

50x2030

DATA-SMART AGRICULTURE

GUIDE D'ÉCHANTILLONNAGE

JUILLET 2021

SÉRIE DE DOCUMENTS TECHNIQUES 50x2030

Contributions

Ce document est un produit de l'Initiative 50x2030, dirigée par Dramane Bako (FAO) avec des contributions de Marcello D'Orazio (Institut national italien de statistique), Flavio Bolliger (FAO) et d'autres membres des équipes de l'Organisation des Nations Unies pour l'alimentation et l'agriculture et de la Banque mondiale travaillant sur l'Initiative. Ce document a été examiné et approuvé par le Groupe consultatif technique de l'Initiative, avec des remerciements particuliers à Samuel Bénin (IFPRI), Cheryl Doss (Université d'Oxford), Mercy Kanyuka (Statistiques Malawi), Miet Maertens (Université de Louvain), Pedro Luis do Nascimento Silva (ancien président de l'Institut international de statistique), Jose Molinas (Institut de développement) et Aberash Tariku (Agence centrale de statistique d'Éthiopie) pour leurs contributions. Les auteurs remercient également les experts ayant mené les revues externes, Emily Berg (Iowa State University), Cristiano Ferraz (Universidade Federal de Pernambuco) et Jon Kastelic (Banque mondiale), pour leurs précieuses contributions.

L'Initiative 50x2030 pour combler les lacunes en matière de données agricoles est un effort de plusieurs organismes visant à soutenir 50 pays à faible et à plus moyen revenu pour produire des données agricoles et rurales fondamentales par l'utilisation d'enquêtes agricoles et rurales intégrées. Pour en savoir plus sur l'Initiative, veuillez visiter www.50x2030.org.

Table des matières

Table des matières	2
I. Introduction.....	4
II. Aperçu des objectifs de mesure et des programmes d'enquête.....	4
II.A. Objectifs du travail de mesure	4
II.B. Programmes d'enquête intégrés 50x2030.....	5
III. Populations concernées	5
III.A. Ménages.....	6
III.B. Exploitations agricoles.....	7
IV. Échantillonnage des cadres.....	9
IV.A. Présentation d'une enquête à base de sondages multiples	9
IV.B. Cadre du Programme intégré d'enquêtes agricoles et rurales	10
IV.C. Cadre du Programme d'enquête agricole	12
V. Plans d'échantillonnage	13
V.A. Plan d'échantillonnage pour le Programme intégré d'enquêtes agricoles et rurales	14
V.B. Plan d'échantillonnage pour le Programme d'enquête agricole	15
VI. Taille de l'échantillon	17
VI.A. Taille de l'échantillon pour le secteur des ménages	18
VI.B. Taille de l'échantillon pour le secteur ne relevant pas des ménages et les exploitations spéciales.....	21
VII. Stratification	21
VII.A. Stratification et répartition des UPE	21
VII.B. Stratification et répartition intra-UPE	23
VIII. Estimation.....	24
VIII.A. Estimateurs et variances	24
VIII.B. Différence entre les unités d'échantillonnage et d'observation dans le secteur des ménages	27
VIII.C. Pondération à l'aide de différentes unités d'observation	28
VIII.D. Estimation pour les cultures rares	29
IX. Sous-échantillonnage et estimation.....	29
IX.A. Taille du sous-échantillon.....	29
IX.B. Sous-échantillonnage pour des modules/questionnaires spécifiques	30
IX.C. Cas des coupes-témoins.....	31
IX.D. Cas de pertes après récolte au niveau de l'exploitation	31
IX.E. Estimation avec sous-échantillonnage.....	32
X. Couverture des exploitations spéciales (commerciales, grandes, modernes, etc.) et estimation sur base de sondage double.....	33
X.A. Approche de filtrage.....	34
X.B. Estimateurs à base double	35
XI. Modèle de collecte de données longitudinales et procédures de suivi	38
XI.A. Longueur du panel/plan de rotation	39
XI.B. Traitement de l'attrition des échantillons.....	39
XII. Intégration des systèmes statistiques existants	41
XII.A. Il existe des enquêtes sur l'agriculture et les ménages dans le pays.....	41
XII.B. Seul un système d'enquête agricole existe	42
XII.C. Seule une enquête auprès des ménages existe	42
XIII. Références.....	43

I. Introduction

L'Initiative 50x2030 pour Comblent les Lacunes en matière de Données Agricoles vise à renforcer les systèmes de données nationaux afin qu'ils soient mieux équipés pour répondre aux demandes de données liées aux systèmes statistiques et aux obligations de déclaration au niveau mondial, régional et national. En particulier, les pays qui adopteront l'approche d'enquête 50x2030 seront favorablement à même de produire des statistiques officielles par l'utilisation d'une méthodologie solide et l'élaboration de rapports sur les indicateurs essentiels des Objectifs de Développement Durable (ODD) liés à l'agriculture. Ils seront également mieux à même de comprendre les moteurs de la productivité et des revenus agricoles, ainsi que leurs liens avec la qualité de vie et le développement en milieu rural. L'approche 50x2030 intègre la collecte de données sur les caractéristiques de base du secteur agricole (y compris les chiffres de production annuelle) à un ensemble plus large de données sur les facteurs économiques, environnementaux et sociaux pertinents pour les zones rurales.

L'Initiative 50x2030 propose des programmes d'enquêtes modulaires et intégrées dans le but de suivre et de comprendre les systèmes agricoles. Les données agricoles clés, c'est-à-dire de production, seront collectées sur une base annuelle, tandis que des données agricoles plus approfondies seront collectées tous les trois ans. Le système s'appuie sur l'expérience du Programme d'enquêtes agricoles intégrées (AGRIS) de la FAO et du Programme d'études sur la Mesure des Niveaux de Vie - Enquêtes intégrées sur l'agriculture (RNA-NVM) de la Banque mondiale. Comme dans le cas de ces programmes, le système deviendra partie intégrante des systèmes statistiques nationaux.

Comme pour tout système d'enquête polyvalent, le plan d'échantillonnage doit être soigneusement élaboré pour produire des estimations fiables de manière rentable afin de réaliser les objectifs de mesure de l'Initiative 50x2030. Le présent document traite des principales caractéristiques techniques d'un modèle d'échantillonnage approprié pour les programmes d'enquête proposés par l'Initiative. Après avoir défini les populations concernées, il sera ensuite discuté de l'élaboration de cadres ou bases d'échantillonnage, des critères de stratification, des calculs de taille de l'échantillonnage, des procédures d'estimation et des approches d'échantillonnage au fil du temps. La dernière section porte sur les questions d'échantillonnage liées aux programmes d'enquête existants dans les pays qui adoptent le système d'enquête de l'Initiative.

II. Aperçu des objectifs de mesure et des programmes d'enquête

II.A. Objectifs du travail de mesure

L'Initiative 50x2030 donne la priorité à l'estimation des statistiques nationales et des indicateurs critiques des ODD liés à l'agriculture dans les pays concernés. En particulier, l'Initiative se concentre sur l'ODD 2 (Faim zéro), l'ODD 5 (Égalité des sexes) et sur la collecte de données pour le calcul de quatre indicateurs ODD hautement prioritaires :

- 2.3.1 - Volume de production par unité de main d'œuvre, par catégorie de taille d'exploitation agricole/pastorale/forestière ;
- 2.3.2 - Revenu moyen des petits producteurs de denrées alimentaires, par sexe et statut autochtone ;
- 2.4.1 - Proportion de la superficie agricole dans le cadre d'une agriculture productive et durable ;

- 5.a.1 - a) Proportion de la population agricole totale détenant des droits de propriété ou des droits garantis sur les terres agricoles, par sexe ; b) Part des femmes parmi les propriétaires ou les titulaires de droits sur les terres agricoles, par type d'occupation.

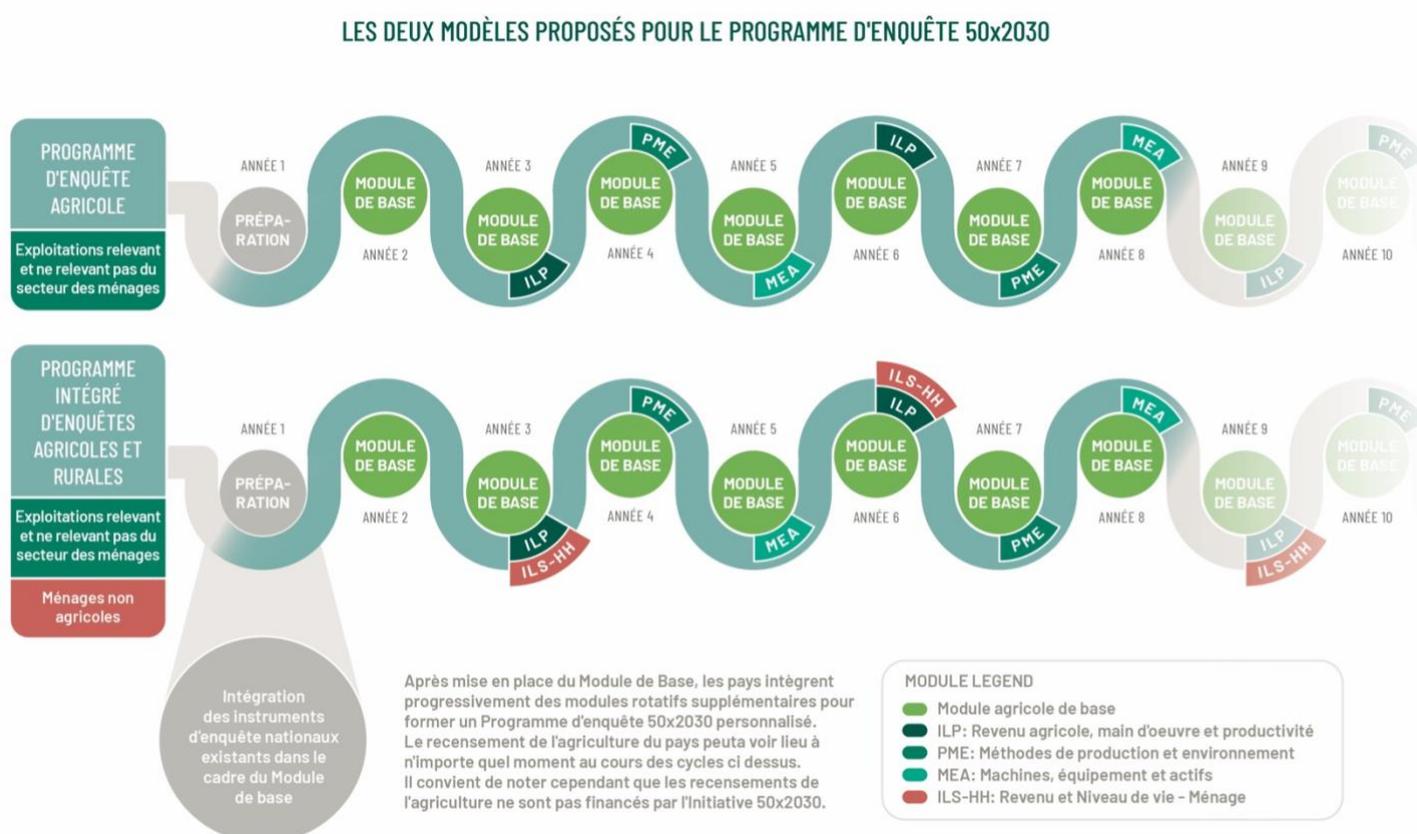
Dans le cadre du programme d'enquête, il peut également être nécessaire de tenir compte des indicateurs supplémentaires requis au niveau régional ou international. Par exemple, les pays africains sont tenus de communiquer les indicateurs du Programme détaillé de développement de l'agriculture africaine (PDDAA). Les instruments d'enquête promus dans le cadre de l'Initiative 50x2030 visent à traiter plusieurs indicateurs du PDDAA et peuvent être adaptés et élargis pour inclure des indicateurs prioritaires nationaux, ainsi que des indicateurs supplémentaires des ODD (Initiative 50x2030, 2020). La couverture d'exigences de mesure aussi larges dans le cadre d'une opération d'enquêtes est un enjeu de taille que le plan d'échantillonnage doit permettre de régler.

II.B. Programmes d'enquête intégrés 50x2030

Deux programmes d'enquête sont appuyés par l'Initiative : le Programme d'enquêtes agricoles (programme agricole), qui ne comporte que des volets agricoles et le Programme intégré d'enquêtes agricoles et rurales (programme intégré), qui intègre les volets agricoles à un volet d'enquêtes sur les ménages. La Figure 1 illustre des exemples de ces deux programmes d'enquête, qui peuvent toutefois être modifiés en fonction des besoins des pays.

Le Programme d'enquête agricole 50x2030 est un système d'enquête modulaire. Il est doté d'un outil d'enquête de base annuel axé sur les cultures, l'élevage, l'aquaculture, la pêche et la production forestière (CORE) et d'un ensemble d'outils spécialisés portant sur des sujets tels que le revenu agricole, la main d'œuvre et la productivité (ILP), les méthodes de production et les aspects environnementaux de l'agriculture (PME), ainsi que les machines, équipements et actifs agricoles (MEA). Ces outils spécialisés sont administrés à des fréquences plus basses, comme illustrés dans la Figure 1 (ligne du haut). D'autres outils spécialisés peuvent être ajoutés en fonction des besoins et de la demande des pays.

FIGURE 1. EXEMPLES DES DEUX MODÈLES PROPOSÉS POUR LES PROGRAMMES D'ENQUÊTE 50x2030



par le programme agricole couvre toutes les exploitations agricoles (dans le secteur des ménages et dans les autres secteurs), tandis que la population concernée par le programme intégré combinant une enquête auprès des ménages avec une enquête agricole sur les exploitations agricoles, couvre tous les ménages dans les zones rurales et toutes les exploitations agricoles, comme illustré dans la Figure 2.

FIGURE 2. COUVERTURE ET POPULATIONS CONCERNÉES DU : PROGRAMME D'ENQUÊTE AGRICOLE ET RURALE INTÉGRÉE EN 50X2030

	Secteur agricole		Secteur non agricole
	Exploitations ne relevant pas du secteur des ménages	Ménages agricoles	Ménages non-agricoles
Zones rurales			
Zones urbaines			

	Programme d'enquête Agricole
	Programme intégré d'enquêtes agricoles et rurales

III.A. Ménages

L'Initiative tiendra compte, à titre de référence, de la définition des ménages établie par le Programme mondial de recensement de la population et des logements 2020 (Nations Unies, 2017) :

« Le ménage est défini en fonction de la manière dont les personnes pourvoient, individuellement ou en groupe, à leurs besoins alimentaires et à leurs autres besoins vitaux. Un ménage peut être : a) soit un ménage composé d'une seule personne, c'est-à-dire une personne qui pourvoit à ses propres besoins alimentaires et autres besoins vitaux sans s'associer avec d'autres personnes pour former un ménage multiple ; b) soit un ménage multiple, c'est-à-dire un groupe de deux ou plusieurs personnes qui, vivant au même foyer, pourvoient en commun à leurs besoins alimentaires et autres besoins vitaux. Les membres du groupe peuvent mettre plus ou moins leurs revenus en commun et avoir un budget unique ; le groupe peut se composer soit de personnes apparentées, soit de personnes non apparentées, soit d'une combinaison des deux catégories.

Si les définitions nationales diffèrent de la définition des Nations Unies, l'Initiative aidera les pays à déterminer les conséquences de l'utilisation d'une définition différente et les aidera à se rapprocher de la définition recommandée.

III.B. Exploitations agricoles

La définition adoptée pour les exploitations agricoles est tirée du Programme mondial de la FAO pour le Recensement de l'agriculture (WCA) 2020 (FAO, 2015a) :

« Une exploitation agricole est une unité économique de production agricole soumise à une direction unique et comprenant tous les animaux qui s'y trouvent et toute la terre utilisée entièrement ou en partie pour la production agricole, indépendamment du titre de possession, du mode juridique ou de la taille. La direction unique peut être exercée par un particulier, par un ménage, conjointement par deux ou plusieurs particuliers ou ménages, par un clan ou par une tribu, ou par une personne morale telle que société, entreprise collective, coopérative ou organisme d'État. L'exploitation peut contenir un ou plusieurs blocs, situés dans une ou plusieurs régions distinctes ou dans une ou plusieurs divisions territoriales ou administratives, à condition qu'ils partagent les mêmes moyens de production tels que la main-d'œuvre, les bâtiments agricoles, les machines ou animaux de trait utilisés pour l'exploitation. »

Le Recensement de l'agriculture (WCA) distingue deux types d'exploitations agricoles : i) les exploitations du secteur des ménages et ii) les exploitations ne relevant pas du secteur des ménages.

