur de terrain (souvent non utilisé, utilisé pour suivre les enquêteurs de terrain affectés à un superviseur de terrain)
ALTERNATIVE	Texte	Néant
ACTIVE*	Texte	O
ALTER_CASEID	Texte	Néant (utilisé par le CMS)
PostDTTM	Date/Heure	Néant (utilisé par le CMS)
POSTED	Texte	Néant (utilisé par le CMS)
Nom	Texte	Néant (utilisé par le CMS)
Phone	Texte	Néant (utilisé par le CMS)
OtherID	Texte	Néant (utilisé par le CMS)

* Indique une information affichée dans la grille du CMS

Annexe C : Collecte et gestion des données

C.1 Fiche d'Information des variables indicatrices du GATS

Nom de la variable	Question du GATS	Taux de réponse à la question
B01	Actuellement, fumez-vous du tabac quotidiennement, moins d'une fois par jour, ou pas du tout ?	
B02	Avez-vous déjà fumé du tabac quotidiennement dans le passé ?	
B03	Dans le passé, avez-vous fumé du tabac quotidiennement, moins d'une fois par jour, ou pas du tout ?	
B06a	En moyenne, combien des produits suivants fumez-vous actuellement, chaque jour ? Dites-moi également si vous fumez le produit, mais pas tous les jours. Cigarettes fabriquées ?	
B06b	Cigarettes roulées ?	
B06c	Kreteks ?	
B06d	Pipes remplies de tabac ?	
B06e	Cigares, cheroots, ou cigarillos ?	
B06f	Nombre de séances d'utilisation d'une pipe à eau (narguilé) par jour ?	
B06g	D'autres sortes de tabac ?	
B10a	Combien des produits suivants fumez-vous actuellement au cours d'une semaine normale ? Cigarettes fabriquées ?	
B13a	Depuis combien de temps avez-vous arrêté de fumer ? [SAISIR L'UNITÉ]	
B13b	(Depuis combien de temps avez-vous arrêté de fumer ?) [ENTRER LE NOMBRE DE {REEMPLIR B13a : ANNÉES/MOIS/SEMAINES/JOURS}]	
B16.	Lors des consultations chez un médecin ou un professionnel de la santé, au cours des 12 derniers mois, vous a-t-on demandé si vous fumiez du tabac ?	
B17	Lors des consultations chez un médecin ou un professionnel de la santé, au cours des 12 derniers mois, vous a-t-on conseillé d'arrêter de fumer du tabac ?	
C01	Actuellement, utilisez-vous du tabac sans fumée quotidiennement, moins d'une fois par jour, ou pas du tout ?	
C02	Avez-vous déjà utilisé du tabac sans fumée quotidiennement dans le passé ?	
C03	Dans le passé, avez-vous utilisé du tabac sans fumée quotidiennement, moins d'une fois par jour, ou pas du tout ?	
C13a	Depuis combien de temps avez-vous arrêté d'utiliser du tabac sans fumée ? [SAISIR L'UNITÉ]	
C13b	(Depuis combien de temps avez-vous arrêté d'utiliser du tabac sans fumée ?) [ENTRER LE NOMBRE DE {REEMPLIR C13a : ANNÉES/MOIS/SEMAINES/JOURS}]	
C16	Lors des consultations chez un médecin ou un professionnel de la santé, au cours des 12 derniers mois, vous a-t-on demandé si vous utilisiez du tabac sans fumée ?	
C17	Lors des consultations chez un médecin ou un professionnel de la santé, au cours des 12 derniers mois, vous a-t-on conseillé d'arrêter d'utiliser du tabac sans fumée ?	
D01	Au cours des 12 derniers mois, avez-vous essayé d'arrêter de fumer ?	

Nom de la variable	Question du GATS	Taux de réponse à la question
D06	Lors des consultations chez un médecin ou un professionnel de la santé, au cours des 12 derniers mois, vous a-t-on demandé si vous fumiez du tabac ?	
D07	Lors des consultations chez un médecin ou un professionnel de la santé, au cours des 12 derniers mois, vous a-t-on conseillé d'arrêter de fumer du tabac ?	
D08	Lequel des énoncés suivants décrit-il le mieux votre pensée au sujet d'arrêter de fumer ? J'ai l'intention d'arrêter avant le mois prochain, je pense arrêter de fumer dans les 12 prochains mois, j'arrêterai un jour, mais pas dans les 12 prochains mois, ou arrêter de fumer ne m'intéresse pas.	
D09	Au cours des 12 derniers mois, avez-vous essayé d'arrêter d'utiliser du tabac sans fumée ?	
D14	Lors des consultations chez un médecin ou un professionnel de la santé, au cours des 12 derniers mois, vous a-t-on demandé si vous utilisiez du tabac sans fumée ?	
D15	Lors des consultations chez un médecin ou un professionnel de la santé, au cours des 12 derniers mois, vous a-t-on conseillé d'arrêter d'utiliser du tabac sans fumée ?	
D16	Lequel des énoncés suivants décrit-il le mieux votre pensée au sujet d'arrêter d'utiliser du tabac sans fumée ? J'ai l'intention d'arrêter avant le mois prochain, je pense arrêter de fumer dans les 12 prochains mois, j'arrêterai un jour, mais pas dans les 12 prochains mois, ou arrêter de fumer ne m'intéresse pas.	
E05	Travaillez-vous habituellement à l'intérieur ou l'extérieur ?	
E08	Au cours des 30 derniers jours, quelqu'un a-t-il fumé dans les espaces intérieurs où vous travaillez ?	
E17	D'après ce que vous savez ou croyez, respirer la fumée des autres peut-il entraîner une maladie grave chez les non-fumeurs ?	
F01a	La dernière fois que vous avez acheté des cigarettes pour vous-même, combien de cigarettes avez-vous achetées ? [SAISIR L'UNITÉ]	
F01b	(La dernière fois que vous avez acheté des cigarettes pour vous-même, combien de cigarettes avez-vous achetées ?) [ENTRER LE NOMBRE DE {REEMPLIR F01a : CIGARETTES/PAQUETS/CARTONS/AUTRES}]	
F02	Au total, combien avez-vous dépensé pour cet achat ?	
G01a*	Au cours des 30 derniers jours, avez-vous remarqué des *informations* sur les dangers de fumer des cigarettes, ou qui encouragent à arrêter dans l'un des supports suivants ? Dans des journaux ou des revues ?	
G01b*	À la télévision ?	
G01c*	À la radio ?	
G01d*	Sur des panneaux d'affichage ?	
G01e*	Ailleurs ?	
G201a1**	Au cours des 30 derniers jours, avez-vous remarqué des informations dans les *journaux ou les magazines* sur les dangers du tabac, ou encourageant à arrêter l'utilisation des produits à base de tabac suivants ? Cigarettes ?	
G201a2**	Tabac sans fumée ?	

Nom de la variable	Question du GATS	Taux de réponse à la question
G201b1**	(Au cours des 30 derniers jours, avez-vous remarqué des informations à la *télévision* sur les dangers du tabac, ou encourageant à arrêter l'utilisation des produits à base de tabac suivants ?) Cigarettes ?	
G201b2**	Tabac sans fumée ?	
G201c1**	(Au cours des 30 derniers jours, avez-vous entendu des informations à la « radio » sur les dangers du tabac, ou encourageant à arrêter l'usage des produits du tabac suivants ?) Cigarettes ?	
G201c2**	Tabac sans fumée ?	
G201d1**	(Au cours des 30 derniers jours, avez-vous remarqué des informations sur les *panneaux d'affichage* sur les dangers du tabac, ou encourageant à arrêter l'utilisation des produits à base de tabac suivants ?) Cigarettes ?	
G201d2**	Tabac sans fumée ?	
G201e1**	(Au cours des 30 derniers jours, avez-vous remarqué des informations *à d'autres endroits* sur les dangers du tabac, ou encourageant à arrêter l'utilisation des produits à base de tabac suivants ?) Cigarettes ?	
G201e2**	Tabac sans fumée ?	
G02*/G202**	Au cours des 30 derniers jours, avez-vous remarqué les avertissements de santé sur les paquets de cigarettes ?	
G202a*	Au cours des 30 derniers jours, avez-vous remarqué des avertissements de santé sur les produits de tabac sans fumée ?	
G03*/G203**	Au cours des 30 derniers jours, les avertissements sur les paquets de cigarettes vous ont-ils amené à envisager d'arrêter ?	
G203a**	Au cours des 30 derniers jours, les avertissements sur les paquets de tabac sans fumée vous ont-ils amené à envisager d'arrêter ?	
G04a*	Au cours des 30 derniers jours, avez-vous remarqué des *publicités ou des affiches faisant la promotion* des cigarettes dans les endroits suivants ? Dans les magasins où les cigarettes sont vendues ?	
G04b*	À la télévision ?	
G04c*	À la radio ?	
G04d*	Sur des panneaux d'affichage ?	
G04e*	Sur des affiches ?	
G04f*	Dans des journaux ou des revues ?	
G04g*	Dans les cinémas ?	
G04h*	Sur Internet ?	
G04i*	Sur les véhicules de transport en commun ou dans les stations ou gares ?	
G04j*	Sur les murs des espaces publics ?	
G04k*	Ailleurs ?	
G204a1**	Au cours des 30 derniers jours, avez-vous remarqué des publicités, ou des affiches, faisant la promotion de produits suivants à base de tabac, dans les « magasins où les produits sont vendus » ? Cigarettes ?	