III.B.1. Exploitations du secteur des ménages

D'une manière générale, dans le secteur des ménages, une exploitation agricole correspond généralement à un ménage ayant des activités de production agricole pour compte propre (soit pour la vente soit pour son propre usage). Dans certains pays, un seuil (limite de taille minimale) est adopté pour définir les exploitations agricoles (par exemple la superficie des terres agricoles exploitées ou le nombre d'animaux élevés).

Les ménages exploitant des exploitations agricoles pour leur propre compte sont souvent appelés « ménages de production agricole », ou simplement « ménages agricoles ». Comme il est mentionné dans le Recensement de l'agriculture (WCA), il existe habituellement une « correspondance un à un » entre un ménage agricole et une exploitation agricole. En d'autres termes, toutes les activités de production agricole pour compte propre des membres d'un ménage agricole donné sont généralement entreprises sous une seule gestion. Il est inhabituel que différents membres du ménage exploitent des terres agricoles ou du bétail de façon indépendante ; ils ont ordinairement tendance à mettre en commun les revenus tirés de ces activités. Même si les activités agricoles des différents membres d'un ménage sont relativement indépendantes, les revenus ou les produits générés par chacun d'entre eux sont généralement mis en commun.

Cependant, bien qu'inhabituels, la FAO (2015a) mentionne deux cas particuliers où les concepts d'exploitation agricole et de ménage peuvent diverger :

- Lorsqu'un ménage est constitué de deux ou plusieurs unités, par exemple lorsqu'un couple marié vit sous le même toit que les parents, les deux unités peuvent gérer les terres de manière indépendante mais, en tant que membres du même ménage, pourvoir ensemble aux besoins alimentaires et mettre leurs revenus en commun. Dans ce cas, il n'y a qu'un seul ménage mais deux exploitations distinctes.
- En plus de ses propres activités individuelles de production agricole, un ménage peut exploiter des terres ou élever des animaux conjointement avec un autre ménage ou groupe de ménages. Dans ce cas, deux unités de production agricole sont associées au ménage, avec deux jeux d'activités : i) les activités de production agricole du ménage individuel lui-même, et ii) les opérations agricoles effectuées conjointement avec le (les) autre(s) ménage(s).

La section VII.B traite de la façon de traiter ces cas au cours de la collecte et de l'estimation des données.

III.B.2. Exploitations ne relevant pas du secteur des ménages

Le Recensement de l'agriculture (WCA) considère les exploitations ne relevant pas du secteur des ménages comme des exploitations agricoles exploitées par des entités telles que des sociétés, des institutions gouvernementales, des coopératives, etc. Il est essentiel de définir clairement les exploitations ne relevant pas du secteur des ménages par rapport aux exploitations du secteur des ménages, car cela a des incidences directes sur le plan d'échantillonnage. Dans certains cas, il peut y avoir confusion entre le secteur des ménages et le secteur ne relevant pas des ménages, en particulier en ce qui concerne les ménages qui exploitent de grandes exploitations commerciales modernes qui pourraient être classées comme des quasi-sociétés dans le Système de comptabilité nationale. À condition que les registres des établissements soient utilisés parmi les principales sources d'échantillonnage pour les exploitations ne relevant pas du secteur des ménages, la FAO recommande d'utiliser ces registres comme critère de classification des quasi-sociétés dans le secteur hors ménages. Par conséquent, les exploitations du secteur ne relevant pas du secteur des ménages pourraient être considérées comme des exploitations agricoles exploitées par :

- Des sociétés au sens défini par le Système de comptabilité nationale.
- Des exploitations d'État (entités de production agricole gérées par une administration centrale ou locale directement ou par l'intermédiaire d'un organisme spécial).
- Des coopératives.
- Des exploitants institutionnels tels que les hôpitaux, les écoles, les prisons, les institutions religieuses, etc.
- Des organismes sans but lucratif.
- Des quasi-sociétés enregistrées : les ménages ayant des exploitations agricoles importantes/modernes ou spécifiques dont les revenus et les dépenses provenant des activités agricoles peuvent être séparés des autres activités du ménage, et dont les opérations sont reflétées dans le registre national des établissements.

Les sociétés et les institutions gouvernementales peuvent avoir des structures complexes, dans lesquelles différentes activités sont entreprises par différentes parties de l'organisation. Dans certains cas, les sociétés peuvent être composées de différents « établissements » qui pourraient constituer des exploitations agricoles différentes. Le Recensement de l'agriculture (WCA) recommande de s'appuyer sur le concept d'établissement utilisé dans la comptabilité nationale, selon lequel un établissement est une unité économique engagée dans une seule activité de production principale et opérant dans une localité unique (FAO, 2015a). Par exemple, une société qui possède deux établissements distincts situés dans des localités différentes serait considérée comme deux exploitations agricoles différentes.

III.B.3. Exploitations spéciales

Certains pays peuvent avoir un intérêt particulier pour des exploitations spéciales (exploitations commerciales, grandes exploitations, exploitations modernes, exploitations spécialisées, etc.) qui pourraient appartenir à la fois au secteur des ménages et au secteur ne relevant pas des ménages. Dans ce cadre, il serait important d'élaborer une définition claire de ces exploitations agricoles à prendre en considération dans les procédures d'échantillonnage et d'estimation. Les procédures d'estimation comportant deux cadres superposés (échantillonnage à base double) sont abordées à la section VII.E.

IV. Échantillonnage des cadres

Un cadre d'échantillonnage de qualité est nécessaire pour produire des estimations fiables. Dans le cadre des enquêtes sur l'agriculture, la FAO (2017) mentionne les problèmes suivants avec les cadres d'échantillonnage qui sont souvent rencontrés et doivent être évités :

- Sous-couverture ou cadre incomplet : non-inclusion de certaines exploitations dans le cadre d'échantillonnage.
- Sur-couverture : certaines unités qui ne sont pas des exploitations agricoles sont incluses dans le cadre. Par exemple, les exploitations qui sont énumérées dans le cadre mais qui n'existent plus.
- Multiplicité : certaines exploitations font l'objet de doublons, ce qui augmente la probabilité de leur inclusion dans l'échantillon.
- Clusters (groupes sectoriels) d'éléments : certaines exploitations du cadre sont en fait des clusters d'exploitations plutôt que des exploitations individuelles. Cela est généralement dû à une mauvaise compréhension de la définition de l'exploitation par certains recenseurs, qui peuvent répertorier deux ou plusieurs exploitations comme une exploitation unique.

Les projets de recherche sur les cadres d'échantillonnage pour les statistiques agricoles, mis en œuvre par la FAO dans le cadre de la Stratégie mondiale pour l'amélioration des statistiques agricoles et rurales, ont souligné l'importance de l'utilisation de cadres d'échantillonnage pour la rentabilité, la cohérence et l'intégration des statistiques agricoles dans les pays. Un cadre directeur d'échantillonnage est un cadre qui permet de sélectionner différents échantillons (y compris ceux provenant de différents modèles d'échantillonnage) à des fins spécifiques : enquêtes agricoles, enquêtes sur les ménages et enquêtes sur la gestion des exploitations agricoles. Un tel cadre permet de tirer des échantillons pour plusieurs enquêtes différentes ou différentes séries d'une même enquête, ce qui permet d'éviter de créer un cadre *ad hoc* pour chaque enquête (FAO, 2015b).

IV.A. Présentation d'une enquête à base de sondages multiples

De façon générale, une enquête à base de sondages multiples fait référence à des enquêtes où deux cadres ou plus sont utilisés et où des échantillons indépendants sont choisis dans chaque cadre. Plus précisément, l'utilisation de deux cadres d'échantillonnage pour une enquête échantillon est appelée « enquête à base de sondage double ». Dans la littérature, l'utilisation de plusieurs cadres dans une enquête est motivée par diverses raisons, notamment :

- **Absence de cadre couvrant l'ensemble de la population concernée** : il y a des situations où il est difficile (ou parfois impossible pour certaines populations) d'obtenir un cadre avec une couverture complète. Une solution consiste à utiliser plusieurs cadres sur les mêmes populations pour améliorer la couverture, en exploitant les forces et en compensant les faiblesses de chaque type de cadre.
- **Production de statistiques fiables sur la sous-population ou les populations rares et difficiles à atteindre** selon l'intérêt du pays.

Les avantages associés à l'utilisation de bases multiples sont les suivants : i) économies découlant du sous-échantillonnage de sondages coûteux et du suréchantillonnage de sondages moins coûteux (par exemple, base de sondage aréolaire et listes) et ii) souplesse accrue dans la conception de l'enquête pour mieux contrôler les coûts de sondage, la couverture, les taux de réponse, l'exactitude (utilisation de différents plans d'échantillonnage pour différentes bases) ; et l'utilisation de différents modes de collecte de données (p. ex., interviews en personne, par téléphone et sur le Web). Dans le cadre d'une enquête sur l'agriculture, cette

approche facilite i) la couverture de toutes les exploitations agricoles, tant dans le secteur des ménages que dans le secteur ne relevant pas des ménages, ii) l'utilisation du modèle d'échantillonnage le plus approprié recommandé pour chaque type d'exploitation, et iii) la production de statistiques fiables sur les exploitations agricoles spéciales qui intéressent les pays. Cependant, les procédures d'estimation peuvent être assez complexes, en particulier lorsque le nombre de bases devient élevé (généralement plus de trois).

Il existe deux exigences importantes pour l'utilisation d'enquêtes à bases multiples :

- **Exhaustivité** : l'union de toutes les bases doit assurer une couverture complète de la population cible. De cette façon, chaque élément doit être recensé dans au moins une des bases.
- **Identifiabilité** : pour toute unité échantillonnée, il doit être possible de comprendre si elle appartient ou non à l'une des bases.

Lorsqu'il n'y a pas de recoupement, les procédures d'estimation sont simples, car des échantillons indépendants sont choisis dans chaque base pour la mise en œuvre de l'enquête. En présence de recoupement, les méthodes proposées dans la littérature pour l'estimation à base de sondage double sont discutées ci-dessous.

Les travaux méthodologiques sur la stratégie d'échantillonnage de l'Enquête intégrée sur l'agriculture (AGRIS) ont recommandé deux types de cadres d'échantillonnage principaux pour les enquêtes agricoles intégrées (FAO, 2017) :

1. Une base de sondage multiple composée de deux bases de type liste : énumérer les exploitations agricoles dans i) le secteur des ménages et ii) le secteur ne relevant pas des ménages.
2. Une base de sondage multiple composée d'une base aréolaire et des deux bases de type liste suivantes pour i) les exploitations sans terre d'élevage de bétail et ii) les grandes exploitations agricoles commerciales.

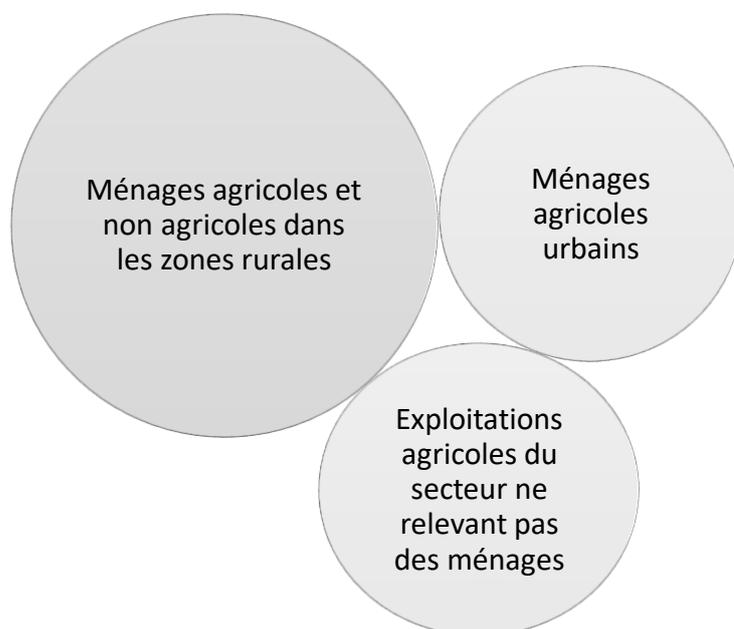
Les deux recommandations en matière de bases de sondage multiples sont valables pour le Programme d'enquête agricole (sans l'échantillon des ménages ruraux non agricoles) de l'Initiative 50x2030. Toutefois, pour le Programme intégré d'enquêtes agricoles et rurales, la première recommandation (à combiner avec une base de ménages ruraux) est la plus appropriée. En effet, elle répond à la nécessité de sélectionner un échantillon représentatif de ménages (indépendamment de l'engagement agricole) dans les zones rurales. L'utilisation d'une base aréolaire permettra de constituer un échantillon indirect de ménages agricoles sans garantie de représentativité quant aux types de ménages couverts (de plus amples détails sont fournis dans la section sur les méthodes d'échantillonnage). Toutefois, afin d'améliorer encore l'estimation de la superficie et de la production agricoles, une base aréolaire peut être associée à la liste des ménages dans le cas d'une méthode à base de sondages multiples.

IV.B. Cadre du Programme intégré d'enquêtes agricoles et rurales

Le cadre du Programme intégré d'enquêtes agricoles et rurales devrait couvrir les ménages ruraux (indépendamment de l'engagement agricole) et toutes les exploitations agricoles du pays. En conséquence, un cadre idéal comprendrait des listes complètes des ménages ruraux et de toutes les exploitations agricoles, tant dans le secteur des ménages que dans le secteur ne relevant pas des ménages. Dans le secteur des ménages, les exploitations agricoles sont généralement échantillonnées indirectement par le biais des ménages agricoles. En effet, l'élaboration directe de listes des exploitations agricoles pose des problèmes pratiques dans la plupart des cas, car les exploitations peuvent comporter de nombreuses parcelles situées à différents endroits. Par conséquent, un cadre directeur d'échantillonnage approprié est une base de sondages multiples composée de (i) la liste de tous les ménages agricoles et non agricoles dans les zones

rurales ; ii) la liste des ménages agricoles urbains ; et iii) la liste des exploitations agricoles du secteur ne relevant pas des ménages.

FIGURE 3. CADRE D'ÉCHANTILLONNAGE POUR LE PROGRAMME INTÉGRÉ D'ENQUÊTES AGRICOLES ET RURALES



IV.B.1. Développer le cadre approprié aux ménages (urbains et ruraux)

Une liste complète des ménages agricoles et non agricoles dans les zones rurales et les ménages agricoles urbains pourrait être établie à l'aide des données du Recensement de la population et des logements (PHC) d'un pays. Toutefois, l'utilisation des données des PHC présente des difficultés, notamment en ce qui concerne l'identification des ménages agricoles. En particulier, il existe un risque de sur-couverture des ménages agricoles dans les zones urbaines si, par exemple, les ménages dont les membres pratiquent l'agriculture en tant qu'employés rémunérés sont considérés comme des ménages agricoles dans le cadre. Lorsque le PHC est mis en œuvre en tenant compte des Principes et recommandations des Nations Unies pour les recensements de la population et des logements, la FAO (2015b) indique que les ménages sont considérés comme des ménages agricoles si au moins un membre du ménage est considéré avoir pour *activité économique* celle d'un *travailleur pour compte propre*, d'un *travailleur indépendant* ou d'un *employeur*, et si sa principale activité est liée à l'agriculture (par exemple, la culture de plantes, la production animale, la multiplication de plantes ou l'agriculture mixte).

Cependant, il est très inhabituel de disposer de données sur la population et les logements (PHP) à jour pour sélectionner directement un échantillon de ménages aux fins de mise en œuvre de l'enquête. Une pratique courante, où le plan d'échantillonnage à plusieurs étapes est approprié (voir la section V sur le plan d'échantillonnage, ci-dessous), consiste à sélectionner un échantillon d'unités primaires d'échantillonnage (p. ex., zones de dénombrement) à partir des données PHP et à établir une nouvelle liste des ménages (micro-recensement) dans chacune d'elles, afin de sélectionner l'échantillon final des ménages. Pour les zones urbaines, il peut être important d'évaluer la pertinence d'un plan d'échantillonnage à plusieurs étapes avant d'utiliser une telle approche, en particulier pour les pays qui n'ont pas d'expérience préalable de réalisation d'enquêtes agricoles. Pour ce faire, on peut utiliser les données d'un recensement précédent des PHP ou d'un recensement agricole pour estimer l'effet du plan. En présence d'un effet de plan élevé, un simple

échantillonnage aléatoire serait plus rentable. Pour ce faire, il faudrait établir une liste exhaustive des ménages agricoles urbains par l'intermédiaire d'une opération de recensement sous forme de liste ou en se fondant sur les sources administratives disponibles. Des données de grande taille pourraient également être explorées (voir Young et al. 2018).

IV.B.2. Couverture des ménages agricoles urbains

L'inclusion de la population des ménages agricoles urbains est facultative, en fonction de l'importance de l'agriculture urbaine dans le secteur des ménages dans le pays. Si l'incidence des ménages agricoles urbains est particulièrement faible, il serait rentable d'ignorer ces populations dans la collecte de données et d'envisager leur contribution au stade de l'estimation. Par exemple, les agrégats finaux peuvent être ajustés en utilisant les proportions de ménages agricoles urbains dans les domaines d'inférence. Dans le plan d'échantillonnage de l'Initiative 50x2030, ces populations sont traitées comme un domaine facultatif que les pays peuvent décider de prendre en compte selon leur importance et des coûts opérationnels de l'enquête. Bien que la proportion de ménages agricoles dans les zones urbaines par rapport à l'ensemble des ménages agricoles soit plutôt faible dans plusieurs pays, ces ménages peuvent être responsables d'une part importante de la production agricole globale ou de la production de produits spécifiques (par exemple, les cultures horticoles, l'élevage). Dans de tels cas, il sera important de les prendre en compte dans l'échantillon de l'enquête.

IV.B.3. Développer le cadre approprié pour les exploitations du secteur ne relevant pas des ménages

La solution idéale serait d'établir une liste complète des exploitations agricoles du secteur ne relevant pas des ménages en utilisant les données d'un recensement récent de l'agriculture qui couvre toutes les exploitations agricoles du pays. À défaut, pour établir le cadre de ces exploitations, les registres d'établissements agricoles sont souvent utilisés comme point de départ, y compris le registre national des établissements et les registres informels des établissements détenus par les organisations d'agriculteurs. Des efforts doivent être faits pour traiter le recoupement probable entre les registres officiels et informels. En outre, tous les autres registres pertinents devraient être pris en considération, y compris la liste des institutions gouvernementales (centres de recherche agricole, écoles, hôpitaux, prisons, etc.) et des organisations non gouvernementales qui gèrent des exploitations agricoles. Les connaissances et les informations locales des agents de vulgarisation et des autorités locales sur les grandes exploitations de type spécialisé sont utiles pour le développement du cadre du secteur ne relevant pas des ménages. La qualité des données provenant de sources administratives mentionnées ci-dessus devrait être évaluée et des opérations spécifiques de traitement des données (y compris le couplage des enregistrements) devraient être planifiées pour produire une base d'échantillonnage de qualité suffisante. L'outil d'évaluation de la qualité des registres administratifs de la Banque mondiale serait utile pour évaluer la qualité de ces données administratives. La FAO (2018c) fournit des directives pratiques sur l'amélioration et l'utilisation des données administratives dans la conception des enquêtes agricoles. Les cadres d'échantillonnage mis au point à partir de sources administratives pourraient être mis à jour avec des opérations de terrain complémentaires.

IV.C. Cadre du Programme d'enquête agricole

La principale exigence de ce cadre est une couverture adéquate de toutes les exploitations agricoles du pays, c'est-à-dire de tous les ménages agricoles et de toutes les exploitations agricoles du secteur ne relevant pas des ménages. Les directives évoquées à la section II.A.1 pour le développement du cadre des ménages peuvent également être utilisées pour les ménages agricoles. Toutefois, les données d'un recensement

récent de l'agriculture peuvent être utilisées pour élaborer le cadre d'échantillonnage du secteur des ménages agricoles et des exploitations ne relevant pas du secteur des ménages, si elles sont couvertes par le recensement. Comme il est expliqué à la section II.A.1, dans la plupart des cas, les données du recensement seront obsolètes. Par conséquent, il faudrait choisir un échantillon d'unités primaires d'échantillonnage à partir des données du plus récent recensement de la population et des logements (PHP) ou du plus récent recensement de l'agriculture, suivi de nouvelles constitutions de listes dans chacun d'eux.

Les recommandations relatives à l'élaboration d'un cadre directeur d'échantillonnage pour les enquêtes agricoles intégrées sont présentées dans le Manuel sur l'enquête agricole intégrée.¹ Les pays sont renvoyés à ces recommandations pour l'élaboration de bases de sondage pour le programme d'enquêtes agricoles de l'Initiative 50x2030, qui a une structure similaire. Les deux principales recommandations sont résumées ci-après :

- 1) Dresser une liste complète des exploitations agricoles i) dans le secteur des ménages et ii) dans le secteur ne relevant pas des ménages. La liste pour le secteur ne relevant pas des ménages peut être développée comme expliqué ci-dessus dans la section II.A.3. Pour le secteur des ménages, une approche coût-efficacité appropriée consiste à relier les recensements de la population et de l'agriculture, comme le suggère le Recensement de l'agriculture 2020 (FAO, 2015a).
- 2) Utiliser une base de sondage multiple composée d'une base aréolaire et des deux bases de liste suivantes pour i) les exploitations sans terre d'élevage de bétail et ii) les grandes exploitations agricoles commerciales. Le processus opérationnel pour l'établissement d'une base aréolaire pour les enquêtes agricoles est expliqué dans FAO (2015b). Étant donné qu'un cadre régional ne couvre pas les exploitations sans terre d'élevage de bétail, il est recommandé d'établir une liste complémentaire de ces exploitations. En outre, s'il arrive que de grandes exploitations agricoles soient échantillonnées à partir d'une base aréolaire, elles peuvent se comporter comme des valeurs aberrantes. Une deuxième liste de grandes exploitations agricoles commerciales est donc également recommandée (FAO, 2017). Les procédures d'élaboration de ces listes (exploitations sans terre et grandes exploitations) sont assez similaires à celles proposées pour les exploitations du secteur ne relevant pas des ménages (voir la section III.B.3 ci-dessus). Le point de départ serait la consultation des registres agricoles (formels et informels) du pays et d'autres sources administratives pour établir une liste préliminaire qui devrait être mise à jour avec les informations des agents de vulgarisation, des autorités locales et des opérations de terrain *ad hoc* visant à compléter les listes. La question des approches d'estimation ayant des cadres ou bases d'échantillonnage qui se recoupent est abordée à la section VII.

V. Plans d'échantillonnage

La stratégie d'échantillonnage de l'Initiative 50x2030 combine les expériences du Programme d'enquête AGRIS de la FAO et du LSMS-ISA de la BM. Par conséquent, le plan d'échantillonnage du programme d'enquête 50x2030 devrait intégrer les principales caractéristiques des stratégies d'échantillonnage des deux systèmes d'enquête.

¹ Chapitre 5 du Programme FAO (2017)

V.A. Plan d'échantillonnage pour le Programme intégré d'enquêtes agricoles et rurales

Le plan intégré d'échantillonnage devrait assurer des estimations fiables des principales variables d'intérêt pour les populations concernées au niveau des domaines d'estimation (également appelés domaines d'inférence ou strates analytiques). Ces domaines sont généralement des zones administratives pour lesquelles les pays attendent des statistiques fiables des données de l'enquête : régions, provinces, districts, etc. Les domaines sont spécifiques au pays et le plan d'échantillonnage doit être adapté à chacun d'entre eux. Les pays devraient être informés que le choix des domaines a une incidence directe sur la taille globale de l'échantillon et, par conséquent, sur le budget : plus le nombre de domaines d'estimation est élevé, plus le budget global est élevé. Le plan d'échantillonnage du Programme intégré d'enquêtes agricoles et rurales est résumé dans le tableau 1 ci-dessous pour les différents secteurs considérés. Ces recommandations doivent être adaptées aux besoins spécifiques des pays.

TABLEAU 1. RÉSUMÉ DES PRINCIPAUX ÉLÉMENTS DU PLAN D'ÉCHANTILLONNAGE POUR LE PROGRAMME INTÉGRÉ

Éléments	Populations concernées		
	Ménage (rural)	Ménage (urbain)	Secteur ne relevant pas des ménages
Unités d'observation	- Ménages - Exploitations agricoles	Exploitations agricoles	Exploitations agricoles
Unités finales d'échantillonnage	Ménages	Ménages	Exploitations agricoles
Cadres/Bases	Liste des ménages du recensement de la population ou liste des ZD issue du recensement de la population et des micro-recensements dans les ZD échantillonnées	Liste des ménages issue du recensement de la population	Liste des exploitations agricoles ne relevant pas du secteur des ménages établie à partir de registres et/ou d'opérations de terrain
Méthode d'échantillonnage	Stratifiée en deux étapes	Spécifique au pays : Stratifiée en une étape ou deux étapes	Stratifiée en une étape
Stratification	Spécifique au pays : - Strates au niveau des USE : zones administratives ; zones agro-écologiques ; intensité de l'activité agricole à l'aide de données sur l'utilisation des terres ; proportion de ménages agricoles - Strates au niveau des USE (intra-UPE) : pratique agricole	Spécifique au pays : zones administratives ; zones agro-écologiques	Spécifique au pays : zones administratives ; systèmes de production (cultures/bétail/mixte) ; catégorisation <i>ad hoc</i> , p. ex. strates selon une mesure de la taille (p. ex. valeur de la production)
Modèle d'échantillonnage	1 ^{ère} étape : PPT des UPE (ZD) 2 ^e étape : Échantillonnage aléatoire systématique ou simple sans remplacement des ménages	Spécifique au pays : en fonction de la méthode d'échantillonnage adoptée	Échantillonnage aléatoire systématique ou simple sans remplacement dans chaque strate

V.A.1. Secteur des ménages (zones rurales)

Un plan d'échantillonnage stratifié standard en deux étapes est recommandé pour ce secteur. Ce modèle d'échantillonnage est largement utilisé dans les enquêtes sur les ménages et les exploitations agricoles. Le cadre idéal, tel que discuté ci-dessus, serait la liste des ménages ruraux tirée d'un recensement récent de la population et des logements (PHC).

Les unités primaires d'échantillonnage (UPE) sont de préférence des zones de dénombrement (ZD) définies dans le cadre du PHC parce qu'elles sont généralement assez homogènes en termes de taille de la population. L'homogénéité de la population servira à améliorer les estimations finales. Dans certains cas, il peut être utile de fusionner plusieurs petites ZD et de diviser les grandes ZD pour améliorer l'homogénéité. La FAO (2017) a examiné les questions liées au choix de l'UPE.

Les UPE doivent être stratifiées comme indiqué à la section VI.A. Dans chaque strate, un échantillon d'UPE est identifié en utilisant une méthode d'échantillonnage proportionnelle à la probabilité (PPT) sans remplacement.² La mesure de la taille des UPE qui sont également des ZD est généralement égale au nombre de ménages de cette zone de dénombrement résultant de la base de sondage.

Les unités secondaires d'échantillonnage (USE) sont les ménages. À l'intérieur de chaque UPE échantillonnée, l'échantillon d'USE peut être sélectionné par échantillonnage aléatoire simple stratifié (ou échantillonnage systématique) sans remplacement. La stratification au sein de l'UPE est abordée à la section VI.B.

V.A.2. Secteur des ménages (zones urbaines)

Étant donné la proportion relativement faible de ménages agricoles urbains dans la plupart des contextes (comme nous l'avons vu plus haut), un échantillonnage aléatoire simple stratifié serait approprié pour ces populations dans la plupart des pays. Dans la plupart des cas, cette opération serait rentable si le cadre recommandé (la liste des ménages du recensement de la population) était récent, permettant une sélection directe des ménages sans faire une nouvelle liste sur le terrain. Dans tous les cas, il faut s'attendre à une dispersion géographique relativement élevée de l'échantillon, ce qui augmenterait le coût de la collecte des données. Si le nombre total de ménages agricoles urbains est relativement important, un échantillonnage stratifié en deux étapes peut être exploré ne tenant compte de la répartition de ces ménages dans la zone urbaine.

V.A.3. Secteur ne relevant pas des ménages

Un plan stratifié en une seule étape est approprié pour les exploitations du secteur ne relevant pas des ménages (FAO, 2017). Les critères de stratification peuvent être le type de système de production agricole (culture/élevage/mixte) ou toute autre typologie ad hoc. La taille des exploitations, si elle est disponible dans le cadre, pourrait être considérée pour la stratification ou même pour une sélection PPT. Parfois, même une strate « universelle » pourrait être envisagée pour s'assurer que des exploitations très grandes ou spéciales sont incluses dans l'enquête.

V.B. Plan d'échantillonnage pour le Programme d'enquête agricole

La principale différence entre le Programme d'enquête agricole et le Programme intégré d'enquêtes agricoles et rurales est l'exclusion des ménages non agricoles dans les zones rurales. Le Programme d'enquête agricole couvre les mêmes populations concernées que le Programme d'enquêtes agricoles intégrées (AGRIS) de la

² La probabilité proportionnelle à la taille est une procédure d'échantillonnage dans laquelle la probabilité de sélection de chaque unité de l'univers est proportionnelle à sa taille, en l'occurrence la population ou le nombre de ménages.

FAO. Deux types de cadres directeurs d'échantillonnage proposés par la FAO (2017) pourraient être utilisés dans ce cadre : une base de sondage multiple fondée sur des listes ou une base de sondage multiple comprenant une base aréolaire et des listes. Les caractéristiques essentielles des modèles d'échantillonnage pour chaque type de cadre, examinées en détail dans FAO (2017), sont présentées dans le tableau 2 ci-dessous. Une personnalisation plus poussée pourrait être nécessaire en fonction des situations propres à chaque pays.

TABLEAU 2. RÉSUMÉ DES PRINCIPAUX ÉLÉMENTS DU PLAN D'ÉCHANTILLONNAGE POUR LE PROGRAMME D'ENQUÊTE AGRICOLE

Type de cadre	Base de sondage multiple avec listes		Base de sondage multiple avec base aréolaire et base de type liste	
Sous-cadre	Exploitations du secteur des ménages	Exploitations ne relevant pas du secteur des ménages	Base aréolaire	Listes
Unités d'observation	Exploitations agricoles	Exploitations agricoles	Exploitations agricoles	Exploitations agricoles
Unités d'échantillonnage	Ménages agricoles	Exploitations agricoles	Segments ou points	-Les exploitations agricoles sans terre d'élevage de bétail ; - Les grandes exploitations commerciales
Méthode d'échantillonnage	Stratifiée en deux étapes	Stratifiée en une étape	Spécifique au pays : i) Échantillonnage en clusters stratifiés, ou ii) échantillonnage en deux étapes	Stratifiée en une étape
Stratification	Spécifique au pays : zones administratives ; zones agro-écologiques ; intensité de l'activité agricole à l'aide de données sur l'utilisation des terres	Spécifique au pays ; zones administratives ; systèmes de production (culture/élevage/mélange) ; catégorisation ad hoc, par exemple avec des strates selon une mesure de la taille (par exemple, la valeur de la production)	Couverture terrestre/utilisation	Catégorisation ad hoc (par exemple, en fonction de la taille, des activités)
Modèle d'échantillonnage	1 ^{ère} étape : PPT des UPE (EZD) ; 2 ^e étape : échantillonnage aléatoire simple sans remplacement des ménages agricoles	Échantillonnage aléatoire simple sans remplacement des exploitations agricoles	Spécifique au pays : i) Sélection selon la méthode PPT des segments et couverture complète de toutes les exploitations des segments sélectionnés, ou ii) échantillonnage en deux étapes : 1 ^{ère} étape : PPT des UPE (segments ou grilles/clusters de points) ; 2 ^e étape : Échantillonnage aléatoire systématique ou simple sans remplacement de segments ou de points	Échantillonnage aléatoire systématique ou simple sans remplacement des exploitations agricoles

VI. Taille de l'échantillon

Le calcul de la taille de l'échantillon est une étape critique du plan de l'enquête. Elle garantit la production de statistiques fiables en limitant au maximum l'erreur d'échantillonnage. L'approche recommandée est une

approche qui tient compte des exigences analytiques de l'enquête, c'est-à-dire qu'elle permet d'estimer de façon fiable les principales variables d'intérêt. La variable d'intérêt peut être choisie parmi les variables clés nécessaires au calcul des indicateurs les plus importants attendus de l'opération d'enquête.

Une mesure de la dispersion statistique (coefficient de variation, variance, écart-type, etc.) de la variable d'intérêt dans la population doit être incluse dans la formule de la taille de l'échantillon. Idéalement, elle est calculée à l'aide de données de recensement, mais cela n'est pas toujours possible compte tenu du périmètre limité des questionnaires de recensement. Une autre solution courante consiste à l'estimer à partir des données recueillies lors d'enquêtes antérieures. La mesure de la dispersion statistique de toute variable fortement corrélée à la variable d'intérêt peut également être utilisée. Le tableau 3 ci-dessous propose des suggestions de variables d'intérêt clés, compte tenu des objectifs de mesure de l'Initiative (voir la section II) et des cadres d'échantillonnage recommandés.