Nom de la variable	Question du GATS	Taux de réponse à la question
G204a2**	Tabac sans fumée ?	
G204b1**	Au cours des 30 derniers jours, avez-vous vu des publicités ou des affiches faisant la promotion des produits à base de tabac suivants, à la « télévision » ? Cigarettes ?	
G204b2**	Tabac sans fumée ?	
G204c1**	Au cours des 30 derniers jours, avez-vous entendu des publicités faisant la promotion des produits à base de tabac suivants, à la « radio » ? Cigarettes ?	
G204c2**	Tabac sans fumée ?	
G204d1**	Au cours des 30 derniers jours, avez-vous remarqué des publicités faisant la promotion les produits à base de tabac suivants, sur des « panneaux d'affichage » ? Cigarettes ?	
G204d2**	Tabac sans fumée ?	
G204e1**	Au cours des 30 derniers jours, avez-vous remarqué des publicités ou des logos faisant la promotion des produits à base de tabac suivants, sur des « affiches » ? Cigarettes ?	
G204e2**	Tabac sans fumée ?	
G204f1**	Au cours des 30 derniers jours, avez-vous remarqué des publicités ou des logos faisant la promotion des produits à base de tabac suivants, dans des « journaux ou des magazines » ? Cigarettes ?	
G204f2**	Tabac sans fumée ?	
G204g1**	Au cours des 30 derniers jours, avez-vous remarqué des publicités ou des logos faisant la promotion des produits à base de tabac suivants, dans les « cinémas » ? Cigarettes ?	
G204g2**	Tabac sans fumée ?	
G204h1**	Au cours des 30 derniers jours, avez-vous remarqué des publicités ou des logos faisant la promotion des produits à base de tabac suivants, sur « Internet » ? Cigarettes ?	
G204h2**	Tabac sans fumée ?	
G204i1**	Au cours des 30 derniers jours, avez-vous remarqué des publicités ou des logos faisant la promotion des produits à base de tabac suivants, sur les « véhicules des transports en commun ou leur gare ou station » ? Cigarettes ?	
G204i2**	Tabac sans fumée ?	
G204j1**	Au cours des 30 derniers jours, avez-vous remarqué des publicités ou des logos faisant la promotion des produits à base de tabac suivants, sur les « murs des espaces publics » ? Cigarettes ?	
G204j2**	Tabac sans fumée ?	
G204k1**	Au cours des 30 derniers jours, avez-vous remarqué des publicités ou des logos faisant la promotion des produits à base de tabac suivants, à « d'autres endroits » ? Cigarettes ?	
G204k2**	Tabac sans fumée ?	

Nom de la variable	Question du GATS	Taux de réponse à la question
G06a*	Au cours des 30 derniers jours, avez-vous remarqué l'un des types de promotions de cigarettes suivants ? Échantillons gratuits de cigarettes ?	
G06b*	Cigarettes à prix réduit ?	
G06c*	Bons pour des cigarettes ?	
G06d*	Cadeaux gratuits ou rabais spécial sur d'autres produits lors de l'achat de cigarettes ?	
G06e*	Vêtements ou autres articles avec le nom d'une marque de cigarette ou un logo ?	
G06f*	Promotions de cigarettes par la poste ?	
G206a1**	Au cours des 30 derniers jours, avez-vous remarqué des échantillons gratuits de produits à base de tabac suivants ? Cigarettes ?	
G206a2**	Tabac sans fumée ?	
G206b1**	Au cours des 30 derniers jours, avez-vous remarqué des produits suivants à base de tabac, à prix réduit ? Cigarettes ?	
G206b2**	Tabac sans fumée ?	
G206c1**	Au cours des 30 derniers jours, avez-vous remarqué des bons pour les produits à base de tabac suivants ? Cigarettes ?	
G206c2**	Tabac sans fumée ?	
G206d1**	Au cours des 30 derniers jours, avez-vous remarqué des cadeaux ou des offres de rabais sur d'autres produits lors de l'achat de l'un des produits à base de tabac suivants ? Cigarettes ?	
G206d2**	Tabac sans fumée ?	
G206e1**	Au cours des 30 derniers jours, avez-vous remarqué des vêtements ou d'autres articles avec un nom de marque ou un logo des produits à base de tabac suivants ? Cigarettes ?	
G206e2**	Tabac sans fumée ?	
G206f1**	Au cours des 30 derniers jours, avez-vous reçu des promotions par la poste pour les produits à base de tabac suivants ? Cigarettes ?	
G206f2**	Tabac sans fumée ?	
H01	D'après ce que vous savez ou croyez, est-ce que fumer du tabac peut entraîner une maladie grave ?	
H03	D'après ce que vous savez ou croyez, est-ce que le tabac « sans fumée » peut entraîner une maladie grave ?	

* Numérotation des questions pour la structure des médias n° 1

** Numérotation des questions pour la structure des médias n° 2

Annexe D : Période Post- Collecte des données

D.1 Nettoyage et préparation des données pour les calculs de pondération de l'échantillon utilisant une méthode de fusion manuelle

Contexte

Cette section fournit des directives pour la fusion des fichiers de données, la validation des variables et des enchaînements, ainsi que la création des codes de décision en utilisant un logiciel statistique¹. Une fois que ces procédures sont effectuées, les données doivent être propres et prêtes pour le calcul de la pondération. Un échantillon de codes de logiciel statistique est fourni dans cette section en exemple où cela est nécessaire, cependant, un ensemble complet de codes est disponible sur demande. Lors du nettoyage des données, il est recommandé qu'un nouvel ensemble de données (un rapport d'erreur) soit créé, composé de tous les enregistrements contenant des données collectées et agrégées sur le terrain. Les exemples SAS fournis dans cette section identifient tous les enregistrements avec des données non valides puis transfèrent ces enregistrements dans un ensemble de données SAS pour un examen plus approfondi (le code SPSS pour créer un rapport d'erreur est disponible auprès du DCC du GATS sur demande). À ce stade, l'analyste peut examiner les enregistrements contenant des erreurs et de tenter de les résoudre. Si une erreur ne peut pas être résolue, contactez le point focal technique du CDC du pays.

Exactitude des fichiers de réponses agrégées et transposées

Alors que la validation des fichiers des réponses HH et IQ transposées doit avoir eu lieu immédiatement après que les fichiers ont été transposés (comme mentionné dans le chapitre Collecte et Gestion des données), il est essentiel de vérifier que les données ont été agrégées/transposées correctement avant le calcul des codes de décisions et des pondérations de l'échantillon.

- Lors de la lecture du fichier des réponses HH transposées et du fichier des réponses IQ transposées avec le logiciel statistique du pays, il est nécessaire d'identifier le nom de chaque variable, l'ordre et le type (caractère ou numérique). Si le type de variable n'est pas spécifié, alors le type sera par défaut un caractère. La suite GSS peut créer une liste des noms de variables, d'ordre, et de type à partir du menu Data Aggregation, Transpose.
- Validez les CaseID dans les fichiers des réponses HH et IQ transposées (décrit dans le paragraphe 4.3.4). Vérifier la validité des champs CaseID avec les contrôles suivants :
 - Le processus d'agrégation ne doit pas permettre les doublons de CaseID dans le sdf maître. Il est toujours souhaitable, cependant, de rechercher les éventuels doublons et anomalies dans les données. Un doublon de CaseID indiquerait une erreur dans le processus d'agrégation/de transposition.
- *Dans SAS le code suivant pourrait être utilisé pour le transfert vers un ensemble de données SAS de tous les enregistrements avec des CaseID doubles (avant l'exécution de cette ligne de code, les données doivent d'abord être triées par CaseID puis organisées par CaseID).*

IF (FIRST.CASEID NE 1 OR LAST.CASEID NE 1) THEN OUTPUT;

¹Le centre de coordination des données (DCC) de la GATS fournit une assistance technique pour l'utilisation des logiciels statistiques suivants : SAS, SPSS et STATA.

- Assurez-vous que chaque enregistrement a un CaseID valide. Les champs de CaseID blancs ou invalides peuvent indiquer une erreur du processus d'agrégation/de transposition.
 - *Dans SAS le code suivant pourrait être utilisé pour transférer des enregistrements avec des CaseID blancs.*

```
IF CASEID = ' ' THEN OUTPUT;
```

Combiner le fichier agrégé DuEvt.csv avec les fichiers des réponses transposées HH et IQ

- Le fichier des événements agrégés (le fichier DuEvt.csv) contient tous les événements (chaque ligne est un événement différent) et le code de résultat correspondant (la variable RESULT) pour le HH et l'IQ. Dans la plupart des cas, chaque CaseID aura un code de résultat final pour le HH et un code de résultat final pour l'IQ. Seuls ces codes de résultat finaux doivent être pris pour le fichier fusionné.
 - Les codes de résultat finaux doivent avoir été attribués à chaque cas (HH et IQ) comme indiqué dans la **section 4.1.3**. Toutefois, si un code de résultat final n'a pas été assigné à un cas (pour l'HH ou l'IQ), une décision devra être prise quant au code de résultat final à affecter puis à prendre pour le fichier fusionné.
 - *Si le HH ou l'IQ ne contient qu'un seul code de résultat temporaire, le code de résultat final correspondant doit être attribué et pris.*
 - *Si le HH ou l'IQ contient plus d'un code temporaire, un examen attentif des codes temporaires devra être effectué afin de déterminer quel code de résultat final doit être attribué. Par exemple, un Enquêteur de terrain a rendu visite à un ménage quatre fois afin de compléter l'IQ et il a saisi les codes de résultats temporaires suivants pour chaque visite : 1) 309-N'est pas au domicile, 2) 304-Refus, 3) 309-N'est pas au domicile, 4) 309-N'est pas au domicile. Dans ce cas, la personne sélectionnée a refusé l'entretien à la deuxième visite. Lorsque l'enquêteur a essayé de revenir deux fois au ménage pour convertir le refus, la personne n'était pas chez elle. Même si les deux dernières visites ont été codées 309-N'est pas au domicile, un code de résultat final de 404-Refus doit être attribué.*
- Lorsque tous les cas ont un code de résultat final, ne transférez que les événements avec les codes de résultat finaux dans de nouveaux fichiers. Transférez les événements HH dans un fichier et les événements IQ dans un autre fichier (il doit alors y avoir qu'une seule ligne par CaseID dans chaque fichier).
 - Dans l'ensemble de données HH, renommez l'ensemble de variables suivantes : la variable code de résultat final (RESULT) doit être renommée HH_RESULT, la variable EVENTDATE doit être renommée HH_EVENTDATE et MODIFYDTTM doit être renommée HH_MODIFYDTTM. Dans l'ensemble de données IQ, renommez l'ensemble des variables suivantes : RESULT doit être renommé IQ_RESULT, EVENTDATE doit être renommé IQ_EVENTDATE et MODIFYDTTM doit être renommé IQ_MODIFYDTTM.
 - Conservez uniquement les variables suivantes : CASEID dans les deux ensembles de données, HH_EVENTDATE, HH_RESULT et HH_MODIFYDTTM dans l'ensemble de données HH, et IQ_EVENTDATE, IQ_RESULT et IQ_MODIFYDTTM dans l'ensemble des données IQ.

Le fichier des événements HH doit être fusionné avec le fichier de réponses HH transposées par la variable CaseID. Le fichier des événements IQ doit être fusionné avec le fichier de réponses IQ transposées par CaseID. Vérifiez les fusions en procédant aux vérifications suivantes :

- Choisissez au hasard 5 à 10 cas dans les ensembles de données fusionnées et revenez aux fichiers utilisés dans chacune des fusions et assurez-vous que ces cas fusionnés sont corrects. Si les enregistrements ne correspondent pas, cela indiquerait que le processus de fusion n'a pas fonctionné correctement.
- Chaque cas dans un fichier de transposition doit avoir un code de résultat final correspondant à partir des fichiers d'événements. Après la fusion, si un enregistrement n'a pas de code de résultat ou vice versa, vérifiez de nouveau le processus de fusion pour vous assurer qu'il a fonctionné correctement. Si cela ne résout pas le problème, vérifiez que le processus de sélection du code de résultat final et le transfert de ce code dans un nouveau fichier a fonctionné correctement.
- Créez une nouvelle variable à partir du CaseID dans chacun des deux ensembles de données fusionnées. La nouvelle variable doit être appelée ID et contiendra tous les chiffres avant l'extension -00 ou -01.
 - Dans SAS, le code suivant peut être utilisé pour créer la nouvelle variable :

```
LENGTH ID $12;  
ID = SCAN(CASEID,1,"-");
```

Fusionner les fichiers de réponses transposées HH et IQ

- Fusionnez ensemble le fichier HH et le fichier IQ (créé en 5.1.2) par la variable ID. Vérifiez la fusion en procédant à la vérification suivante :
 - Choisissez au hasard 5 à 10 cas dans l'ensemble de données fusionnées et revenez aux fichiers IQ et HH et assurez-vous que ces cas fusionnés sont corrects. Si les enregistrements ne correspondent pas, cela indiquerait que le processus de fusion n'a pas fonctionné correctement.

Supprimer les variables d'identification des foyers et valider les CaseID du fichier maître de sélection de l'échantillon

- Créez un nouveau fichier à partir du fichier maître de sélection de l'échantillon, dans lequel toutes les informations d'identification ont été enlevées. Reportez-vous à l'annexe B.4 pour plus d'informations sur le fichier maître de sélection de l'échantillon. Les variables qui doivent rester dans le nouveau fichier maître de sélection de l'échantillon sont les suivantes : NUMÉRO DE CAS, TYPE, SEXE (si applicable), STRATE, PSU, SEGMENT, et FOYERS.
- Validez les CaseID du fichier maître de sélection de l'échantillon. Vérifiez qu'aucun CaseID n'apparaît plusieurs fois dans le fichier maître de sélection de l'échantillon. Les enregistrements avec des doublons de CaseID indiquent que deux cas ont reçu des CaseID identiques.
 - Le code SAS du paragraphe 5.1.1 peut être utilisé pour identifier les enregistrements avec des CaseID doubles dans le fichier maître de sélection de l'échantillon.

- Assurez-vous que chaque répondant a un CaselD valide. Le fichier maître de sélection de l'échantillon est structuré afin que seules les données au niveau des foyers soient présentes et que les extensions 00 ou 01 ne soient pas utilisées. Les champs CaselD vides ou invalides indiquent une erreur de saisie de données.
- Le code SAS du paragraphe 5.1.2 peut être utilisé pour identifier les enregistrements avec des CaselD vides dans le fichier maître de sélection de l'échantillon.
- Créez une nouvelle variable à partir du CaselD. Créez une nouvelle variable appelée ID qui est égale à CASEID. Cette variable est nécessaire pour la prochaine étape où le fichier maître de sélection de l'échantillon est fusionné avec le fichier des réponses transposées créé au paragraphe 5.1.3 par la variable ID.

Fusionner un fichier de réponses transposées avec le fichier maître de sélection de l'échantillon

- Le fichier de réponses transposées contenant les réponses HH et IQ (de 5.1.3) doit être fusionné avec le fichier maître de sélection de l'échantillon par la variable ID. Le contrôle suivant est recommandé afin de valider l'exactitude du fichier des données fusionné :
 - Choisissez au hasard 5 à 10 cas dans les ensembles de données fusionnées et revenez au fichier de réponses transposées et au fichier maître de sélection de l'échantillon et assurez-vous que ces cas fusionnés sont corrects. Si les enregistrements ne correspondent pas, cela indiquerait que le processus de fusion n'a pas fonctionné correctement.

D.2 Codes de décision finale et calculs des taux de réponse

Après un examen attentif de la littérature disponible sur les systèmes de surveillance, les enquêtes nationales auprès des ménages sur le tabac ou non, utilisant le mode de collecte de données par téléphone ou en personne dans les ménages, une liste des codes de décision finale et les calculs de taux de réponse ont été finalisés pour le GATS. Plusieurs enquêtes ont été examinées à cette fin, y compris les enquêtes de l'Institut brésilien de géographie et de statistique (IBGE), l'enquête nationale de démographie et de santé sexuelle et reproductive, 2004 (ENDSSR-2004), l'enquête sur la santé reproductive de Géorgie (RHS), les enquêtes démographiques et sanitaires (DHS), l'enquête nationale sur la santé familiale (NFHS), le système de surveillance des facteurs de risques comportementaux (BRFSS), le système de surveillance des comportements à risque de la jeunesse (YRBSS), et l'Association américaine pour la recherche sur l'opinion publique (AAPOR).

Les codes de décision finale et les rapports de taux de réponse sont décrits dans les tableaux suivants dans cette section.

Les tableaux D-1 et D-2 présentent les codes de décisions finales HH et IQ qui doivent être affectés pour différents codes de résultat final des questionnaires HH et IQ. (De plus amples détails sur les codes de résultat figurent dans le *Manuel de l'Enquêteur de terrain du GATS*).

Le tableau D-3 donne un exemple de modèle pour le calcul et la notification des taux de réponse et des codes de décision dans le *Rapport du Pays*.

Tableau D-1 : Codes des décisions et des résultats finaux du Questionnaire Ménage

Codes des décisions finales	Codes de résultat final	Nom et description
1	200	Questionnaire Ménage rempli, un individu sélectionné -Le Questionnaire Ménage est considéré comme complet si la liste est complète
2	201	Questionnaire Ménage rempli, aucun individu sélectionné -Personne n'est âgé de 15 ans ou plus -Personne ne considère le ménage comme étant leur lieu habituel de résidence -Le ménage a été attribué aux hommes et aucun individu de sexe masculin ne réside dans le ménage ou le ménage a été affecté aux femmes et aucun individu de sexe féminin ne réside dans le ménage (si la randomisation des sexes a été utilisée)
3	202	Questionnaire Ménage partiellement rempli, liste interrompue -Entretien incomplet
4	203	Questionnaire Ménage incomplet, le répondant de présélection approprié n'a pas pu être identifié -Personne âgé de 18 ou plus n'est pas au domicile -Le membre du ménage âgé de 18 ou plus est incompetent - Ces ménages n'ont peut-être pas de répondants éligibles.
5	209	Personne au domicile
6	204	Refus du ménage
7	205	Ménage vide
8	206	Adresse sélectionnée n'est pas un ménage
9	208	Autre non-réponse du ménage

Tableau D-2 : Codes des décisions et des résultats finaux du questionnaire individuel

Codes des décisions finales	Codes de résultat final	Nom et description
11	400, 402*	Questionnaire individuel rempli -Questionnaire rempli jusqu'à la question E01 -Aucune des questions clés (par exemple, B01/B02/B03 et C01/C02/C03 selon le cas) = Ne sais pas ou Refusé
12	402*	Incomplet -Questionnaire interrompu avant la question E01
13	403	L'individu sélectionné s'est avéré inadmissible à l'enquête ultérieurement -Âge < 15 -L'individu ne considère pas le ménage comme étant son lieu de résidence habituel -Erreur de sexe (si la randomisation des sexes a été utilisée)
14	409	Le répondant sélectionné n'est pas chez lui
15	404	Refus du répondant sélectionné
16	407	Répondant choisi incompetent
17	408	Autre non-réponse individuelle

* Cas avec un code de résultat final de 402 (*questionnaire individuel* complété partiellement) recevra un code de décision de 11 (IQ terminé) ou 12 (incomplet). Voir les règles décrites dans la **section 5.1.7**.