TABLEAU 3. VARIABLES D'INTÉRÊT SUGGÉRÉES POUR LA DÉTERMINATION DE LA TAILLE DE L'ÉCHANTILLON

Unités d'échantillonnage	Variables d'intérêt suggérées pour la détermination de la taille de l'échantillon
Ménages (agricoles et non agricoles)	Revenu Zone agricole Valeur de la production agricole (cultures/élevage)
Exploitation agricole du secteur ne relevant pas des ménages	Zone agricole Valeur de la production agricole (cultures/élevage)
Exploitation sans terre de bétail	Valeur du cheptel Unité de taille du cheptel
Grandes exploitations agricoles	Zone agricole
Segments/points (base aréolaire)	Zone agricole

Dans le cas où il existe plusieurs variables d'intérêt, il peut être nécessaire de déterminer une valeur maximum aux tailles d'échantillon minimales requises pour chacune d'elles. C'est ce qui est proposé à la section VI.A.1 ci-dessous pour la taille de base des ménages utilisée dans le cadre du Programme intégré d'enquêtes agricoles et rurales.

Si la variable d'intérêt clé est très hétérogène dans la population, la taille de l'échantillon nécessaire sera très grande et peut être trop coûteuse. Dans une telle situation, la solution serait d'utiliser la taille maximale de l'échantillon que le budget peut offrir. Bien sûr, cela ne garantirait pas une estimation fiable, bien que les méthodes statistiques avancées pour une stratification et une répartition optimales puissent être utilisées de façon intéressante, comme par exemple les procédures optimales de stratification et de répartition multivariées proposées par Barcaroli et autres (2020) dans le cadre du programme R, « Échantillonner les Strates ». Une autre alternative pourrait être le programme R « stratbr », décrit dans Brito et autres (2019).

VI.A. Taille de l'échantillon pour le secteur des ménages

Pour ce secteur, il est recommandé d'utiliser un plan d'échantillonnage stratifié en deux étapes. Les unités primaires d'échantillonnage (UPE) recommandées sont généralement les zones de dénombrement et les unités secondaires d'échantillonnage (USE) sont soit des ménages (Programme intégré d'enquêtes agricoles et rurales), soit des ménages agricoles (Programme d'enquête agricole). Les méthodes de calcul de la taille des échantillons d'UPE et d'USE sont traitées dans la présente section.

VI.A.1. Nombre d'unités d'échantillonnage secondaires

Programme intégré d'enquêtes agricoles et rurales

Dans les zones rurales, le Programme intégré d'enquête agricole et rurale vise deux objectifs principaux : produire des estimations pour l'ensemble de la population des ménages ruraux et des estimations pour le sous-ensemble des ménages agricoles au niveau national et infranational. Pour atteindre ces objectifs, la stratégie d'échantillonnage optimale exigerait une liste complète des ménages ruraux d'un PHC récent, classés selon qu'ils sont agricoles (désignés A ci-après) ou non agricoles (désignés B).

Dans l'enquête intégrée, la taille de l'échantillon du secteur des ménages devrait permettre d'estimer de manière fiable une variable clé liée aux ménages (par exemple, le revenu) dans la population des ménages ruraux (A et B) et une estimation fiable d'une variable agricole clé (par exemple, la superficie agricole) dans la sous-population des ménages agricoles (A), les ménages de la sous-population B n'exploitant pas de terres agricoles. Pour calculer la taille minimale de l'échantillon de ménages nécessaire pour atteindre cet objectif, on peut utiliser la formule habituelle approximative fondée sur le coefficient de variation.

Considérons pour chaque domaine d'estimation U_d :

- M_{Ad} et M_{Bd} est le nombre total de ménages respectivement de type A et B.
- cv_{Aincd}^2 et cv_{Bincd}^2 est le coefficient de variation du revenu des ménages de type A et B, respectivement
- cv_{Aland}^2 est le coefficient de variation de la superficie agricole du ménage agricole
- cv_d^{*2} est l'erreur relative maximale acceptable (écart-type) pour l'estimation du revenu total (moyen) et de la superficie agricole
- \overline{def}_{Aincd} , \overline{def}_{Bincd} et \overline{def}_{land} sont des estimations de l'effet de plan pour le revenu des ménages de type A et B et de la zone agricole, respectivement
- g est le taux de réponse attendu

La taille minimale de l'échantillon des ménages (m_d) dans le domaine U_d est de

$$m_d = \frac{1}{g} \left[\text{Max} \left(\overline{def}_{land} \frac{cv_{Aland}^2}{cv_d^{*2} + \frac{cv_{Aland}^2}{M_{Ad}}}, \overline{def}_{Aincd} \frac{cv_{Aincd}^2}{cv_d^{*2} + \frac{cv_{Aincd}^2}{M_{Ad}}} \right) + \overline{def}_{Bincd} \frac{cv_{Bincd}^2}{cv_d^{*2} + \frac{cv_{Bincd}^2}{M_{Bd}}} \right]$$

Ou :

$$m_d = \max(m_{dA,inc}, m_{dA,land}) + m_{dB,inc} = m_{dA} + m_{dB,inc}$$

Cette procédure exige que toutes les variables de la formule des ménages des types A et B (ménages ruraux agricoles et non agricoles) soient incluses dans chaque domaine d . Toutefois, il peut arriver que le coefficient de variation du revenu ne puisse être estimé pour chaque sous-population si l'exercice est entrepris à l'aide de données provenant d'une enquête sur les ménages qui ne couvrirait pas les activités agricoles. Dans ce cas, si $m_{d,inc}$ est la taille minimale globale des ménages ruraux pour une estimation fiable du revenu, nous avons :

$$m_{d,inc} = \frac{1}{g} \overline{def}_{incd} \frac{cv_{incd}^2}{cv_d^{*2} + \frac{cv_{incd}^2}{M_{Ad} + M_{Bd}}}$$

Et:

$$m_d = \max(\tilde{W}_{Ad}m_{d,inc}, m_{d,land}) + (1-\tilde{W}_{Ad})m_{d,inc}$$

Où :

- cv_{incd}^2 est le coefficient de variation du revenu des ménages ruraux dans le domaine d
- \widehat{def}_{incd} est une estimation de l'effet de conception sur le revenu des ménages ruraux
- \tilde{W}_{Ad} est une estimation de la proportion des ménages agricoles dans le domaine d .

Programme d'enquête agricole

Dans le cadre de ce programme, les ménages agricoles (A) étant les seuls concernés, la taille de l'échantillon sera simplement :

$$m_d = \frac{1}{g} \widehat{def}_{land} \frac{cv_{Aland}^2}{cv_d^{*2} + \frac{cv_{Aland}^2}{M_{Ad}}}$$

VI.A.2. Nombre d'UPE

Lorsque les UPE sont sélectionnées selon la méthode d'échantillonnage PPT, la sélection d'un nombre fixe de ménages m_0 par UPE permettra d'obtenir des poids constants. Cela signifie que le nombre d'UPE à sélectionner dans d serait donné en divisant la taille de l'échantillon des ménages par m_0 . Avec cette approche, le nombre d'UPE à sélectionner dans le domaine d est donné par :

$$n_d = \left\lceil \frac{m_d}{m_0} \right\rceil + 1$$

où $\left\lceil \frac{m_d}{m_0} \right\rceil$ est la partie intégrale.

La valeur de m_0 peut être déterminée en tenant compte à la fois des coûts et de l'homogénéité des ménages dans les UPE (corrélation intraclasse $\bar{\rho}$) :

$$m_0 = \sqrt{\frac{c_p \times (1 - \bar{\rho})}{c \times \bar{\rho}}}$$

où c_p et c sont respectivement le coût de l'ajout d'une UPE supplémentaire dans l'échantillon et le coût unitaire d'une interview. La corrélation intraclasse $\bar{\rho}$ peut être estimée à partir d'enquêtes précédentes. Puisque deux variables sont prises en compte, le revenu et la superficie des terres agricoles, la valeur minimale, $\bar{\rho} = \min(\bar{\rho}_{cons}, \bar{\rho}_{land})$, doit être prise en compte (c'est un choix conservateur). Il est à noter que cette formule est une approximation basée sur un échantillonnage aléatoire simple en deux étapes, à la fois des UPE et des USE, où la taille des UPE ne varie pas beaucoup.

Une autre pratique courante pour faciliter l'organisation du travail sur le terrain consiste à fixer le nombre de ménages m_0 à sélectionner dans chaque UPE, en tenant compte de la charge de travail maximale de l'enquêteur pendant le déroulement de l'enquête. Une valeur arbitraire variant généralement entre 10 et 15 est souvent prise en compte.

VI.B. Taille de l'échantillon pour le secteur ne relevant pas des ménages et les exploitations spéciales

Le plan d'échantillonnage recommandé pour les exploitations agricoles du secteur ne relevant pas des ménages est un échantillonnage aléatoire simple stratifié. Le même plan est recommandé pour les listes complémentaires d'exploitations spéciales recommandées dans certains cas (voir la section IV) afin d'améliorer la couverture des cadres d'échantillonnage pour inclure, par exemple, les exploitations agricoles sans terre élevant du bétail ou les grandes exploitations commerciales. Pour ces populations, la valeur de la production agricole serait une variable appropriée à prendre en considération pour le calcul de la taille de l'échantillon. Toutefois, pour les unités de superficie agricole et de taille du cheptel, il peut également être approprié de calculer la taille de l'échantillon pour les grandes exploitations et les exploitations sans terre élevant du bétail, respectivement.

Considérons pour chaque domaine d'estimation U_d :

- M_d est le nombre total d'exploitations
- cv_{yd} est le coefficient de variation de la variable clé d'intérêt
- cv_d^* est l'erreur maximale acceptable (écart-type)
- g est le taux de réponse attendu

La taille de l'échantillon m_d peut être calculée à l'aide de la formule suivante :

$$m_d = \frac{1}{g} \frac{cv_{yd}^2}{cv_d^{*2} + \frac{cv_{yd}^2}{M_d}}$$

VII. Stratification

La stratification consiste à diviser la population en sous-populations (strates) et à effectuer une sélection d'échantillon indépendante dans chacune d'elles. La stratification peut contribuer dans une large mesure à réduire les erreurs d'échantillonnage, en particulier lorsque les sous-populations sont homogènes et que la taille de l'échantillon est bien répartie dans les strates. Parmi les autres avantages importants de la stratification, mentionnons le contrôle de la taille des échantillons pour différentes sous-populations aux fins de déclaration et d'analyse, et la possibilité d'adopter des stratégies d'échantillonnage différentes pour chaque strate. Dans de nombreux pays, les ménages et les enquêtes agricoles devraient produire des estimations fiables au niveau infranational (régions, provinces, districts, etc.) qui sont les domaines d'estimation ou les strates analytiques. Pour améliorer l'estimation dans les domaines d'estimation, une stratification supplémentaire, discutée ici, est habituellement effectuée à l'intérieur de ces domaines. La discussion porte sur la stratification dans le cadre d'un plan en deux étapes avec une base de sondage de type liste, car les procédures de stratification sont assez simples avec les bases aréolaires et dans les contextes où un plan d'échantillonnage stratifié à une seule étape est recommandé.

VII.A. Stratification et répartition des UPE

Lors de la mise en œuvre de l'échantillonnage en deux étapes, la FAO (2017) recommande une stratification des ZD par zones administratives (par exemple, régions, provinces, etc.) et zones agro-écologiques. Cela devrait se faire avant la première phase de sélection, afin d'améliorer les estimations des statistiques agricoles. La stratification des UPE devrait être contrôlée avec attention, car il n'est pas souhaitable d'avoir trop de strates (un échantillon indépendant doit être sélectionné dans chaque strate). Pour éviter un trop grand nombre de strates, la stratification explicite peut être associée à une stratification implicite. Il s'agit de trier le cadre d'échantillonnage selon des critères pertinents (généralement géographiques) dans chaque

strate et de sélectionner un échantillon indépendant dans chaque strate selon une méthode d'échantillonnage systématique.

Stratification dans le cadre du Programme intégré d'enquêtes agricoles et rurales

Lorsque la liste des ménages du PHP est obsolète, la structure réelle des ménages au sein des UPE échantillonnées ne peut être connue qu'après établissement d'une nouvelle liste des ménages dans ces UPE. Un inconvénient majeur est le manque de contrôle sur l'échantillon final, en particulier le nombre de ménages agricoles requis dans le domaine. Étant donné que la sélection se fait au niveau des UPE, elle peut présenter une situation différente en termes de proportion des ménages agricoles.

Pour maintenir le contrôle de la taille de l'échantillon final par type de ménage (A et B), il est préférable de procéder à une stratification de premier niveau des ZD en fonction de la proportion de ménages agricoles dans chacun d'eux, estimée à partir du dernier PHC ou d'une autre source appropriée. Même si les données sur les PHP sont considérées comme désuètes, cette information structurelle (proportion de ménages agricoles) ne variera probablement pas beaucoup dans toutes les UPE et pourra être utile à des fins de stratification. La stratification de premier niveau ci-dessous peut être envisagée en utilisant un seuil de proportion ρ ($\frac{1}{2} < \rho < 1$).

Strates composées du premier niveau d'UPE	Définition
Agricole	$Proportion\ de\ ménages\ agricoles\ dans\ l'UPE \geq \rho$
Mixte	$1 - \rho < Proportion\ de\ ménages\ agricoles\ dans\ l'UPE < \rho$
Non agricole	$Proportion\ de\ ménages\ agricoles\ dans\ l'UPE \leq 1 - \rho$

L'échantillon d'UPE dans le domaine d (n_d) peuvent être répartis en utilisant des paramètres θ_a, θ_m et θ_{na} avec $\theta_a + \theta_m + \theta_{na} = 1$

Répartition de premier niveau	
Strates composées du premier niveau d'UPE	Répartition de l'échantillon d'UPE
Agricole	$\theta_a n_d$
Mixte	$\theta_m n_d$
Non agricole	$\theta_{na} n_d$

Si m_0 ménages sont sélectionnés dans chaque UPE échantillonnée en utilisant un échantillonnage aléatoire systématique ou simple sans remplacement, le nombre attendu de ménages agricoles dans l'échantillon final (m_{dAexp}) est :

$$m_{dAexp} = \rho m_0 \theta_a n_d + (1 - \rho) m_0 \theta_m n_d + \delta m_0 \theta_{na} n_d = (\rho \theta_a + (1 - \rho) \theta_m) m_0 n_d + \delta \theta_{na} m_0 n_d$$

$\delta < 1$ est inconnu avant la sélection de l'échantillon de ménages, contrairement aux autres paramètres qui sont fixés par le concepteur de l'échantillon.

Considérons τ la proportion de ménages agricoles dans l'échantillon prévu :

$$\tau = \frac{m_{dA}}{m_d} = \frac{m_{dA}}{m_0 n_d} \Rightarrow m_{dA} = \tau m_0 n_d$$

Pour garantir la réalisation de l'échantillon prévu de ménages agricoles dans l'échantillon final de ménages, les paramètres θ_a , θ_m et θ_{na} pourraient être fixés de manière à obtenir $m_{dAexp} \geq m_{dA}$. Cela correspond à la formule :

$$(\rho\theta_a + (1 - \rho)\theta_m)m_0n_d + \delta\theta_{na}m_0n_d \geq \tau m_0n_d$$

δ étant inconnus, les paramètres θ_a , θ_m et θ_{na} peuvent donc être fixé dans les conditions suivantes :

$$\rho\theta_a + (1 - \rho)\theta_m \geq \tau$$

$$\theta_{na} = 1 - (\theta_a + \theta_m)$$

Une stratification de deuxième niveau des UPE peut être effectuée à l'intérieur des strates de premier niveau. Les critères communs de stratification pour améliorer les estimations dans les enquêtes sur les exploitations agricoles et les ménages sont les suivants : les zones agro-écologiques ; les classes d'utilisation des terres ; les catégories de taille en fonction de la population ; la superficie agricole ; l'intensité de l'activité agricole, etc.). Pour les pays qui s'intéressent fortement aux statistiques ventilées par sexe, on peut envisager de stratifier les UPE en fonction du nombre de ménages dirigés par une femme. La répartition dans ces strates de deuxième niveau peut suivre différents critères. En règle générale, dans les enquêtes sur les ménages, une répartition proportionnelle à la population des strates est prise en compte. La FAO (2017) recommande la répartition optimale de Neyman pour les enquêtes agricoles. Kish (1987, p. 228) propose une solution de compromis entre la répartition égale et la répartition proportionnelle :

$$n_{dh} = n_d \times \frac{\theta_{dh}}{\sum_{h=1}^{H_d} \theta_{dh}}$$

Où :

$$\theta_{dh} = \sqrt{\left(W_{dh}^2 + \frac{1}{H_d^2}\right)}$$

H_d est le nombre de strates dans le domaine d, tandis que W_{dh} est la taille relative de la strate h dans le domaine d, pouvant correspondre à la proportion d'UPE dans la strate h par rapport au total du domaine, $W_{dh} = N_{dh}/N_d$ (taille relative en termes de population). Une stratification et une répartition multivariée (Barcaroli, 2020) ou une répartition de puissance de compromis (Bankier, 1988) pourrait également être explorée si la base de sondage contient des variables pertinentes corrélées avec le revenu des ménages ou la superficie agricole (taille du ménage, bétail, production agricole, etc.) au niveau de l'UPE.