Tableau D-3 : Tableau 3.1 du rapport national du GATS — Modèle pour les rapports des taux de réponse

	Résidence				Total	
	Urbaine		Rurale		Nombre	Pourcentage
	Nombre	Pourcentage	Nombre	Pourcentage		
<i>Ménage sélectionné</i>						
Rempli(HC)						
Rempli– Aucun individu admissible (HCNE)						
Incomplètement rempli (HINC)						
Aucun répondant à la présélection (HNS)						
Personne au domicile (HNS)						
Refusé [de répondre] (HR)						
Inoccupé (HUO)						
L'adresse n'est pas une habitation (HAND)						
Autre ¹ (HO)						
Nombre total de ménages sélectionnés		100		100		100
<i>Taux de réponse des ménages (HRR) (%)²</i>						
<i>Personne sélectionnée</i>						
Rempli [Répondu] (PC)						
Partiellement rempli [Incomplet] (PINC)						
Non éligible (PNE)						
N'est pas au domicile (PNH)						
Refusé [de répondre] (PR)						
Incapable [de répondre] (PI)						
Autre ¹ (PO)						
Nombre total de personnes échantillonnées		100		100		100
<i>Taux de réponse au niveau des personnes (PRR) (%)³</i>						
<i>Taux de réponse total (TRR) (%)⁴</i>						
¹ Autre inclus tout autre résultat ne figurant pas dans la liste.						
² Le taux de réponse des ménages (HRR) est calculé comme suit :						
		$\frac{HC * 100}{HC + HINC + HNS + HNH + HR + HO}$				
³ Le taux de réponse au niveau des personnes (PRR) est calculé comme suit :						
		$\frac{PC * 100}{PC + PINC + PNH + PR + PI + PO}$				
⁴ Le taux de réponse total (TRR) est calculé comme suit : (HRR x PRR) / 100						
Remarques :						
– Un entretien des foyers incomplets (c.-à-d., la liste n'a pas pu être terminée) était considéré comme un non-répondant à la GATS. Ainsi, ces cas (HINC) n'ont pas été inclus dans le numérateur du taux de réponse des foyers.						
– Le nombre total de personnes échantillonnées doit être égal au nombre d'entretiens des foyers terminés [HC].						
– Un entretien individuel terminé [PC] inclut les répondants qui avaient complété au moins la question E01 et qui ont fourni des réponses valides aux questions B01/B02/B03 (et C01/C02/C03 le cas échéant). Les répondants qui ne remplissaient pas ces critères ont été considérés comme des non-répondants incomplets (PINC) à la GATS et par conséquent, n'ont pas été inclus dans le numérateur du taux de réponse au niveau des personnes.						

D.3 Modèle d'ajustements de l'étalonnage de la pondération poste-stratification entre les cellules d'ajustement

Contexte

La dernière étape de production des pondérations de l'échantillon implique l'étalonnage de la pondération du dénombrement de la population par des corrélats connus de mesures de résultats d'étude clés, appelées variables d'étalonnage (par exemple, le sexe, l'éducation, l'âge, environnement urbain/rural et la région, comme le suggère le *Manuel de pondération de l'échantillon du GATS*). En termes de procédure, l'étalonnage sous la forme de post stratification consiste à former des « cellules d'ajustement » par la classification croisée des mesures de corrélat. La « correction post stratification » (PSA) dans chacune de ces cellules d'ajustement est $< 1,00$ si l'échantillon dans cette catégorie était surreprésenté dans l'échantillon après comptabilisation pour la sélection de l'échantillon et des non-réponses, et la PSA $> 1,00$ dans les catégories où l'échantillon était sous-représenté. Examiner la tendance de la PSA parmi les cellules d'ajustement permet de comprendre où le déséquilibre de l'échantillon demeure, après la correction du déséquilibre de l'échantillon dû à la non-réponse.

Source(s) des données

Un fichier de données contenant l'ensemble des pondérations finales de l'échantillon corrigé pour la non-réponse, ainsi que tous les composants de calcul de la pondération finale de chaque répondant, doit être utilisé pour ces calculs.

Logiciel de calcul

Aucun en particulier puisque les calculs nécessaires peuvent facilement être programmés dans n'importe quel logiciel.

Calcul

Les étapes de la procédure de calcul des corrections post-stratification sont décrites dans la **section 3.3** du *Manuel de pondération de l'échantillon du GATS*. En utilisant les résultats finaux de ces calculs, organisez l'ensemble des PSA sous la forme d'un tableau multi varié par rapport à la corrélation des variables qui définissent les cellules d'ajustement (c.-à-d., les variables d'étalonnage). On espère trouver que tous PSA seront proches de 1,00, certaines un peu supérieures à 1,00, les autres un peu inférieures à 1,00.

Interprétation

Les écarts importants par rapport à 1,00 de ces PSA suggèrent qu'il y a eu des problèmes considérables de déséquilibre de l'échantillon en raison de la couverture différentielle de l'échantillon des foyers, des problèmes non détectés de calcul des pondérations de l'échantillon, ou des grands changements démographiques de la population cible entre la date du dénombrement de la population à laquelle l'échantillon a été étalonné et la date de l'échantillonnage et de collecte des données de l'enquête du GATS.

Tableau D-4 : Tableau d'un exemple hypothétique de correction de l'étalonnage de la pondération poste-stratification, suivant les régions, le milieu urbain/rural, le sexe et l'âge.

Région	Zone urbaine/rurale	Sexe	Âge	Correction	
1	Urbaine	Masculin	15 à 24 25 à 44 45 à 59 60 et plus		
		Féminin	15 à 24 25 à 44 45 à 59 60 et plus		
	Rurale	Masculin	15 à 24 25 à 44 45 à 59 60 et plus		
		Féminin	15 à 24 25 à 44 45 à 59 60 et plus		
	2	Urbaine	Masculin	15 à 24 25 à 44 45 à 59 60 et plus	
			Féminin	15 à 24 25 à 44 45 à 59 60 et plus	
Rurale		Masculin	15 à 24 25 à 44 45 à 59 60 et plus		
		Féminin	15 à 24 25 à 44 45 à 59 60 et plus		
3		Urbaine	Masculin	15 à 24 25 à 44 45 à 59 60 et plus	
			Féminin	15 à 24 25 à 44 45 à 59 60 et plus	
	Rurale	Masculin	15 à 24 25 à 44 45 à 59 60 et plus		
		Féminin	15 à 24 25 à 44 45 à 59 60 et plus		

D.4 Effet de la variable de pondération de l'échantillon sur la précision des estimations de l'enquête

Contexte

La variation de pondération de l'échantillon crée souvent une augmentation de la variance (c.-à-d., une réduction de la précision) des estimations de l'enquête. Ce type de variation sera un problème important pour les conceptions d'échantillon des pays qui ont sur échantillonné par zone urbaine/rurale ou par région. Kish (1965, section 11.7) a proposé un modèle simple, mais daté pour mesurer l'augmentation multiplicative de la variance lorsque la taille de pondération n'est pas corrélée avec la mesure de l'enquête utilisée pour estimer une caractéristique (θ) de population simple. Nous allons utiliser $Meff_w$ pour désigner cet effet multiplicatif.

$Meff_w$ n'est pas spécifique de l'estimation, mais il est spécifique du domaine de rapport des estimations. Il s'appliquera donc à toutes les estimations de la GATS dans le domaine concerné. Par exemple, $Meff_w$ calculé pour les estimations de la population globale doit être calculé à partir de l'échantillon entier, mais $Meff_w$ calculé pour les estimations rapportées des personnes vivant dans les zones rurales doit être calculé à partir de répondants des régions rurales seulement.

Source(s) des données

Le fichier final des données pondérées utilisé pour l'analyse doit être utilisé pour ces calculs.

Logiciel de calcul

Aucun logiciel spécial n'est requis. Le calcul de $Meff_w$ peut facilement être programmé en utilisant la sortie du logiciel d'analyse de l'enquête utilisé.

Calcul

En se concentrant sur la pondération finale (w_i) calculée pour chacun des n répondants auxquels elle $Meff_w$ s'applique, calculez la moyenne simple des pondérations pour obtenir \bar{w} , puis utilisez la formule,

$s_w^2 = \sum_{i=1}^n (w_i - \bar{w})^2 / (n - 1)$ pour calculer la variance des pondérations de l'échantillon. Nous pouvons

calculer $Meff_w$ à partir de l'ensemble final des pondérations de l'échantillon des répondants de la GATS, comme suit :

$$Meff_w = 1 + \frac{s_w^2}{\bar{w}^2}, \quad (1)$$

Notez que $Meff_w \geq 1$ puisque \bar{w}^2 et s_w^2 seront toujours positifs.

Interprétation

Alors qu'il est clairement préférable que $Meff_w$ soit aussi proche que possible de 1, $Meff_w > 2.00$ peut être considéré comme considérable et nécessitant, par conséquent, une certaine forme de mesures correctives. La réduction de la pondération ou les stratégies de censure sont le remède le plus fréquent

pour les pondérations trop variables (Potter, 1988). La décision finale de réduire la pondération dépend de l'obtention d'un équilibre entre l'avantage de la précision résultant de la réduction et de l'effet sur les estimations de l'enquête entraîné par les modifications de pondération. Si la réduction de la pondération diminue $Meff_w$, mais ne change pas sensiblement les estimations pondérées des résultats des mesures clés de l'étude, l'étape de réduction peut être justifiée, étant donné le temps nécessaire et les efforts à le faire.

Exemple

Pour un échantillon de la GATS, la moyenne des pondérations de l'échantillon final parmi les $X\ XXX$ répondants de l'enquête est $\bar{w} = YYY.YYYY$ et la variance simple est : $s_w^2 = ZZZZ.ZZ$. L'effet multiplicatif de la variable de pondération sur l'ensemble des estimations nationales d'un pays est

$$Meff_w = 1 + \frac{YYY.YYYY}{XXXX.ZZ} = U.UU$$

Un échantillon avec $Meff_w = U.UU < 2.00$ peut être considéré comme un échantillon dont la variation de pondération est suffisamment modeste pour ne pas nécessiter d'effort pour réduire l'effet de la variable de pondération sur les estimations, tel qu'une réduction de la pondération (voir Potter, 1988).

D.5 Effet global de la conception sur la précision des estimations de l'enquête et sur l'homogénéité intragroupe dans les PSU des estimations clés correspondantes de l'enquête

Contexte

Kish (1965), à l'origine, a défini un « effet de conception » désigné par 'deff' pour mesurer le changement multiplicatif de la variance d'une estimation de l'enquête due à l'utilisation de l'échantillonnage en grappes pour une conception d'échantillon donnée, par rapport à la variance d'une estimation comparable provenant d'un échantillon aléatoire simple de la même taille. Chacun des résultats des principales mesures de l'enquête GATS est une proportion (p), ou un taux exprimable sous la forme d'un pourcentage ou d'une moyenne. Le modèle original de Kish de l'effet de conception (dû à l'échantillonnage en grappes) d'une estimation (\hat{p}) de p est simplement,

$$Deff_{CS} = 1 + \rho_\theta(\bar{m} - 1) ,$$

où ρ_θ est une mesure de l'homogénéité relative intragroupe des répondants dans une PSU pour la mesure d'enquête utilisée pour estimer p (mesurant le degré de comparabilité des membres de la même grappe d'échantillon avec les membres en général dans la population), $\bar{m} = n/a$ est le nombre moyen de répondants par PSU de l'échantillon, n est la taille de l'échantillon global des répondants, et a est le nombre de PSU de l'échantillon.

Quand un échantillon de l'enquête implique des grappes d'échantillonnage et une pondération inégale de l'échantillon, l'effet global de la conception de l'échantillon sur une estimation de l'enquête, désignée par $Deff(\hat{\theta})$ (1999) est le produit de dû à la fois à l'échantillonnage en grappes (voir ci-dessus) et à l'effet

multiplicatif de la variable de pondération de l'échantillon (voir la discussion sur $Meff_w$) comme démontré par Gabler, et al., c-à-d.,

$$Deff(\hat{\theta}) \equiv \frac{V(\hat{p}; design, n)}{V(\hat{p}_{srs}; srs, n)}$$

$$= [Meff_w][Deff_{CS}(\hat{\theta})] = \left[1 + \frac{S_w^2}{\bar{w}^2}\right][1 + \rho_\theta(\bar{m} - 1)] . \quad (1)$$

Source(s) des données

Le fichier final des données pondérées utilisé pour l'analyse doit être utilisé pour ces calculs.

Logiciel de calcul

Les mêmes logiciels d'analyse d'enquête comme présentés dans le calcul de la marge d'erreur (voir l'**annexe D.6**).

Calcul

La variance estimée ou erreur type (c.-à-d., $v(\hat{p}; design, n)$ ou $\sqrt{v(\hat{p}; design, n)}$ dans l'équation 1, respectivement) est établie par le logiciel d'analyse d'enquête. Pour estimer ρ_θ nous pouvons simplement résoudre pour ρ_θ dans l'équation 1 et nous obtenons,

$$\rho_\theta = \frac{\frac{Deff(\hat{\theta})}{Meff_w} - 1}{\bar{m} - 1} \quad (2)$$

Pour estimer ρ_θ pour chaque estimation de l'enquête, nous pouvons calculer $Meff_w$ à partir de l'échantillon, nous connaissons \bar{m} par le nombre de PSU (a) et la taille de l'échantillon global des répondants (n) et pour de nombreux types de logiciels d'analyse d'enquête (par exemple, SUDAAN), nous pouvons demander une estimation de $Deff(\hat{\theta})$ ainsi que l'estimation ($\hat{\theta}$). Si une estimation de $Deff(\hat{\theta})$ n'est pas utilisable directement à partir du produit du logiciel en déterminant $Deff(\hat{\theta})$ tel que défini dans l'équation 1 ou ρ_θ comme le montre l'équation 2, alors pour les proportions estimées, nous pouvons estimer l'effet global de la conception comme suit :

$$deff(\hat{\theta}) = \frac{v(\hat{p}; design, n)}{v(\hat{p}_{srs}; srs, n)} = \frac{v(\hat{p}; design, n)}{\hat{p}(1 - \hat{p})/n} ,$$

puisque $v(\hat{p}_{srs}; srs, n) = \hat{p}(1 - \hat{p})/n$. L'estimation de ρ_θ peut alors être obtenue directement à partir de l'équation 2.

Interprétation

On espère généralement que l'effet global de la conception pour une estimation, ou $Deff(\hat{\theta})$, sera moins de 2,50. Les estimations de ρ_{θ} (de l'homogénéité dans une PSU) sont généralement de petits nombres positifs entre 0 et 1 (le plus souvent entre 0 et 0,05). Dans certains cas, les estimations de ρ_{θ} seront des nombres négatifs faibles.

Pour résumer les niveaux de $deff(\hat{\theta})$ parmi toutes les estimations de l'enquête GATS, il est recommandé d'estimer $Deff(\hat{\theta})$ pour plusieurs résultats de mesures clés de l'enquête (par exemple, la prévalence du tabagisme actuel, prévalence du tabagisme ancien et actuel, le taux de sevrage, etc.) par des domaines de rapports divers (par exemple, par sexe, âge, zones urbaine/rurale, région, et diverses classifications croisées de ces variables). Ensuite, pour chaque résultat de mesure clé, résumez les valeurs estimées de $Deff(\hat{\theta})$ parmi tous les domaines de rapports en rapportant les synthèses de statistiques suivantes pour l'estimation des valeurs $Deff(\hat{\theta})$: médiane, minimum et maximum.

Résumez les valeurs de $\hat{\rho}_{\theta}$ (médiane, minimum et maximum) parmi toutes les estimations d'enquête du GATS de la même manière que vous l'avez fait pour $deff(\hat{\theta})$.

Certains systèmes d'enquête ou organisations préfèrent rapporter l'effet de conception désigné par 'deft' plutôt que 'deff'. Deft est calculé pour chaque estimation par le rapport de l'erreur type utilisant la conception d'un échantillon donné sur l'erreur type qui se produirait si un échantillon aléatoire simple avait été utilisé.

D.6 Marge d'erreur pour les estimations clés de l'enquête

Contexte

Quelques définitions de base de mesures de précision statistique des échantillons de l'enquête sont nécessaires. Tout d'abord, notez que toutes les estimations à partir d'échantillons sont susceptibles de différer quelque peu de ce qui est estimé dans la population. Cette différence est généralement dénommée **erreur d'échantillonnage**. Des échantillons de probabilité comme ceux utilisés dans la GATS nous permettent de produire des mesures sommaires de l'erreur d'échantillonnage qui indiquent la précision des estimations. Il existe plusieurs mesures sommaires de la précision des estimations d'enquête. Pour définir ces mesures, utilisons le symbole, θ , pour désigner la caractéristique de la population que nous estimons (par exemple, le taux de prévalence des personnes qui fument actuellement des cigarettes). Nous utiliserons $\hat{\theta}$ pour représenter l'estimation de θ basée sur l'échantillon.

La première mesure sommaire fréquente de précision est la **variance de l'estimation**, qui s'écrit $V(\hat{\theta})$. Comme nous allons le voir, toutes les autres mesures sont d'une manière ou d'une autre liées à $V(\hat{\theta})$. La variance des estimations de l'enquête, et toutes les autres mesures connexes, sont des indicateurs quantitatifs de combien les estimations de l'échantillon pourraient varier parmi tous les échantillons possibles que la conception de l'échantillon utilisée dans l'enquête pourrait produire. Une deuxième mesure de la précision est l'**erreur type de l'estimation**, définie comme $SE(\hat{\theta}) = \sqrt{V(\hat{\theta})}$. L'erreur type

d'une estimation est donc simplement la racine carrée de la variance de cette estimation, et est calculée dans les mêmes unités que l'estimation $\hat{\theta}$. Une troisième mesure de la précision statistique est l'**erreur type relative** de l'estimation, définie comme $RSE(\hat{\theta}) = SE(\hat{\theta}) / \theta = \sqrt{V(\hat{\theta})} / \theta$. Puisque $RSE(\hat{\theta})$ mesure la précision par rapport à la taille de ce qui est estimé, elle n'a pas d'unité et est donc un indicateur de précision plus comparable parmi les estimations que $SE(\hat{\theta})$.

La mesure recommandée de la précision pour que les pays rendent compte des estimations clés de la GATS (par exemple, le taux de prévalence du tabagisme, etc.) est la **marge d'erreur**, définie comme $MOE(\hat{\theta}) = [Z][SE(\hat{\theta})]$, où Z est une mesure du niveau de confiance pour la mesure et $(SE(\hat{\theta}))$ est l'erreur type de $\hat{\theta}$. Pour la plupart des estimations nationales du GATS, nous pourrions utiliser $Z = 1.96$ pour un niveau de confiance de 95 %. La valeur estimée ($moe(\hat{\theta})$) rapportée pour $\hat{\theta}$ est interprétée comme suit :

Nous sommes sûrs à 95 % que la valeur déclarée ($\hat{\theta}$) se situe entre la valeur $moe(\hat{\theta})$ de θ .