Stratification dans le Programme d'enquête agricole

Dans le cadre du Programme d'enquête agricole, la même stratégie de stratification en UPE que celle proposée ci-dessus pourrait être utilisée, mais la répartition de premier niveau n'est pas nécessaire puisque ce programme d'enquête ne couvre pas les ménages non agricoles.

VII.B. Stratification et répartition intra-UPE

Bien que la stratification intra-UPE ait un effet limité sur la variance totale, elle peut être utile pour assurer une bonne couverture de populations d'intérêt spécifiques. Pour le Programme d'enquêtes agricoles et rurales intégrées, un critère de stratification pertinent pourrait être de savoir si l'agriculture est pratiquée ou non. Les pays qui s'intéressent à l'obtention de statistiques ventilées par sexe peuvent également tenir compte du sexe du chef du ménage. Dans le cadre du Programme d'enquête agricole, les systèmes de

production (cultures/élevage/mixte) devraient être explorés si ces informations sont disponibles dans le cadre. Une répartition proportionnelle de l'échantillon devrait être utilisée dans l'UPE pour s'assurer que les unités finales ont une probabilité égale de sélection dans l'UPE. Une autre possibilité serait d'envisager une stratification implicite, c'est-à-dire de classer les ménages par type et de sélectionner ensuite l'échantillon aléatoire de m_0 au moyen d'un critère systématique.

VIII. Estimation

Les estimateurs, les variances et les questions particulières liées à l'estimation sont abordés dans la présente section.

VIII.A. Estimateurs et variances

Les principaux estimateurs et les principales variances des modèles d'échantillonnage proposés sont présentés ici.

VIII.A.1. Plan d'échantillonnage stratifié en deux étapes

Notation

h = strate

H = nombre total de strates

i = UPE

N = nombre total d'UPE

I_h = nombre total d'UPE dans la h^e strate

j = USE

M_{hi} = nombre total d'USE trouvées dans la i -ème UPE dans la strate h ($j = 1, 2, \dots, M_{hi}$)

$M = \sum_h \sum_i M_{hi}$ = nombre total d'USE dans le pays

F_{hi} = nombre total d'USE répertoriées dans la base de sondage comme appartenant à la i -ème USE de la strate h

$F_h = \sum_i F_{hi}$, est le nombre total d'USE répertoriées dans la base de sondage de la strate h

n_h = nombre d'UPE de l'échantillon sélectionnées dans la strate h

m_{hi} = nombre d'USE de l'échantillon sélectionnées dans la i -ème USE de la strate h

y_{hij} = valeur de la variable cible Y observée sur la j -ème USE, dans la i -ème USE de la strate h

Estimateurs

La probabilité de sélection de l'USE j dans l'échantillon est le produit de la probabilité de sélection de l'UPE i dans laquelle elle se trouve ($n_h \frac{F_{hi}}{F_h}$) et de sa probabilité de sélection dans l'UPE i ($\frac{m_{hi}}{M_{hi}}$).

Ainsi, le *poids* attribué à l'USE f sélectionnée dans la i -ème UPE de la strate h est :

$$w_{hij} = \left(n_h \frac{F_{hi}}{F_h} \right) * \left(\frac{M_{hi}}{m_{hi}} \right)$$

Les poids sont à peu près constants dans chaque strate lorsque m_{hi} est constant ($m_{hi} = m_{h0}$) et lorsque le nombre d'USE trouvées dans une UPE échantillonnée est approximativement égal au nombre d'USE résultant de la base de sondage ($M_{hi} = F_{hi}$).

Une estimation du montant total de Y pour l'ensemble de la population peut être calculée à l'aide de la formule suivante :

$$\hat{Y} = \sum_h \sum_i \sum_j w_{hij} y_{hij}$$

La moyenne de Y peut être estimée avec deux estimateurs différents :

- *Moyenne simple*

$$\hat{\bar{Y}} = \hat{Y} / M$$

- *Moyenne pondérée de l'échantillon*

$$\tilde{\bar{Y}} = \frac{\hat{Y}}{\sum_h \sum_i \sum_j w_{hij}}$$

Ce dernier estimateur tend à être préférable à la moyenne simple lorsque la taille totale de la population est inconnue ou incertaine (comme c'est le cas, par exemple, si le cadre est désuet) et lorsque l'estimation de la moyenne de Y est faite pour un domaine d'intérêt imprévu (p. ex., la sous-population).

Variance

La variance de l'estimateur de la quantité totale de Y dans la population est une formule assez complexe et on peut la trouver, par exemple, dans Cochran (1977, équation 11.42). On peut obtenir une estimation simple et approximative d'un tel écart en utilisant uniquement les UPE grâce à l'estimateur suivant, fourni par Särndal, Swensson et Wretman (1992, p. 154), qui surestime cette variance

$$\tilde{v}(\hat{Y}) = \sum_{h=1}^H (M_h^2 \frac{1}{m_h(m_h - 1)} \sum_{i=1}^{I_h} (\hat{Y}_{hi} - \frac{1}{m_h} \sum_{i=1}^{I_h} \hat{Y}_{hi})^2)$$

où \hat{Y}_{hi} et \hat{Y}_h sont les estimations du montant total de Y aux niveaux des UPE et de strate, respectivement.

Une estimation approximative de la variance de la moyenne est :

$$\tilde{v}(\hat{\bar{Y}}) = \frac{1}{M^2} \tilde{v}(\hat{Y})$$

Coefficient de variation du total :

$$\bar{cV}(\hat{Y}) = \frac{\sqrt{\tilde{v}(\hat{Y})}}{\hat{Y}}$$

Coefficient de variation de la moyenne :

$$\bar{cV}(\hat{\bar{Y}}) = \frac{\sqrt{\tilde{v}(\hat{\bar{Y}})}}{\hat{\bar{Y}}}$$

VIII.A.2. Plan d'échantillonnage stratifié à une étape

Considérons la notation suivante :

h = strate

H = nombre total de strates

i = exploitation agricole

N_h = nombre total d'exploitations agricoles dans la strate h

N = nombre total d'exploitations agricoles dans le cadre

n_h = nombre d'exploitations agricoles échantillons sélectionnées dans la strate h

y_{hi} = valeur de la variable cible Y observée dans la i -ème exploitation agricole, dans la strate h

Le poids de l' i -ème exploitation agricole dans la strate h est simplement l'inverse de sa probabilité de sélection :

$$w_{hi} = \frac{N_h}{n_h}$$

Une estimation du montant total de Y pour la population des exploitations agricoles est :

$$\hat{Y} = \sum_h \sum_i w_{hi} y_{hi}$$

Une estimation de la variance de l'échantillonnage est fournie par :

$$\tilde{v}(\hat{Y}) = \sum_{h=1}^H N_h(N_h - n_h) \frac{1}{n_h(n_h - 1)} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2$$

Où :

$$\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$$

Une estimation de la moyenne de la population de Y est fournie par :

$$\hat{Y} = \hat{Y}/N$$

Et la variance de l'échantillonnage correspondante est estimée de la façon suivante :

$$\tilde{v}(\hat{Y}) = \frac{1}{N^2} \tilde{v}(Y)$$

Le coefficient de variation du total et de la moyenne peut être estimé à l'aide des mêmes formules que celles prévues pour le plan d'échantillonnage stratifié en deux étapes.

VIII.B. Différence entre les unités d'échantillonnage et d'observation dans le secteur des ménages

L'unité d'échantillonnage proposée pour les programmes d'enquête dans le secteur des ménages est le ménage. Toutefois, l'unité d'observation de la composante agricole est l'exploitation agricole. Il est donc important de traiter la relation entre les ménages et les exploitations agricoles à prendre en considération dans le calcul des poids d'échantillonnage des exploitations. Quatre types de liens sont prévus entre les unités d'observation et d'échantillonnage (Lessler et Kalsbeek, 1992). Dans le cadre des enquêtes intégrées, elles peuvent être décrites comme suit :

- Un-à-un : chaque ménage est associé à une exploitation agricole unique et chaque exploitation agricole est associée à un ménage unique.
- Un-à-plusieurs : un ménage peut être associé à de nombreuses exploitations agricoles, mais chaque exploitation agricole est associée à un seul ménage.
- Plusieurs-à-un : de nombreux ménages peuvent être associés à une seule exploitation agricole, mais chaque ménage est associé à une seule exploitation agricole.
- Plusieurs-à-plusieurs : chaque exploitation agricole peut être associée à plusieurs ménages et chaque ménage peut être associé à plusieurs exploitations agricoles.

Le nombre d'unités d'échantillonnage qui conduiraient à la collecte de données auprès de la même unité déclarante est appelé « multiplicité » (FAO, 2015b). Ici, la multiplicité se produit lorsqu'une exploitation agricole est exploitée par deux ménages distincts ou plus, ou lorsqu'un seul ménage exploite deux exploitations distinctes ou plus. Dans le secteur des ménages, ce dernier cas est très rare, et il est complexe sur le plan opérationnel de le saisir dans l'enquête en tenant compte de la définition de l'exploitation. Reconnaissant la rareté de cette situation et la complexité de traiter les exploitations présentes dans le ménage comme deux entités distinctes, les instruments d'enquête 50x2030 recueillent les informations comme si les exploitations étaient des entités uniques.

L'existence de multiplicités conduit à des estimations biaisées (Lavalée, 2007). Falorsi et autres (2015) recommandent l'utilisation de la méthode généralisée de partage du poids (MGPP), proposée par (Lavalée, 2007), pour traiter les multiplicités entre les exploitations et les ménages.

Les recommandations opérationnelles suivantes peuvent être formulées pour l'utilisation de la MGPP dans le cadre du plan d'échantillonnage intégré 50x2030 :

- (i). Identification des multiplicités lors du recensement en liste des ménages : lors de l'inscription des ménages dans l'UPE, inclure des questions pour identifier les exploitations à ménages multiples.
- (ii). Après l'échantillonnage et pendant l'enquête proprement dite, identifier les ménages échantillonnés liés à chaque exploitation à ménages multiples.
- (iii). Au cours du traitement des données, calculer les poids d'échantillonnage des exploitations à ménages multiples à l'aide de la méthode généralisée du partage des poids.

Considérons :

w_i le poids d'échantillonnage du ménage i ($w_i = 0$ si le ménage n'est pas échantillonné)

$l_{ij} = 1$ si le ménage i exploite l'exploitation agricole j et $l_{ij} = 0$ sinon

Le poids d'échantillonnage w_j' de l'exploitation agricole j peut être exprimé comme suit :

$$w_j' = \frac{\sum_i l_{ij} w_i}{\sum_i l_{ij}}$$

VIII.C. Pondération à l'aide de différentes unités d'observation

Pour atteindre les principaux objectifs de l'Initiative, deux unités d'observation sont considérées : les ménages et les exploitations agricoles. Toutefois, comme il a été indiqué ci-dessus, il existe des preuves de l'existence d'exploitations à ménages multiples et elles doivent être identifiées dans le cadre d'échantillonnage. Dans les pays où il existe de nombreux cas de ménages exploitant des exploitations agricoles en partenariat, compte tenu de l'obligation de collecter des données tant au niveau des ménages qu'au niveau des exploitations, les actions suivantes doivent être entreprises :

- Déterminer les informations agricoles devant être obtenues au niveau des ménages et des exploitations, y compris les revenus et les dépenses, les actifs, les investissements, etc.
- Lors de l'interview des ménages exploitant une exploitation à ménages multiples, collecter ces informations séparément au niveau du ménage et de l'exploitation. À l'évidence, cela augmentera la pénibilité de l'interview pour les personnes interrogées, mais nous pouvons espérer, comme mentionné ci-dessous, que ces cas seront inhabituels.
- Enfin :
 - l'estimation pondérée des données sur les ménages au niveau national sera calculée en utilisant les coefficients de pondération directs de l'échantillonnage des ménages comme coefficients de pondération. Les poids initiaux sont ensuite corrigés par des ajustements d'étalonnage, de post-stratification, d'absence de réponse, etc., pour calculer les poids finaux.
 - l'estimation pondérée des données sur les exploitations au niveau national sera effectuée en utilisant les poids d'échantillonnage des exploitations (w_j') calculés comme expliqué dans la section précédente et en utilisant les poids finaux des ménages.

Par conséquent, le poids final des ménages et le poids final des exploitations ne sera différent que dans le cas d'exploitations tenues par plusieurs ménages.

VIII.D. Estimation pour les cultures rares

La couverture des éléments rares est une des difficultés de chaque enquête car elle peut nécessiter un échantillon très vaste. Avec la méthode d'enquête 50x2030, il devrait être possible d'effectuer des estimations fiables pour la plupart des cultures au niveau national, mais pas pour les domaines infranationaux. Pour les cultures très rares, une telle garantie n'existe pas, même au niveau national. Toutefois, à la demande des pays, des méthodes d'estimation avancées (par exemple, des modèles d'estimation de petites superficies) pourraient être explorées en utilisant les informations auxiliaires disponibles provenant de données administratives ou de recensements agricoles antérieurs. Ces informations auxiliaires pourraient également être utilisées au stade du plan dans le cadre d'un travail de stratification *ad hoc* afin d'améliorer la couverture des cultures rares et obtenir des estimations fiables.

IX. Sous-échantillonnage et estimation

Dans les enquêtes sur l'agriculture et les ménages, le sous-échantillonnage peut être utilisé comme un outil rentable à diverses fins, notamment :

L'utilisation de domaines d'estimation différents pour des informations spécifiques

Si les principaux domaines de l'enquête sont usuellement les zones administratives infranationales (régions, provinces, districts, etc.), le pays pourrait considérer que certaines informations ne sont nécessaires que pour des estimations de niveau national. Par conséquent, il ne sera pas nécessaire de les recueillir dans l'échantillon complet, au sein de chaque domaine d'estimation. La taille de l'échantillon, dans le cas d'une estimation nationale, peut être calculée pour les informations concernées et les questions correspondantes seront administrées uniquement par le biais d'un sous-échantillon dans chaque domaine d'estimation. Par exemple, si la plupart des informations recueillies par un module rotatif ne sont pas demandées pour l'élaboration des politiques infranationales, les questionnaires rotatifs pourraient être administrés via un sous-échantillon.

La collecte d'informations avec des coûts opérationnels et/ou des délais plus élevés

Certaines méthodes de collecte de données fournissent des informations de grande qualité mais leur mise en œuvre peut être trop coûteuse pour être faite à grande échelle, comme le métrage objectif des terrains (par exemple, en utilisant le GPS) ou le calcul objectif du rendement (« coupes-témoins »). De telles opérations peuvent être effectuées sur un sous-échantillon lorsque la mise en œuvre à grande échelle n'est pas possible, et les résultats peuvent ensuite être utilisés pour corriger l'erreur de mesure (écart-type) sur l'ensemble de l'échantillon. La réussite de la construction des facteurs de correction dépend de la nature de l'erreur d'estimation et peut ne pas être possible dans tous les cas.

IX.A. Taille du sous-échantillon

La taille du sous-échantillon peut être calculée en se fondant sur la variabilité des variables clés destinées à être collectées par le biais du sous-échantillon. Une option consiste à stratifier l'échantillon principal avant de sélectionner le sous-échantillon. Dans ce cas, Fuller (2009) propose une formule pour calculer la taille optimale du sous-échantillon. Supposons par exemple que le sous-échantillonnage soit effectué pour une mesure objective du rendement et que y soit le rendement collecté par voie de déclaration. Si l'échantillon est stratifié en H strates, la taille optimale de l'échantillon du sous-échantillon, selon Fuller (2009), sera :

$$m_{\text{sous-échantillon}} = m \sqrt{\frac{\sigma_w^2 C_1}{\sigma_b^2 C_2}}$$

Où :

$$\sigma_w^2 = \frac{\sum_{h=1}^H \sigma_{yh}^2}{H}$$

$$\sigma_b^2 = \sigma_y^2 - \sigma_w^2$$

σ_y^2 et σ_{yh}^2 sont les variances de y dans l'échantillon complet et dans la strate h respectivement.

m est la taille de l'échantillon de l'enquête.

C_1 et C_2 sont les coûts de l'interview par déclaration et mesure objective, respectivement. En cas d'informations fiables sur les coûts unitaires C_1 et C_2 ne sont pas disponibles, une valeur de remplacement de $\frac{C_1}{C_2}$ peut être prise en compte ici sur la base d'avis d'experts et de simulations de coûts.

IX.B. Sous-échantillonnage pour des modules/questionnaires spécifiques

Globalement, l'Initiative 50x2030 propose un système d'enquête modulaire avec un outil d'enquête de base annuel axé sur les cultures, l'élevage, l'aquaculture, la pêche et la production forestière (CORE), et un ensemble de modules spécialisés (voir section II) :

- Revenu agricole, main-d'œuvre et productivité
- Méthodes de production et environnement (PME)
- Machines, équipement et actifs (MEA)
- Revenu non agricole et niveau de vie des ménages (RNA-NVM)

Ces modules spécialisés sont administrés à des fréquences plus basses et font tous partie du Programme intégré d'enquêtes agricoles et rurales. Toutefois, l'enquête RNA-NVM n'est pas administrée dans le cadre du Programme d'enquête agricole.

Étant donné que les modules spécialisés sont généralement considérés pour les indicateurs et les analyses nécessaires au niveau national, il est suggéré d'en tenir compte dans les domaines d'estimation plus vastes, peut-être même uniquement au niveau national seulement et non au niveau des zones administratives infranationales. Par exemple, alors que le questionnaire CORE peut rechercher des données représentatives au niveau cantonal ou départemental et potentiellement représentatives de cultures spécifiques, des questionnaires comme le questionnaire RNA-NVM peuvent être utilisés avec un échantillon représentatif au niveau régional ou national, en fonction des besoins des pays. Cette recommandation présente certains avantages puisqu'elle réduira la charge pesant sur les personnes interrogées dans une partie importante de l'échantillon. La réduction de la durée des interviews a un effet direct sur le coût de la collecte des données et sur la réduction des erreurs d'estimation en raison de la pénibilité éprouvée par les personnes interrogées. Toutefois, l'utilisation d'un sous-échantillon dans ce cadre peut avoir une incidence sur la précision des estimateurs, puisque l'estimation sera effectuée dans un modèle d'échantillonnage en deux phases.

Afin d'améliorer l'estimation du revenu total (agricole et non agricole), il est recommandé d'envisager un sous-échantillon unique de ménages agricoles pour les modules RNA-NVM et ILP (les mêmes ménages agricoles recevront les modules RNA et ILP). Étant donné que les ménages non agricoles ne seront administrés que par le module RNA-NVM, il n'est évidemment pas nécessaire de sélectionner un sous-échantillon de ces populations pour l'enquête.

Le sous-échantillon peut être sélectionné par un simple échantillonnage aléatoire sans remplacement à l'aide de l'échantillon de grande taille. Toutefois, s'il existe des informations pertinentes dans l'échantillon de grande taille pour la stratification, il peut être stratifié avant la sélection de l'échantillon. Par exemple, dans le cadre d'un échantillonnage en deux étapes, les renseignements recueillis avec le questionnaire d'établissement des listes dans les unités primaires d'échantillonnage peuvent être utiles pour une telle stratification.

IX.C. Cas des coupes-témoins

La production de données sur des coupes-témoins est une activité exigeante, mais qui ajoute une valeur significative aux opérations d'enquête. Elle nécessite l'acquisition d'équipements spécifiques et au moins deux visites de l'agent recenseur de l'exploitation (pendant les périodes de plantation et de récolte). Certains pays, comme le Niger et le Burkina Faso, procèdent à des coupes-témoins dans le cadre de leurs enquêtes agricoles sur l'ensemble de l'échantillon des exploitations et sur toutes les parcelles couvertes par l'opération. Toutefois, un sous-échantillon d'exploitations et/ou de parcelles peut être utilisé, en particulier lorsque seules des estimations nationales des rendements des cultures sont attendues de l'enquête ou si les résultats des coupes-témoins sont destinés à être utilisés pour corriger les estimations de rendement fondées sur les déclarations des exploitants.

Méthodes de sous-échantillonnage

Les options de sous-échantillonnage pour les coupes-témoins sont les suivantes :

- Sélection directe d'un sous-échantillon de parcelles (en contournant le niveau individuel de l'exploitation) pour l'utilisation de coupes-témoins. Il s'agit d'une option efficace (en termes de qualité d'estimation) mais qui comporte certaines contraintes opérationnelles. En fait, elle ne peut être réalisée qu'après l'établissement d'une liste de toutes les parcelles des exploitations. Par conséquent, cette liste doit être complétée ponctuellement pour permettre le traitement des données et la sélection du sous-échantillon de parcelles. En outre, de parcelles de grande taille ou des parcelles très éloignées de l'exploitation peuvent apparaître dans l'échantillon, ce qui augmente les coûts opérationnels.
- Sélection d'un sous-échantillon d'exploitations et couverture de la totalité ou de certaines parcelles de chaque exploitation sous-échantillonnée par l'opération de coupes-témoins. Cela peut permettre de couvrir plus de parcelles que l'option précédente et à un coût moindre, puisque l'exploitation est un cluster (groupe) de parcelles. Cependant, cela est moins efficace que l'option précédente, car l'échantillonnage par unité sectorielle ou « cluster » entraîne une plus grande variance.

IX.D. Cas de pertes après récolte au niveau de l'exploitation

La FAO examine les principales opérations de post-production au cours desquelles les pertes de récolte se produisent à différents stades de la chaîne de valeur (FAO, 2018c). Au niveau de l'exploitation, dans le cas des céréales et des légumineux, les pertes se produisent principalement au cours des opérations suivantes : la récolte, le battage ou le bombardement, le nettoyage ou la vannerie et le séchage et le stockage dans l'exploitation (FAO, 2018a). Les Directives recommandent d'utiliser les enquêtes par sondage probabilistes comme base de toute évaluation des pertes, complétée par d'autres méthodes qui peuvent être utilisées

principalement comme évaluations préliminaires ou pour analyser plus en détail certains aspects liés aux pertes après récolte. Avec de telles enquêtes, les mesures des pertes peuvent être (i) objectives - tirées des coupes-témoins faites sur le terrain ou de l'analyse en laboratoire des grains échantillonnés dans les installations de stockage, ou (ii) subjectives - en demandant aux personnes interrogées (exploitant agricole, exploitant de l'installation de stockage, etc.) de fournir leur propre estimation des pertes.

Dans le cadre de l'Initiative, il est recommandé de procéder aux évaluations des pertes après récolte à l'aide d'un sous-échantillon pour plusieurs raisons, notamment la réduction des coûts opérationnels et la pénibilité pour les personnes interrogées. En fait, les deux options de mesure des pertes (objective et subjective) sont relativement coûteuses, prennent beaucoup de temps et nécessitent un personnel bien formé (FAO, 2018a). Les mesures objectives sont particulièrement coûteuses et nécessitent de nombreuses visites des recenseurs dans les exploitations. Les évaluations subjectives sont moins coûteuses mais nécessitent des visites supplémentaires dans l'exploitation, notamment pour recueillir des informations sur les pertes de stockage, de sorte que les coûts supplémentaires peuvent être importants. En outre, les tests sur le terrain effectués par la FAO dans le cadre de la Stratégie mondiale ont révélé des erreurs importantes lors de la comparaison de mesures objectives et subjectives au Ghana, au Malawi, en Namibie et au Zimbabwe (FAO, 2018b).

Un autre motif de sous-échantillonnage est que les informations sur les pertes après récolte sont généralement nécessaires au niveau national, ce qui signifie qu'une estimation infranationale n'est pas nécessaire. En outre, en ce qui concerne la demande internationale, l'ODD 12.3.1 sur l'Indice Mondial des Perte Alimentaires doit être satisfait au niveau des pays.

Les indicateurs importants concernant les pertes après récolte sont les proportions des pertes subies par les cultures et les exploitations agricoles. La culture concernée doit être précisée, bien qu'il s'agisse généralement de céréales et de légumineux. Par conséquent, le sous-échantillonnage devrait être effectué parmi les exploitations produisant les cultures cibles.

IX.E. Estimation avec sous-échantillonnage

Si un sous-échantillonnage est utilisé pour recueillir des informations spécifiques, l'estimation peut être faite en utilisant i) un modèle d'échantillonnage en trois étapes ou ii) un modèle d'échantillonnage en deux étapes.

Si l'ensemble de l'échantillon d'exploitations est couvert et qu'un sous-échantillon de parcelles est sélectionné dans chacune d'elles, cela correspond à une sélection d'échantillons en trois étapes. Les nouveaux poids d'échantillonnage devraient être calculés pour les parcelles sous-échantillonnées en multipliant l'inverse de leur probabilité de sélection et des poids d'échantillonnage des exploitations. Les estimateurs de variances fournis pour la conception en deux étapes peuvent encore être utilisés ici avec peu d'adaptations.

Toutefois, le choix d'un sous-échantillon d'exploitations correspondrait à un échantillonnage en deux phases. L'estimation peut être faite en utilisant des estimateurs de régression ou de ratio. Les estimateurs de régression sont jugés plus efficaces dans ce contexte (Cochran, 1977). Considérons l'utilisation du sous-échantillonnage pour des mesures objectives (métrages GPS ou coupes-témoins) dans un modèle en deux étapes (sous-échantillon d'exploitations). Par souci de simplicité, nous examinerons le cas d'un échantillon d'exploitations sélectionné par une méthode aléatoire simple sans remplacement, à partir duquel un sous-échantillon aléatoire simple d'exploitations est sélectionné pour les mesures objectives. Considérons y , le rendement mesuré à l'aide des données des coupes-témoins dans le sous-échantillon et x , le rendement recueilli par déclaration sur l'ensemble de l'échantillon. D'après Sitter (1997), l'estimateur de régression utilisé pour estimer un rendement moyen plus précis est :

$$\bar{y}_{reg} = \bar{y} + \beta(\bar{x} - \bar{x}_{\text{sous-échantillon}})$$

Où :

$\bar{x}_{\text{sous-échantillon}}$ est la moyenne de x dans le sous-échantillon et β le coefficient de régression des moindres carrés de Y sur x dans le sous-échantillon.

Un estimateur de la variance de \bar{y}_{reg} est :

$$\tilde{v}(\bar{y}_{reg}) = \left(\frac{1}{m_{\text{sous-échantillon}}} - \frac{1}{m} \right) s_d^2 + \left(\frac{1}{m} - \frac{1}{M} \right) s_y^2$$

Où :

- $m_{\text{sous-échantillon}}$ = taille du sous-échantillon
- m = taille de l'ensemble de l'échantillon
- M = taille de la population des exploitations
- s_d^2 = la variance de l'échantillon des quantités $d_i = y_i - \bar{y}_i - \beta(x_i - \bar{x}_i)$

X. Couverture des exploitations spéciales (commerciales, grandes, modernes, etc.) et estimation sur base de sondage double

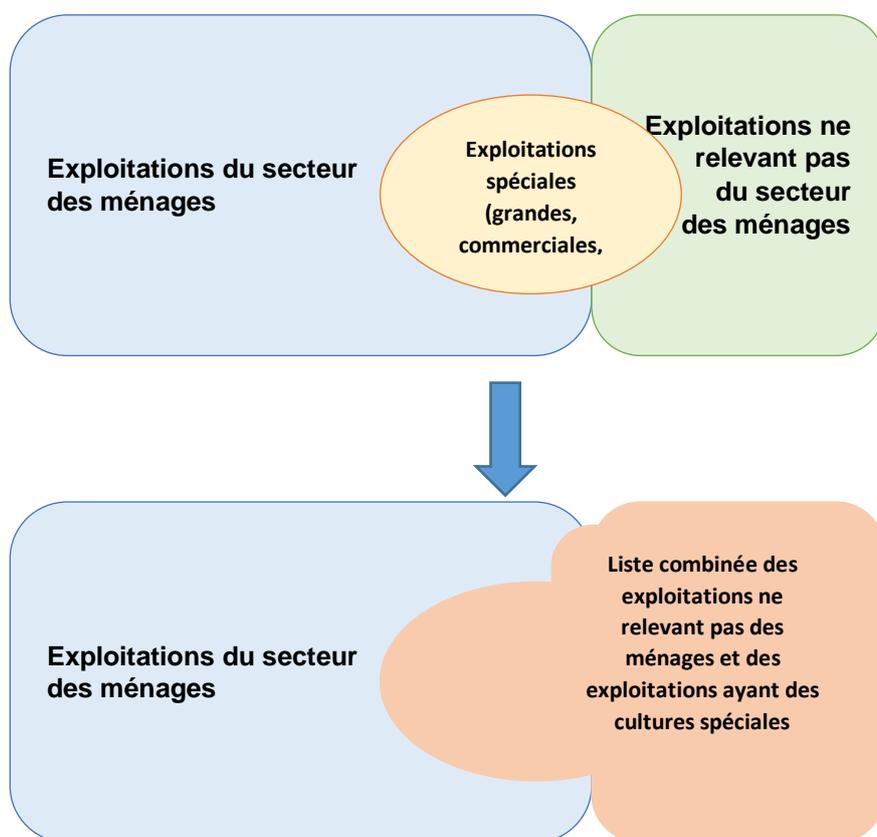
La définition adoptée par l'Initiative pour les exploitations agricoles du secteur ne relevant pas des ménages permet d'éviter le recoupement de cette population avec la population des exploitations agricoles du secteur des ménages. Cependant, il arrive souvent que des pays s'intéressent à certaines exploitations spéciales (par exemple, les exploitations commerciales, les exploitations modernes, les grandes exploitations, les exploitations biologiques, les exploitations produisant des cultures spécifiques, etc.) pour diverses raisons. En particulier, ils ont besoin d'informations pour prendre des décisions relatives à des types d'exploitations spécifiques ou au suivi des programmes agricoles d'aide à ces exploitations. En outre, des listes complémentaires d'exploitations ayant des exploitations spéciales sont recommandées dans certains cas (voir section IV), afin d'améliorer la couverture des cadres d'échantillonnage, comme les exploitations agricoles sans terre d'élevage de bétail ou les grandes exploitations commerciales. Les exploitations agricoles qui ont la charge de ces exploitations spéciales peuvent appartenir à la fois au secteur des ménages et au secteur ne relevant pas des ménages.

Lorsqu'un pays s'intéresse à des exploitations agricoles spéciales, il convient d'élaborer une définition claire et de dresser une liste des exploitations spéciales, établie à l'aide de données administratives et/ou d'opérations d'établissement de listes comme base d'échantillonnage de ces exploitations. Cette liste se recoupera soit avec le cadre des exploitations du secteur des ménages, soit avec celui des exploitations ne relevant pas du secteur des ménages ou les deux, ce qui mènera inévitablement à une enquête sur base de double sondage.

En ce qui concerne la méthode d'échantillonnage des exploitations agricoles spéciales, on peut observer que certains pays optent pour la création d'une strate explicite de ménages agricoles exploitant des exploitations agricoles spéciales, après avoir inscrit les opérations sur la liste de l'échantillon d'unités de production utilisées pour la sélection des exploitations du secteur des ménages. Toutefois, cette approche présente certaines faiblesses qui affectent l'efficacité de l'échantillon, surtout lorsque la proportion de ménages exploitant des exploitations spéciales est faible et/ou que la répartition géographique de ces exploitations est asymétrique. Cela ne garantit pas une estimation fiable de la population des exploitations spéciales, étant donné que la stratification explicite est effectuée dans un échantillon et peut conduire à un sous-échantillonnage des exploitations du secteur des ménages, ce qui affecte la précision des estimations.

Dans la pratique, lorsque des exploitations spéciales présentent un intérêt et que des bases de sondage de type liste sont utilisées comme cadres d'échantillonnage, une liste unique peut être établie à la fois pour les populations ayant des exploitations spéciales et pour les populations ayant des exploitations ne relevant pas du secteur des ménages, étant donné que le même plan d'échantillonnage (stratifié à une étape) est recommandée pour les deux populations et que les stratégies d'établissement des listes sont assez semblables dans les deux cas. Ainsi, en fin de compte, cette dernière liste et la liste des exploitations du secteur des ménages constitueront une base à double sondage qui recoupera et représentera le cadre d'échantillonnage couvrant toutes les exploitations et permettant la production de statistiques fiables sur les exploitations spéciales. Les unités qui se recoupent sont simplement des exploitations du secteur des ménages exploitant des exploitations agricoles spéciales.