Les analystes du GATS sont priés de signaler la valeur de $moe(\hat{\theta})$ pour toutes les estimations clés, y compris toutes les estimations nationales des principaux indicateurs du tabagisme. $moe(\hat{\theta})$ pour les estimations régionales correspondantes doit également être signalé si la conception de l'échantillon demande des tailles d'échantillon régional accrues pour faciliter la production d'estimations régionales qui répondront aux normes de précision de la GATS.

Comme les mesures de précision peuvent être facilement obtenues avec des logiciels statistiques, les pays du GATS sont fortement incités à demander régulièrement des estimations de la précision pour l'ensemble de leurs estimations produites, même si elles ne doivent être rapportées que pour des estimations d'envergure nationale (et éventuellement régionale). Les estimations dont $MOE(\hat{\theta})$ approche ou dépasse la taille de l'estimation elle-même doivent être signalées avec prudence, voire pas du tout. Il est également important de noter que tout comme $\hat{\theta}$, les estimations de $V(\hat{\theta})$, $SE(\hat{\theta})$, $MOE(\hat{\theta})$, ou $RSE(\hat{\theta})$ sont sujettes à des erreurs d'échantillonnage. Ceci est important, car toutes ces estimations peuvent être imprécises si elles sont fondées sur de petits échantillons, en particulier ceux impliquant moins de 10 à 20 PSU.

Source(s) des données

Le fichier final des données pondérées utilisé pour l'analyse doit être utilisé pour ces calculs.

Logiciel de calcul

Les estimations des caractéristiques de la population ($\hat{\theta}$) et leurs valeurs correspondantes de $moe(\hat{\theta})$ doivent tenir compte des caractéristiques statistiquement importantes de la conception de l'échantillon. Les caractéristiques estimées doivent être pondérées et les mesures de précision associées doivent permettre l'utilisation de la stratification, d'échantillonnage en grappes, d'échantillonnage sans remplacement, et de pondération de l'échantillon. Il est bien connu que le non-respect de cette recommandation contribue généralement à des estimations biaisées (en particulier pour la précision) et

par conséquent à des estimations d'intervalle et des tests de signification inappropriés. Pour cette raison, les analystes du GATS du pays sont fortement incités d'utiliser un logiciel d'analyse qui permet de tenir pleinement compte de la conception de l'échantillon utilisée pour produire les données de l'enquête. Cela signifie utiliser la pondération de l'échantillon pour produire toutes les estimations (de $\hat{\theta}$ et $moe(\hat{\theta})$) pour l'analyse descriptive. Cela signifie également utiliser un logiciel qui suit une approche largement acceptée pour estimer les variances et les erreurs types des estimations de l'enquête. Les trois principales approches de l'estimation de la variance sont la linéarisation par série de Taylor (TSL), les répliques répétées équilibrées (BRR), et le jackknife (JRR) (Wolter, 1985). La plupart des logiciels d'analyse d'enquête utilisent la linéarisation par série de Taylor (TSL) pour l'estimation de la variance, et donc pour des raisons de comparabilité entre les résultats des pays du GATS, un logiciel qui suit l'approche TSL est recommandé.

Plusieurs organismes statistiques du monde entier ont développé des logiciels pour analyser les données à partir d'échantillons complexes, comme ceux du GATS¹. Ces logiciels produiront non seulement des estimations de l'usage du tabac (c.-à-d., $\hat{\theta}$) mais ils peuvent aussi générer des estimations de la précision (c.-à-d., généralement $V(\hat{\theta})$ ou $SE(\hat{\theta})$) qui tiennent compte de manière appropriée des caractéristiques de conception clés de la GATS, à savoir l'échantillonnage en grappes, l'utilisation de la stratification, et la variation des probabilités de sélection (c.-à-d., la pondération de l'échantillon) :

Informations supplémentaires

Pour plus de détails sur les définitions données ci-dessus, consultez l'*Encyclopedia of Survey Research Methods*, (2007). Pour accéder à des informations en ligne sur l'ensemble des logiciels d'analyse d'enquête, visitez le site Web parrainé par la Survey Research Methods Section of the American Statistical Association (<http://www.hcp.med.harvard.edu/statistics/survey-soft/>). En plus de ces liens vers des logiciels d'analyse, le site de l'ASA-SRMS inclut également plusieurs autres liens et ressources utiles sur la conception de l'enquête et sa mise en œuvre (visitez <http://www.amstat.org/sections/SRMS/index.html>).

Calcul

Les produits du logiciel énumérés ci-dessus rendront compte d'une valeur de $\hat{\theta}$, ainsi que de sa variance estimée, notée $v(\hat{\theta})$, ou de son erreur type, $se(\hat{\theta})$. À partir de ces valeurs rapportées, on peut calculer la marge d'erreur estimée de $\hat{\theta}$, comme suit

$$moe(\hat{\theta}) = [t][se(\hat{\theta})] = [t]\sqrt{v(\hat{\theta})} \quad (1)$$

Interprétation

La valeur de $moe(\hat{\theta})$ rapportée pour $\hat{\theta}$ est interprétée comme suit :

Nous sommes sûrs à 95 % que la valeur estimée ($\hat{\theta}$) se situe entre la valeur $moe(\hat{\theta})$ de θ .

¹Le centre de coordination des données (DCC) de la GATS fournit une assistance technique pour l'utilisation des logiciels statistiques suivants : SAS, SPSS et STATA.

D.7 Estimations des erreurs d'échantillonnage

Les estimations d'un échantillon d'enquête sont affectées par deux types d'erreur : (1) les erreurs non dues à l'échantillonnage et (2) les erreurs d'échantillonnage. *Les erreurs non dues à l'échantillonnage* sont le résultat d'erreurs ou de fautes qui ne peuvent être attribuables à l'échantillonnage et qui ont été commises lors de la réalisation la collecte de données et du traitement des données, par exemple des erreurs de couverture, des erreurs de réponse, de non-réponse, des questionnaires erronés, des erreurs d'enregistrement de l'enquêteur, des erreurs de traitement de données, etc. Bien que de nombreux efforts aient été faits lors de la mise en œuvre du GATS en [pays] pour minimiser ces erreurs, les erreurs non dues à l'échantillonnage sont impossibles à éviter et difficiles à évaluer statistiquement.

L'échantillon des répondants sélectionnés pour le GATS en [pays] n'est que l'un des échantillons qui pouvaient être sélectionnés dans la même population, en utilisant la même conception et la même taille d'échantillon. Chacun de ces échantillons pourrait donner des résultats qui pourraient différer d'une manière ou d'une autre des résultats de l'échantillon réel sélectionné. *Les erreurs d'échantillonnage* sont une mesure de la variabilité entre tous les échantillons possibles. L'étendue de la variabilité n'est pas connue exactement, mais peut être estimée statistiquement à partir des résultats de l'enquête.

Les mesures suivantes des erreurs d'échantillonnage sont présentées pour chacun des indicateurs choisis :

- **Valeur (R)** : Estimation de la prévalence pondérée de l'indicateur
- **Erreur type (SE)** : Les erreurs d'échantillonnage sont généralement mesurées en termes d'erreurs types pour une estimation particulière ou d'un indicateur (R). L'erreur type d'une estimation est donc simplement la racine carrée de la variance de cette estimation, et est calculée dans les mêmes unités que l'estimation.
- **Taille de l'échantillon (n)** : Nombre total d'observations utilisées pour calculer l'estimation de la prévalence (R).
- **Effets de la conception** : L'effet de conception noté 'deff' est le rapport de la variance réelle d'un indicateur, selon la méthode d'échantillonnage utilisée dans l'enquête, sur la variance calculée sous l'hypothèse d'un échantillonnage aléatoire simple. La racine carrée de l'effet de conception noté 'deft' est utilisée pour monter l'efficacité de la conception de l'échantillon. Deft est calculé pour chaque estimation par le rapport de l'erreur type utilisant la conception d'un échantillon donné sur l'erreur type qui se produirait si un échantillon aléatoire simple avait été utilisé. Une valeur de DEFT de 1,0 indique que la conception de l'échantillon est aussi efficace qu'un échantillon aléatoire simple, alors qu'une valeur de DEFT supérieure à 1,0 indique l'augmentation de l'erreur type due à l'utilisation d'une conception de l'échantillon plus complexe. En général, pour une enquête bien conçue, DEFT varie entre 1 et 3. Il est commun, cependant, que DEFT soit beaucoup plus élevé, jusqu'à 7 ou 8.
- **Erreur type relative (RSE)** : L'erreur type relative aussi connue comme le coefficient de variation (CV) est le rapport de l'erreur type sur la valeur de l'indicateur.
- **Marge d'erreur (MOE)** : La marge d'erreur correspond à la mesure de la confiance désirée multipliée par l'erreur type de l'estimation. Le niveau de confiance est généralement basé sur une valeur (Z) de la distribution normale standard. Par exemple, pour un niveau de confiance de 95 %, nous pouvons utiliser $Z = 1,96$.

- **Les limites de confiance** ($R \pm 1,96SE$) sont calculées pour montrer l'intervalle dans lequel la vraie valeur pour la population peut raisonnablement se situer. Pour toute statistique donnée calculée à partir de l'enquête, la valeur de cette statistique se situera dans une fourchette de plus ou moins deux fois l'erreur type de la statistique dans 95 pour cent de tous les échantillons possibles de taille et de conception identiques.