FIGURE 4. PROCÉDURE D'ÉLABORATION DES CADRES POUR LA COUVERTURE DES EXPLOITATIONS SPÉCIALES



X.A. Approche de filtrage

L'approche de l'estimateur de filtrage consiste à retirer les unités qui se recoupent, soit d'un cadre avant la sélection de l'échantillon, soit d'un échantillon avant la collecte des données. Il peut être considéré comme un cas spécial d'échantillonnage stratifié et, par conséquent, l'estimation est simple. Les inconvénients communs des approches de filtrage sont qu'elles peuvent être gourmandes en ressources, sujettes à erreur et constituer une occasion manquée de recueillir des données auprès d'une personne disposée à participer.

X.A.1. Filtrage au niveau des cadres ou bases

Cette approche de filtrage consiste à supprimer le recouplement entre les cadres (en dédoublant ces cadres) avant la sélection de l'échantillon. Par exemple, dans le cas d'une enquête à double base sur les exploitations familiales et les exploitations spéciales, l'examen préalable consistera à retirer les ménages ayant des exploitations spéciales des listes complètes des ménages agricoles des ZD avant de sélectionner l'échantillon des ménages agricoles. Avant de procéder à de tels retraits, il est important de s'assurer que les exploitations spéciales filtrées dans les ZD, sont incluses dans la base des exploitations spéciales. Cela peut augmenter le temps d'inscription dans les listes des ZD, lorsque l'identification d'exploitations agricoles spéciales nécessite de nombreuses questions supplémentaires dans le formulaire de recensement.

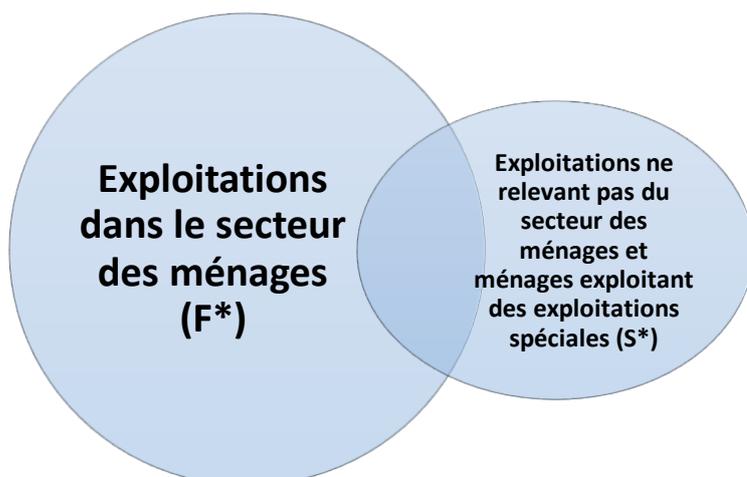
X.A.2. Filtrage au niveau des échantillons

Dans cette approche de filtrage, le recouplement entre une base et l'échantillon sélectionné à partir de l'autre base (habituellement le cadre le plus coûteux) est supprimé. Par exemple, lorsqu'on utilise une base à double sondage avec une base aréolaire et une base de type liste, les exploitations agricoles identifiées dans l'échantillon aréolaire (p. ex., échantillon de segments) sont retirées de la base de type liste avant de sélectionner l'échantillon à partir de cette liste. Le filtrage au niveau de l'échantillon est jugé acceptable lorsque le recouplement est de relativement faible importance. Un inconvénient particulier de cette approche est qu'elle peut augmenter les erreurs de non-réponse.

X.B. Estimateurs à base double

Prenons la situation typique d'un pays qui s'intéresse à des exploitations agricoles spéciales et qui utilise une base à sondage double constituée d'une liste d'exploitations du secteur des ménages et d'une liste combinée d'exploitations non familiales et de ménages exploitant des exploitations agricoles spéciales.

FIGURE 5. ILLUSTRATION D'UNE BASE À DOUBLE SONDRAGE COUVRANT LES EXPLOITATIONS AGRICOLES ET LES EXPLOITATIONS SPÉCIALES



Soit N le nombre total d'exploitations agricoles dans le domaine de l'estimation (ou domaine d'inférence), F^* la population des exploitations dans le secteur des ménages et S^* la population des exploitations hors ménages et des ménages exploitant des exploitations spéciales. La population des exploitations agricoles de ces deux bases ($F^* \cup S^*$) peut être divisée en trois sous-populations qui s'excluent mutuellement :

F : la population des exploitations familiales n'exploitant pas d'exploitations spéciales, avec une taille N_f

FS : le recoupement de la population des exploitations familiales exploitant des cultures spéciales, avec une taille N_{fs}

S : la population des exploitations ne relevant pas du secteur des ménages ayant une taille N_s

Par conséquent : $N = N_f + N_{fs} + N_s$

Si Y est une variable d'intérêt (par exemple, la superficie agricole plantée) dans la population des exploitations agricoles et si y_k est sa valeur sur l'unité k , pour $k = 1, \dots, N$.

L'objectif est de faire, à partir des données de deux sondages indépendants (ménages et exploitations agricoles enregistrées), une estimation de la population totale $Y = \sum_{k=1}^N y_k$, selon la formule suivante :

$$Y = Y_f + Y_{fs} + Y_s = \sum_{k=1}^{N_f} y_k + \sum_{k=1}^{N_{fs}} y_k + \sum_{k=1}^{N_s} y_k$$

La simple somme des deux totaux de Y , estimée à partir des deux échantillons, est biaisée en raison du recoupement entre les deux bases de sondage utilisées pour sélectionner les échantillons. Différentes méthodes sont proposées dans la littérature pour effectuer une estimation sur enquête réalisée sur une base à double sondage. Les plus célèbres ont été proposées par Hartley (1962 ; 1974), Fuller et Burmeister (1972), Bankier (1986), Kalton et Anderson (1986), Skinner (1991), Skinner et Rao (1996), Singh et Wu (1996) ; 2003), Lohr et Rao (2006), Mecatti (2007) et Singh et Mecatti (2011).

L'estimation en général, et l'estimation de la variance en particulier, ne sont pas toujours simples dans les enquêtes à base de sondage double. Arcos et autres (2015) notent que les logiciels standards pour les enquêtes complexes ne peuvent pas être utilisés directement lorsque l'échantillon est obtenu à partir d'une enquête à base de sondage double parce que les estimateurs classiques fondés sur le plan sont fortement biaisés et qu'il y a une sous-estimation des écarts-types. Les auteurs ont développé le programme R « *Bases 2* » qui peut être utilisé pour les estimations en cas de base double en utilisant la plupart des méthodes mentionnées ci-dessus.

La méthode d'estimation en cas de base double proposée par Hartley (1962) consiste à pondérer les deux estimations de Y pour le domaine de recoupement (ici, la population de recoupement FS) afin d'éviter le biais de surestimation. L'estimation de Hartley est donnée par la formule suivante :

$$\hat{Y}_H = \hat{Y}_f + \theta \hat{Y}_{fs}^f + (1 - \theta) \hat{Y}_{fs}^s + \hat{Y}_s$$

Où :

- \hat{Y}_{fs}^f est l'estimateur sur le domaine FS , basé uniquement sur les données de l'échantillon des exploitations agricoles du secteur des ménages
- \hat{Y}_{fs}^s est l'estimateur sur le domaine FS , basé uniquement sur les données de l'échantillon des exploitations ne relevant pas du secteur des ménages ou des exploitations spéciales
- θ est une constante arbitraire, telle que $0 \leq \theta \leq 1$

On remarque aisément que l'utilisation de $\theta = 0$ ou $\theta = 1$ est équivalente à une méthode de filtrage au niveau de l'échantillon (voir section 5.3.2.). Hartley (1974) a proposé une valeur optimale de θ , minimisant la variance de \hat{Y}_H :

$$\theta_{opt} = \frac{V(\hat{Y}_{fs}^s) + Cov(\hat{Y}_s, \hat{Y}_{fs}^s) - Cov(\hat{Y}_f, \hat{Y}_{fs}^f)}{V(\hat{Y}_{fs}^f) + V(\hat{Y}_{fs}^s)}$$

Cette valeur optimale peut être estimée à l'aide du programme R « *Bases 2* » (Arcos et al., 2015). Le principal inconvénient est qu'elle est spécifique à une variable, c'est-à-dire qu'elle doit être estimée pour chaque indicateur.

L'estimateur de Hartley utilisant $\theta = 1/2$ (également appelé estimateur moyen) est équivalent à l'estimateur de multiplicité proposé par Mecatti (2007) pour les enquêtes à bases multiples et contient des caractéristiques qui pourraient intéresser certains pays. Lohr (2011) mentionne que la valeur de $\theta = 1/2$ est fréquemment recommandée avec l'estimateur de Hartley. L'estimateur moyen n'est peut-être pas le plus efficace dans de nombreux contextes par rapport à d'autres estimateurs pour base double, mais il peut être recommandé aux pays parce qu'il offre des avantages opérationnels. Premièrement, il est simple à calculer et à mettre en œuvre car la valeur de θ ne dépend pas de la quantité d'intérêt (Baffour et al., 2016), et sa variance est assez facile à estimer. En outre, il a donné de bons résultats dans un certain nombre d'expériences retracées dans la littérature, y compris l'étude à base double sur les numéros de téléphone fixe et mobile de Brick et autres (2006) et l'étude de Ferraz et autres (2017) sur l'utilisation des bases à double sondage avec base aréolaire et base de type liste dans les enquêtes agricoles. Ferraz et autres (2017) notent que l'estimateur moyen est plus efficace que l'estimateur de filtrage. Chauvet et Tandeau de Marsac (2014) ont entrepris des simulations comparant l'efficacité de plusieurs estimateurs pour base double dans des plans d'échantillonnage en deux étapes et ont conclu qu'un estimateur simple est parfois préférable, même s'il n'utilise qu'une partie des renseignements recueillis.

L'estimation est simple à l'aide de l'estimateur moyen (ou de multiplicité), qui peut être présenté comme suit :

$$\hat{Y} = \hat{Y}_f + \frac{1}{2}\hat{Y}_{fs}^f + \frac{1}{2}\hat{Y}_{fs}^s + \hat{Y}_s$$

Considérons :

S^f et S^s les échantillons sélectionnés parmi les populations F^* et S^* , respectivement

w_i^f : poids d'échantillonnage de l'unité i sélectionnée dans S^f

w_i^s : poids d'échantillonnage de l'unité i sélectionnée dans S^s

L'utilisation de l'estimateur moyen consiste simplement à calculer le poids ajusté comme suit :

$$w_i^{*f} = \begin{cases} w_i^f & \text{if } i \in F \\ \frac{1}{2}w_i^f & \text{if } i \in FS \end{cases}$$

$$w_i^{*s} = \begin{cases} w_i^s & \text{if } i \in S \\ \frac{1}{2}w_i^s & \text{if } i \in FS \end{cases}$$

Les poids ajustés devraient être utilisés pour estimer la population des exploitations agricoles (p. ex. superficie, production, etc.) à l'aide de l'estimateur standard d'Horvitz-Thompson. Toutefois, les poids initiaux w_i^f et w_i^s sont conservés dans les deux échantillons pour être utilisés pour les estimations spécifiques aux populations F et S , respectivement. Par exemple, l'estimation de la production des exploitations spéciales serait réalisée en utilisant l'échantillon S^s et les poids initiaux w_i^s .

Étant donné que les échantillons ont été sélectionnés indépendamment des deux bases de sondage, l'estimation de la variance est faite de façon simple à l'aide de tout logiciel statistique standard, comme suit :

$$V(\hat{Y}) = V(\hat{Y}_{hf} + \frac{1}{2}\hat{Y}_{hfr}^{hf}) + V(\frac{1}{2}\hat{Y}_{hfr}^r + \hat{Y}_r).$$

XI. Modèle de collecte de données longitudinales et procédures de suivi

Les programmes d'enquête 50x2030 recommandent la collecte annuelle de données sur le secteur agricole. D'une année à l'autre, il y a trois solutions alternatives pour les échantillons à réaliser sur ce type d'enquêtes répétées : i) choisir un nouvel échantillon chaque année (souvent appelé « coupe transversale répétée ») ; ii) utiliser le même échantillon pendant un certain nombre d'années (panel) ; et iii) changer une proportion de l'échantillon d'une année à l'autre (rotation partielle).

L'approche par panel présente généralement des coûts opérationnels moindres puisque le même échantillon est interrogé chaque année sur une certaine période de temps, en particulier pour les enquêtes qui ne nécessitent pas d'opérations de suivi intensives. Le groupe est également bien adapté à l'estimation des changements, mais l'échantillon du groupe peut ne plus être représentatif après un certain nombre d'années en raison de l'attrition de l'échantillon et des changements structurels dans la population.

Le modèle de rotation partielle est donc une bonne alternative, en particulier pour un plan d'enquête dont la durée de mise en œuvre est relativement longue, bien qu'il puisse également pâtir d'une attrition des échantillons. Il est moins coûteux que l'approche transversale répétée et permet des analyses longitudinales et des estimations plus précises des changements.

La première option améliorerait les estimations transversales annuelles si le cadre d'échantillonnage était entièrement mis à jour chaque année avant la sélection de l'échantillon. Toutefois, par rapport aux autres options, le programme d'enquête entraînera des coûts opérationnels plus élevés, y compris les coûts annuels de mise à jour du cadre d'échantillonnage et de localisation des unités d'échantillonnage pour la mise en œuvre de l'enquête. De plus, comme il n'y a pas ou peu de recoupement entre les échantillons successifs, les coupes transversales répétées présentent habituellement plus de différences entre les données des séries chronologiques, les estimations des changements sont moins précises et les analyses longitudinales sont très limitées et parfois impossibles.

Dans le cadre de l'Initiative 50x2030, les approches par panel et par rotation partielle sont conseillées comme méthodes d'échantillonnage rentables sur le long terme. Dans les pays où le taux de non-réponses unitaires est généralement élevé (comme on l'a vu dans les enquêtes précédentes), l'approche par panel devrait être évitée en raison du risque de taux de non-réponse élevé au fil du temps dû à la pénibilité ressentie par les personnes interrogées.

TABLEAU 4. AVANTAGES ET INCONVÉNIENTS DES MÉTHODES D'ÉCHANTILLONNAGE AU FIL DU TEMPS

Approche	Avantages	Inconvénients
Coupes répétées	<ul style="list-style-type: none"> ▪ Meilleure représentativité des échantillons (cadre et échantillon mis à jour) ▪ Estimations transversales plus précises 	<ul style="list-style-type: none"> ▪ Coûts opérationnels annuels élevés : mise à jour du cadre, nouvel échantillon à interviewer ▪ Impossible d'obtenir des estimations moins précises des changements et des études longitudinales ▪ Peut nécessiter un rapprochement des données d'une année à l'autre
Panel	<ul style="list-style-type: none"> ▪ Réduire la variation des estimations ▪ Estimations précises des changements ▪ Données de séries chronologiques plus lisses ▪ Faible coût opérationnel (dans le cas où sans suivi lourd) 	<ul style="list-style-type: none"> ▪ Faible représentativité de l'échantillon en cas de changements structurels importants dans la population ▪ Attrition de l'échantillon : poids pour les personnes interrogées, changement ou déplacement des unités
Rotation	<p>Comparé à la méthode des coupes répétées :</p> <ul style="list-style-type: none"> ▪ Amélioration de la précision des estimations des changements ▪ Coût opérationnel réduit 	<ul style="list-style-type: none"> ▪ Affecte la représentativité de l'échantillon en fonction de la fraction de l'échantillon

XI.A. Longueur du panel/plan de rotation

Dans le cas du Programme d'enquêtes agricoles et rurales intégrées, il serait rentable d'établir un groupe d'experts ou une rotation de trois ans, car les ménages non agricoles sont pris en compte tous les trois ans dans ce programme d'enquête. Dans le cas du Programme d'enquête agricole, il y a plus de souplesse dans le choix de la longueur du panel/de la rotation. Cependant, comme l'efficacité du panel ou de la rotation de l'échantillon diminuera avec le temps, il est important d'éviter que le panel ou la rotation s'étalent sur de longues périodes. Une période de trois à cinq ans serait appropriée, en fonction des contextes nationaux spécifiques.