Calcul de l'erreur type

Si l'échantillon de répondants avait été choisi comme un échantillon aléatoire simple, il aurait été possible d'utiliser des formules simples pour calculer les erreurs d'échantillonnage. Toutefois, l'échantillon de la GATS de [année] est le résultat d'une conception de stratifiée en plusieurs étapes, et, par conséquent, il était nécessaire d'utiliser des formules plus complexes. Pour le calcul des erreurs d'échantillonnage provenant des données de la GATS en [pays], [version du logiciel de statistique] a été utilisé. La méthode de linéarisation de Taylor d'estimation de la variance a été utilisée pour les estimations d'enquête qui sont des moyennes ou des proportions.

La méthode de linéarisation de Taylor traite tout pourcentage ou moyenne comme une estimation de rapport, $r = y/x$, où y représente la valeur totale de l'échantillon pour la variable y , et x représente le nombre total de cas dans le groupe ou sous-groupe en considération. La variance de r est calculée en utilisant la formule donnée ci-dessous :

$$SE^2(r) = \text{var}(r) = \frac{1-f}{x^2} \sum_{h=1}^2 \left[\frac{m_h}{m_h - 1} \left(\sum_{i=1}^{m_h} Z_{hi}^2 - \frac{Z_h^2}{m_h} \right) \right]$$

dans laquelle, $Z_{hi} = y_{hi} - rx_{hi}$, et $Z_h = y_h - rx_h$

où h (= 1 ou 2) représente la strate qui est urbaine ou rurale,
 m_h est le nombre total de PSUs sélectionnées dans la h e strate,
 y_{hi} est la somme des valeurs pondérées de la variable y dans la i e PSU dans la h e strate,
 x_{hi} est la somme du nombre pondéré de cas dans la i e PSU dans la h e strate, et
 f est la fraction d'échantillonnage global, qui est si petit qu'il est ignoré.

Les résultats sont présentés dans cette annexe pour le pays tout entier, pour les zones urbaines et rurales, et pour le sexe. Pour chaque variable ou indicateur, le type de statistique (moyenne, proportion ou taux) et la population de base sont donnés dans le **tableau D-5**. En plus de l'erreur type (SE) décrite ci-dessus, le **tableau D-6** comprend la valeur de l'estimation (R), la taille de l'échantillon, l'effet de conception (DEFF ou DEFT), l'erreur type relative (SE/R), la marge d'erreur (MOE) et les limites de confiance à 95 pour cent ($R \pm 1,96SE$), pour chaque variable ou indicateur.

Tableau D-5 : Liste des indicateurs pour les erreurs d'échantillonnage, GATS [pays] [année]

Indicateur	Estimation	Population de base
Utilisateurs actuels de tabac	Proportion	Adultes âgés de 15 ans et plus
Fumeurs actuels de tabac	Proportion	Adultes âgés de 15 ans et plus
Fumeurs actuels de cigarette	Proportion	Adultes âgés de 15 ans et plus
Utilisateurs actuels de tabac sans fumée	Proportion	Adultes âgés de 15 ans et plus
Utilisateurs quotidiens de tabac	Proportion	Adultes âgés de 15 ans et plus
Fumeurs quotidiens de tabac	Proportion	Adultes âgés de 15 ans et plus
Fumeurs de cigarettes quotidiens	Proportion	Adultes âgés de 15 ans et plus
Utilisateurs quotidiens de tabac sans fumée	Proportion	Adultes âgés de 15 ans et plus
Anciens utilisateurs quotidiens de tabac parmi tous les adultes	Proportion	Adultes âgés de 15 ans et plus
Anciens utilisateurs de tabac parmi tous les utilisateurs quotidiens de tabac, actuels et anciens	Proportion	Utilisateurs quotidiens de tabac anciens et actuels de 15 ans et plus
Anciens fumeurs quotidiens de tabac parmi tous les adultes	Proportion	Adultes âgés de 15 ans et plus
Anciens fumeurs de tabac parmi les fumeurs quotidiens anciens et actuels	Proportion	Fumeurs quotidiens de tabac anciens et actuels de 15 ans et plus
Anciens utilisateurs quotidiens de tabac sans fumée parmi tous les adultes	Proportion	Adultes âgés de 15 ans et plus
Anciens utilisateurs quotidiens de tabac sans fumée parmi les utilisateurs quotidiens de tabac sans fumée actuels et anciens	Proportion	Utilisateurs quotidiens de tabac sans fumée anciens et actuels de 15 ans et plus
Temps jusqu'à la première consommation de tabac dans les 5 minutes du réveil	Proportion	Utilisateurs quotidiens de tabac anciens de 15 ans et plus
Temps jusqu'à la première consommation de tabac dans les 6 à 30 minutes du réveil	Proportion	Utilisateurs quotidiens de tabac anciens de 15 ans et plus
Tentative d'arrêt du tabac à fumée dans les 12 derniers mois	Proportion	Les fumeurs actuels et anciens qui se sont abstenus de fumer pendant moins de 12 mois
Tentative d'arrêt du tabac sans fumée dans les 12 derniers mois	Proportion	Les utilisateurs actuels et anciens de tabac sans fumée qui ont été abstinents pendant au moins 12 mois
Professionnel de la santé a posé des questions sur le tabac à fumée	Proportion	Les fumeurs actuels et anciens qui ont été abstinents depuis moins de 12 mois et qui se sont rendus chez un professionnel de la santé au cours des 12 derniers mois
Professionnel de la santé a conseillé d'arrêter de fumer	Proportion	Les fumeurs actuels et anciens qui ont été abstinents depuis moins de 12 mois et qui se sont rendus chez un professionnel de la santé au cours des 12 derniers mois
Utilisation de pharmacothérapie pour cesser de fumer	Proportion	Les fumeurs actuels et anciens qui se sont abstenus de fumer pendant moins de 12 mois
Utilisation de séances de conseils ou de lignes téléphoniques d'aide à l'arrêt du tabac	Proportion	Les fumeurs actuels et anciens qui se sont abstenus de fumer pendant moins de 12 mois
Prévoit d'arrêter, pense à arrêter ou arrêtera de fumer	Proportion	Fumeurs actuels de 15 ans et plus
Exposition au tabagisme passif (SHS) à la maison	Proportion	Adultes âgés de 15 ans et plus
Exposition au tabagisme passif (SHS) sur le lieu de travail	Proportion	Adultes qui travaillent à l'intérieur

Indicateur	Estimation	Population de base
Exposition au tabagisme passif (SHS) dans les bâtiments/bureaux du gouvernement	Proportion	Adultes âgés de 15 ans et plus
Exposition au tabagisme passif (SHS) dans les établissements de soins de santé	Proportion	Adultes âgés de 15 ans et plus
Exposition au tabagisme passif (SHS) dans les restaurants	Proportion	Adultes âgés de 15 ans et plus
Exposition au tabagisme passif (SHS) dans les transports publics	Proportion	Adultes âgés de 15 ans et plus
Dernier achat de cigarette en magasin	Proportion	Fumeurs actuels de cigarettes fabriquées de 15 ans et plus
Dernier achat de cigarette auprès d'un vendeur ambulant	Proportion	Fumeurs actuels de cigarettes fabriquées de 15 ans et plus
Dernier achat de cigarette dans un kiosque	Proportion	Fumeurs actuels de cigarettes fabriquées de 15 ans et plus
Remarqué des informations antitabac à la radio ou la télévision	Proportion	Adultes âgés de 15 ans et plus
Remarqué des étiquettes d'avertissement sur la santé sur les paquets de cigarettes	Proportion	Fumeurs actuels de 15 ans et plus
Pense à arrêter en raison des mises en garde de santé sur les paquets de cigarettes	Proportion	Fumeurs actuels de 15 ans et plus
A remarqué des publicités ou des promotions de cigarettes	Proportion	Adultes âgés de 15 ans et plus
Estime que fumer du tabac peut provoquer une maladie grave	Proportion	Adultes âgés de 15 ans et plus
Estime que fumer du tabac peut causer des attaques	Proportion	Adultes âgés de 15 ans et plus
Estime que fumer du tabac peut causer des crises cardiaques	Proportion	Adultes âgés de 15 ans et plus
Estime que fumer du tabac peut causer un cancer du poumon	Proportion	Adultes âgés de 15 ans et plus
Estime que le tabac sans fumée peut provoquer une maladie grave	Proportion	Adultes âgés de 15 ans et plus
Estime que le tabagisme passif (SHS) peut provoquer une maladie grave chez les non-fumeurs	Proportion	Adultes âgés de 15 ans et plus
Nombre de cigarettes fumées par jour (par les fumeurs quotidiens)	Moyenne	Fumeurs quotidiens actuels de cigarettes de 15 ans et plus
Temps écoulé depuis l'arrêt de l'utilisation du tabac à fumée (en années)	Moyenne	Anciens fumeurs de 15 ans et plus
Dépenses mensuelles pour les cigarettes fabriquées	Moyenne	Fumeurs actuels de cigarettes fabriquées de 15 ans et plus
Âge au commencement de l'utilisation quotidienne du tabac à fumée	Moyenne	Fumeurs quotidiens anciens et actuels de 15 ans et plus

Tableau D-6 : Exemple de tableau pour le rapport des erreurs d'échantillonnage

Indicateur	Estimation (R)	Erreur type (SE)	Taille de l'échantillon (n)	Effet de conception (DEFF ou DEFT)	Erreur relative (SER)	Marge d'erreur (MOE)	Limites de confiance	
							Limite inférieure (R-1,96SE)	Limite supérieure (R+1,96SE)
Utilisateurs actuels de tabac								
Fumeurs actuels de tabac								
Fumeurs actuels de cigarette								
Utilisateurs actuels de tabac sans fumée								
Utilisateurs quotidiens de tabac								
Fumeurs quotidiens de tabac								
Fumeurs de cigarettes quotidiens								
Utilisateurs quotidiens de tabac sans fumée								
Anciens utilisateurs quotidiens de tabac parmi tous les adultes								
Anciens utilisateurs de tabac parmi tous les utilisateurs quotidiens de tabac, actuels et anciens								
Anciens fumeurs quotidiens de tabac parmi tous les adultes								
Anciens fumeurs de tabac parmi les fumeurs quotidiens actuels et actuels								
Anciens utilisateurs quotidiens de tabac sans fumée parmi tous les adultes								
Anciens utilisateurs quotidiens de tabac sans fumée parmi les utilisateurs quotidiens de tabac sans fumée actuels et anciens								
Temps jusqu'à la première consommation de tabac dans les 5 minutes du réveil								
Temps jusqu'à la première consommation de tabac dans les 6 à 30 minutes du réveil								
Tentative d'arrêt du tabac à fumée dans les 12 derniers mois								
Tentative d'arrêt du tabac sans fumée dans les 12 derniers mois								
Professionnel de la santé a posé des questions sur le tabac à fumer								
Professionnel de la santé a conseillé d'arrêter de fumer								
Utilisation de pharmacothérapie pour cesser de fumer								
Utilisation de séances de conseils ou de lignes téléphoniques d'aide à l'arrêt du tabac								

Indicateur	Estimation (R)	Erreur type (SE)	Taille de l'échantillon (n)	Effet de conception (DEFF ou DEFT)	Erreur relative (SE/R)	Marge d'erreur (MOE)	Limites de confiance	
							Limite inférieure (R-1,96SE)	Limite supérieure (R+1,96SE)
Prévoit d'arrêter, pense à arrêter ou arrêtera de fumer								
Exposition au tabagisme passif (SHS) à la maison								
Exposition au tabagisme passif (SHS) sur le lieu de travail								
Exposition au tabagisme passif (SHS) dans les bâtiments/bureaux du gouvernement								
Exposition au tabagisme passif (SHS) dans les établissements de soins de santé								
Exposition au tabagisme passif (SHS) dans les restaurants								
Exposition au tabagisme passif (SHS) dans les transports publics								
Dernier achat de cigarette en magasin								
Dernier achat de cigarette auprès d'un vendeur ambulancier								
Dernier achat de cigarette dans un kiosque								
Remarqué des informations anti-tabac à la radio ou la télévision								
Remarqué des étiquettes d'avertissement sur la santé sur les paquets de cigarettes								
Pense à arrêter en raison des mises en garde de santé sur les paquets de cigarettes								
A remarqué des publicités ou des promotions de cigarettes								
Estime que fumer du tabac peut provoquer une maladie grave								
Estime que fumer du tabac peut causer des attaques cardiaques								
Estime que fumer du tabac peut causer un cancer du poumon								
Estime que le tabac sans fumée peut provoquer une maladie grave								
Estime que la fumée passive peut provoquer une maladie grave chez les non-fumeurs								
Nombre de cigarettes fumées par jour (par les fumeurs quotidiens)								
Temps écoulé depuis l'arrêt de l'utilisation du tabac à fumer (en années)								

Indicateur	Estimation (R)	Erreur type (SE)	Taille de l'échantillon (n)	Effet de conception (DEFF ou DEFT)	Erreur relative (SE/R)	Marge d'erreur (MOE)	Limites de confiance	
							Limite inférieure (R-1,96SE)	Limite supérieure (R+1,96SE)
Dépenses mensuelles pour les cigarettes fabriquées Âge au commencement de l'utilisation quotidienne du tabac à fumée								

Remarques :

1. Les tableaux doivent être développés pour des sous-groupes (par exemple, le sexe, zone urbaine/rurale, la région et les groupes d'âge). Les groupes d'âge recommandés pour le GATS sont les suivants : 15 à 24, 25 à 44, 45 à 64 et 65 et plus.
2. Reportez-vous aux **Définitions des indicateurs du GATS** pour des définitions précises et des calculs des indicateurs suivants.
3. Les indicateurs fournis dans ce tableau sont produits à partir du questionnaire principal. Ceux-ci peuvent être adaptés à chaque pays au besoin. En outre, les pays peuvent souhaiter produire les estimations pour un sous-ensemble d'indicateurs, plutôt que pour tous, selon le cas.
4. Les pays peuvent souhaiter générer ' deff ' ou ' deff ' suivant la pratique standard applicable dans le pays/l'organisation.
5. Les tableaux ci-dessus peuvent être inclus dans l'annexe du rapport national, comme estimations des erreurs d'échantillonnage pour décrire la qualité des données.

D.8 Taux de couverture de la base des foyers

Contexte

Un élément clé du processus d'échantillonnage dans une enquête auprès des foyers est le dénombrement des foyers au cours de la phase d'échantillonnage lorsque les adresses des foyers sont choisies. Différentes approches de dénombrement des foyers peuvent être utilisées, et le résultat final ordinaire est un sous-dénombrement net des foyers et donc une couverture incomplète de l'échantillon final des foyers. Comme l'effet de biais du sous-dénombrement dépend des facteurs suivants : (i) la différence cumulée entre les mesures clés de l'étude pour les foyers couverts par la base de sondage et ceux non couverts, et (ii) l'ampleur de la non-couverture (ou 100 % moins le taux de couverture des foyers) dans l'échantillon, il est utile d'estimer le taux de couverture de la base de sondage pour les foyers. Un taux de couverture de la base (FCR) est défini comme,

$$FCR = \frac{N_F}{N} = \frac{\# \text{ of households accounted for by the household frame}}{\# \text{ of households in the population}} \quad (1)$$

Source(s) des données

Deux sources de données seront nécessaires pour ces calculs. L'un est le fichier final de données pondérées utilisées pour l'analyse, qui est utilisé pour estimer N_F dans l'équation 1. L'autre est la meilleure source disponible de dénombrement des foyers à l'échelle du pays, qui est utilisée comme mesure de N dans l'équation 1. Cette dernière source peut être le plus récent recensement, les plus récentes mises à jour du dernier recensement en utilisant des méthodes démographiques, ou une récente enquête nationale dont la taille de l'échantillon et la qualité sont au moins de la même ampleur que celle de l'enquête réalisée pour la GATS.

Logiciel de calcul

Aucun en particulier puisque les calculs nécessaires peuvent facilement être programmés dans n'importe quel logiciel.

Calcul

Le nombre de foyers dénombrés pour la base des foyers (N_F) peut être estimé par l'une des deux façons ci-dessous. Toutes deux nécessitent certaines des probabilités de sélection spécifiques à l'étape qui ont été utilisées pour calculer la pondération de base de chaque répondant. L'une est une estimation pondérée de N_F en utilisant une pondération de base pour l'échantillonnage de « segments » de la conception de l'échantillon (*l'approche de la somme pondérée*), l'autre est une somme de pondération de base pour tous les foyers sélectionnés dans l'échantillon de la GATS, ce qui est aussi une estimation de N_F (*l'approche de la somme pondérée des foyers*).

Approche de la somme pondérée – Dans la plupart des conceptions de la GATS, les segments seront les unités secondaires d'échantillonnage (SSU), mais en général ils sont les unités d'échantillonnage de la zone dans laquelle les foyers sont dénombrés pour produire les bases d'échantillonnage à partir desquelles les foyers sont choisis. En supposant que le segment est la SSU, définissez W_j comme la

pondération de base pour le j -e segment de l'échantillon, calculé comme celui divisé par le produit des probabilités de sélection pour la PSU dans laquelle le segment est situé multiplié par la probabilité de sélection du segment intra-PSU. Si M_j désigne le nombre de foyers dénombrés dans le j -e segment de l'échantillon, alors l'estimation de N_F en vertu de cette approche est alors calculée comme étant,

$$\hat{N}_F^{(i)} = \sum_j^{\text{All sample segments}} W_j M_j \quad (2)$$

Approche de la somme de la pondération des foyers - Dans cette approche, on utilise la pondération de la base pour chaque foyer sélectionné. L'ensemble des foyers sélectionnés comprend ceux qui sont affectés sur le terrain pour le recrutement des foyers pendant le recueil des données de la GATS. W_{jk} désigne la pondération de base pour le k -e foyer sélectionné dans le j -e segment de l'échantillon. W_{jk} est W_j divisé par la probabilité de sélection du k -e foyer au sein du segment. L'estimation de N_F suivant cette approche est alors calculée comme suit :

$$\hat{N}_F^{(ii)} = \sum_{j,k}^{\text{All selected households}} W_{jk} \quad (3)$$

Les meilleures sources pour une valeur appropriée du nombre total réel de foyers dans la population (N) sont les suivantes : (1) le recensement de la population le plus récent (à condition qu'il soit généralement considéré comme de haute qualité), ou (2) un nombre de foyers estimé pour N_F (obtenu en utilisant les équations 2 ou 3) à partir d'un échantillon d'une enquête récente et statistiquement supérieure (par exemple, l'échantillon maître de la GATS à partir duquel l'échantillon a été aléatoirement subsélectionné).

Interprétation

Une estimation de (FCR) de 95 % ou plus indique que la qualité du dénombrement des foyers dans l'échantillon de la GATS est d'une qualité acceptable.



GLOBAL TOBACCO SURVEILLANCE SYSTEM (GTSS)