XI.B. Traitement de l'attrition des échantillons

L'attrition dans une enquête répétée correspond à la disparition des unités d'échantillonnage d'une enquête à l'autre. Les échantillons d'un panel et d'une rotation peuvent souffrir d'un biais d'attrition. L'attrition influe sur les estimations transversales et longitudinales en raison de la réduction de la taille de l'échantillon et de la nature systématique potentielle des unités sujettes à attrition. Le biais d'attrition est habituellement bien traité avec les ajustements de poids pour l'estimation transversale lorsque le taux d'attrition n'est pas très élevé. Toutefois, les estimations des changements longitudinaux sont véritablement moins précises en présence d'un phénomène d'attrition. Par conséquent, il est préférable d'élaborer des stratégies pour éviter ce problème autant que possible.

La documentation comporte un certain nombre d'approches pour traiter l'attrition des échantillons, y compris la sensibilisation, l'incitation des participants, le remplacement, le suréchantillonnage, le suivi et les interviews de suivi. Dans le cadre de l'Initiative 50x2030, les remplacements ne sont pas conseillés en raison de problèmes de partialité dans la sélection (Witoelar, 2011). Les approches de suréchantillonnage et de suivi, expliquées ci-dessous, sont des solutions habituellement utilisées.

Toutefois, en présence d'une forte attrition, il serait approprié de mettre à jour l'échantillon au cours de la série d'enquêtes successives. En fait, l'adoption de la méthode par panel ou par rotation suppose implicitement qu'un changement important de la population d'intérêt ne soit pas attendu à très court terme (un à trois ans). Une telle hypothèse ne serait pas valable si une forte attrition est enregistrée.

XI.B.1. Suréchantillonnage

Le suréchantillonnage permet d'anticiper l'attrition en augmentant la taille de l'échantillon en conséquence. Il peut être effectué en multipliant la taille de l'échantillon par l'inverse du taux de réponse attendu. Les taux de réponse des campagnes d'enquête précédentes ou d'enquêtes similaires dans le pays sont généralement pris en compte. Cela atténue l'effet de l'attrition en termes d'erreur d'échantillonnage, en raison de la taille plus petite de l'échantillon. Pour éviter complètement les non-réponses unitaires lors de la première série d'enquêtes, une approche « d'échantillonnage inverse » pourrait être adoptée. D'une manière générale, il s'agit d'un modèle dans lequel les unités sont sélectionnées successivement jusqu'à l'obtention d'un nombre prédéterminé d'unités présentant une certaine caractéristique (Tillé, 2016). Afin de maximiser le taux de réponse, l'application de l'échantillonnage inverse consiste à sélectionner un échantillon plus grand et à mener des interviews jusqu'à ce que le nombre cible d'interviews valides soit atteint (Vasconcellos et al., 2005). Pour un effort plus ciblé, cette solution offre une alternative par rapport au suréchantillonnage général effectué auparavant.

XI.B.2. Suivi

Dans les enquêtes longitudinales, le suivi fait référence au suivi et aux interviews des personnes interrogées qui quittent la localité où elles ont été recensées une première fois. Les procédures de suivi visent à traiter non seulement la question des unités manquantes, mais aussi les questions liées à des changements importants dans les unités d'observation. Le suivi dans le cadre de l'Initiative 50x2030 concerne les exploitations agricoles (c'est-à-dire les ménages agricoles et les exploitations du secteur ne relevant pas des ménages).

a) Quand effectuer un suivi ?

Lorsqu'il s'agit de déterminer les conditions de suivi, il est important de tenir compte des objectifs de l'enquête (Witoelar, 2011). Dans le cadre des enquêtes agricoles, il peut être nécessaire de suivre les exploitations agricoles en cas de disparition de l'exploitation ou de modifications importantes de l'unité qui pourraient affecter, dans une large mesure, ses productions agricoles.

Disparition d'une exploitation

- Un ménage agricole qui déménage d'une localité où il a été une première fois recensé vers une autre location (p. ex. pour des raisons de migration ou autre).
- Une entreprise agricole qui change de siège social.

- Une exploitation agricole (ne relevant pas du secteur des ménages) qui fusionne complètement avec une autre exploitation. Dans le secteur des ménages, par exemple, deux ménages peuvent fusionner en un seul par mariage.

Modification spécifique de l'exploitation

- Une exploitation agricole est scindée en deux exploitations différentes ou plus. Les membres du ménage peuvent quitter le ménage en emportant avec eux une parcelle agricole ou d'autres actifs importants et constituer des ménages indépendants.
- Une exploitation agricole est scindée en de nombreuses entités, certaines étant des exploitations indépendantes et d'autres entités fusionnant avec d'autres exploitations. Par exemple, un membre d'un ménage agricole peut rejoindre un autre ménage par mariage.
- De même, une exploitation agricole peut accueillir des parties d'une autre exploitation pour joindre des actifs agricoles tels que des terres, des cheptels, etc.

b) Jusqu'où pousser le suivi ?

Afin d'éviter les biais dans l'estimation transversale, il est préférable de limiter la zone géographique du suivi au domaine géographique dans lequel l'exploitation a été échantillonnée. Par exemple, dans le cadre d'un plan en deux étapes, les zones de suivi seraient les principales unités d'échantillonnage. Si l'unité a été échantillonnée dans une zone de dénombrement considérée comme une UPE, le suivi ne sera pas effectué si l'unité est sortie de l'UPE. Cette approche pourrait influencer sur la qualité des estimations des changements ou d'autres analyses thématiques, notant toutefois que les estimations transversales sont une priorité pour les programmes d'enquête.

XII. Intégration des systèmes statistiques existants

Les pays qui souhaitent adhérer à l'Initiative 50x2030 peuvent déjà avoir une enquête sur l'agriculture et/ou les ménages qui est menée régulièrement dans le cadre du système statistique national. Selon la situation de chaque pays, certaines recommandations sont présentées ci-dessous.

XII.A. Il existe des enquêtes sur l'agriculture et les ménages dans le pays

Si le pays veut tenir des enquêtes séparées pour l'agriculture et les ménages, il peut utiliser le modèle d'intégration des données avec les deux sources de micro-données afin de développer des micro-données intégrées pour l'agriculture et les ménages aux fins d'analyses. Un document distinct est en cours d'élaboration sur l'intégration des micro-données.

Si le pays est disposé à intégrer les deux opérations d'enquête, deux options pourraient être explorées :

- (i) les deux échantillons peuvent être fusionnés (au moins dans les zones rurales) tous les trois ans pour une seule opération d'enquête à l'aide du questionnaire intégré (qui doit être adapté au contexte du pays). Cela suppose que l'enquête sur les ménages porte sur un échantillon représentatif des ménages dans le pays et que l'enquête agricole porte à la fois sur les ménages et sur les secteurs ne relevant pas des ménages. Cette option peut ne pas être rentable car la nouvelle dispersion de l'échantillon final augmentera les coûts opérationnels par rapport au plan d'échantillonnage intégré proposé ici. La probabilité de sélection des unités dans l'échantillon final peut être difficile à calculer (surtout si les plans d'échantillonnage des deux échantillons

sont différents), alors que d'autres options d'estimation, comme l'utilisation d'estimateurs à bases multiples, peuvent être explorées.

- (ii) un nouvel échantillon intégré d'exploitants agricoles et de ménages peut être sélectionné conformément à la stratégie d'échantillonnage proposée ici. Il s'agit de l'option privilégiée et la plus rentable.

Il est important de noter que l'intégration des travaux au titre des deux enquêtes peut entraîner des difficultés institutionnelles si les enquêtes sont mises en œuvre par différents ministères.

XII.B. Seul un système d'enquête agricole existe

Outre l'échantillon de ménages agricoles couvert par l'enquête agricole existante, un échantillon complémentaire de ménages non agricoles peut être sélectionné pour les années au cours desquelles l'enquête RNA-NVM est administrée. L'échantillon final de ménages peut ne pas être pleinement représentatif. En fait, la taille de l'échantillon existant de ménages agricoles peut ne pas être suffisante pour estimer de manière fiable certains indicateurs relatifs aux ménages, en particulier si le pays utilise une base aréolaire pour les enquêtes agricoles.

Là encore, la meilleure option consiste à sélectionner un nouvel échantillon intégré d'exploitants agricoles et de ménages.

XII.C. Seule une enquête auprès des ménages existe

En fonction du nombre de ménages agricoles dans l'échantillon de ménages par rapport au nombre requis de ménages agricoles pour une estimation fiable des statistiques agricoles, l'échantillon existant peut être utilisé pour la collecte de données pendant les années au cours desquelles l'enquête RNA-NVM est administrée et devra, à défaut, être modifié par un nouveau plan d'échantillonnage intégré. Si l'échantillon existant peut être utilisé, la sous-population de ménages agricoles de l'échantillon peut également être utilisée pour des enquêtes agricoles annuelles, complétées par un échantillon d'exploitations du secteur ne relevant pas des ménages.

XIII. Références

- 50x30 Initiative. (2020). A Guide to the 50x2030 Data Collection Approach: Questionnaire Design. Technical Paper Series #2. Rome.
- Abreu, D.A., Lawson, L. A., and Hickman, S. (2018). *Assessment of a Review Process for the 2017 Census of Agriculture*. Proceedings of the Joint Statistical Meetings Survey Research Methods Section.
- Arcos, A., Molina, D., Rueda, M., & Ranalli, M. G. (2015). *Frames2: A Package for Estimation in Dual Frame Surveys*. The R Journal, 7-1, 52-72.
- Baffour, B., Haynes, M., Western, M., Pennay, D., Misson, S., Martinez, A. (2016). *Weighting strategies for combining data from dual-frame telephone surveys: Emerging evidence from Australia*. Journal of Official Statistics, 32, 549-578.
- Bankier, M. D. (1986). *Estimators based on several stratified samples with applications to multiple frame surveys*. Journal of the American Statistical Association, 81(396):1074–1079.
- Bankier, M.D. (1988). "Power allocations: determining sample sizes for subnational areas." *The American Statistician* 42, 174-177.
- Barcaroli, G. Ballin, M. Odendaal, H. Pagliuca, D. Willighagen, E. Zardetto D. (2020). *SamplingStrata: optimal stratification of sampling frames for multipurpose sampling surveys*, R package. Version 1.5-1
- Bethel, J. (1989). *Sample Allocation in Multivariate Surveys*, Survey Methodology, 15,47-57
- Brick, J. M., Dipko, S., Presser, S., Tucker, C. and Yuan, Y. (2006). Nonresponse bias in a dual frame sample of cell and landline numbers. *Public Opinion Quarterly*, 70(5), 780—793.
- Brito, J. A. D. M., Veiga, T. M. da, & Silva, P. L. do N. (2019). *An optimisation algorithm applied to the one-dimensional stratification problem*. Survey Methodology, 45(2), 295–315.
- Chauvet G., Tandeau de Marsac, G. (2014). *Estimation methods on multiple sampling frames in two-stage sampling designs*. Survey Methodology, Vol. 40, No. 2, p. 335-346.
- Cochran, W.G. (1977). *Sampling Techniques*. 3rd Edition. John Wiley & Sons: New York, USA.
- Falorsi, P.D. Bako, D. Righi, P. Piersante, A. (2015). *Integrated Survey Framework*. FAO Publication. Rome
- FAO (2015a). *World Census of Agriculture 2020. Volume 1: Programme, concepts and definitions*. FAO Publication. Rome.
- FAO (2015b). *Handbook on Master Sampling Frames for Agricultural Statistics*. FAO Publication. Rome.
- FAO (2017). *Handbook on the Agricultural Integrated Survey (AGRIS)*. FAO Publication. Rome
- FAO (2018a). *Guidelines on the measurement of harvest and post-harvest losses. Recommendations on the design of a harvest and post-harvest loss statistics system for food grains (cereals and pulses)*. FAO Publication. Rome
- FAO (2018b). *Accelerated Technical Assistance Plan for Africa. Global Office Final report*. Technical Report Series GO-44-2018. FAO Publication. Rome. Available at <http://gsars.org/en/accelerated-technical-assistance-plan-for-africa-global-office-final-report/>
- FAO. (2018c). *Guidelines on improving and using administrative data in agricultural statistics*. FAO Publication. Rome. Italy.

- Ferraz, C. 2018. Master Sampling Frame: The Field Experiments conducted in Brazil. GSARS Technical Report: Rome
- Fuller, W. A. Burmeister, L. F. (1972). *Estimation for samples selected from two overlapping frames*. In ASA Proceedings of the Social Statistics Sections, pages 245–249.
- Fuller, W.A. (2009). *Sampling Statistics*. Wiley
- Hartley, H. O. (1962). *Multiple frame surveys*. In Proceedings of the American Statistical Association, Social Statistics Sections, pages 203–206.
- Hartley, H. O. (1974). *Multiple frame methodology and selected applications*. Sankhya C, 36(3):99–118.
- Kalton, G. Anderson, D. W. (1986). *Sampling rare populations*. Journal of the Royal Statistical Society A, 149 (1):65–82.
- Kish, L. (1965). *Survey Sampling*. John Wiley & Sons: New York, USA.
- Kish, L. (1987). *Statistical design for research*. New York, NY: John Wiley & Sons.
- Kokan, A. R. (1963). *Optimum Allocation in Multivariate Surveys*. Journal of the Royal Statistical Society. Series A (General), 126(4), 557.
- Kozak, M. (2006). *On Sample Allocation in Multivariate Surveys*. *Communications in Statistics - Simulation and Computation*, 35(4), 901–910.
- Lavallée, P. (2007). *Indirect Sampling*. Springer: Ottawa
- Lessler, J.T. Kalsbeek, W.D. (1992). *Nonsampling errors in surveys*. John Wiley & Sons Inc.: New York, US
- Lohr, S.L., and Rao, J.N.K. (2006). *Estimation in multiple-frame surveys*. Journal of the American Statistical Association, 101, 1019-1030.
- Lohr, S. L. (2009). *Sampling: design and analysis*. Nelson Education
- Lohr, S.L. (2011). *Alternative survey sample designs: Sampling with multiple overlapping frames*. *Survey Methodology*, Vol.37 no.2, p. 197-213.
- Mecatti, F. (2007). *A single frame multiplicity estimator for multiple frame surveys*. *Survey Methodology*, 33, 151-157.
- Singh, A. and Wu, S. (1996). *Estimation for multiframe complex surveys by modified regression*. Proceedings of the Statistical Society of Canada, Survey Methods Section, n, pages 69–77.
- Singh, A. and Wu, S. (2003). *An extension of generalized regression estimator to dual frame surveys*. Proceedings of the Joint Statistical Meeting - Section on Survey Research Methods, pages 3911–3918.
- Singh, A. and Mecatti, F. (2011) *Generalized Multiplicity-Adjusted Horvitz-Thompson Type Estimation as a Unified Approach to Multiple Frame Surveys*. *Journal of Official Statistics*, 27, 633-650.
- Sitter, R. R. (1997). *Variance Estimation for the Regression Estimator in Two-Phase Sampling*, Journal of the American Statistical Association, 92:438, 780-787
- Skinner, C. J. (1991). *On the efficiency of raking ratio estimation for multiple frame surveys*. Journal of the American Statistical Association, 86(415):779–784.
- Skinner, C. J. Rao J. N. K. (1996). *Estimation in dual frame surveys with complex designs*. Journal of the American Statistical Association, 91(443):349–356.

- Tillé, Y. (2016). *Unequal probability inverse sampling*. *Survey Methodology*, 42(2), 283–295.
<https://doi.org/10.1002/9781118162934.ch6>
- U.N. (2017). *Principles and Recommendations for Population and Housing Censuses- Revision 3*. United Nations Publication. New York, US
- Valliant, R., Dever, J. A., & Kreuter, F. (2015). Effects of cluster sizes on variance components in two-stage sampling. *Journal of Official Statistics*, 31(4), 763-782.
- Valliant, R., Dever, J. A., and Kreuter, F. (2018). *Practical tools for designing and weighting survey samples*. 2nd edition. New York: Springer.
- Vasconcellos, M. T. L. de, Silva, P. L. do N., & Szwarcwald, C. L. (2005). *Sampling design for the World Health Survey in Brazil*. *Cadernos de Saúde Pública*, 21(S), S89–S99.
- Witoelar, F. (2011). *Tracking in Longitudinal Household Surveys*. The World Bank. Washington, DC:
- Young, L.J., Lamas, A.C., & Abreu, D. A. (2017). The 2012 Census of Agriculture: a capture–recapture analysis. *Journal of Agricultural, Biological and Environmental Statistics*, 22(4), 523-539.
- Young, L. J., Hyman, M., & Rater, B. R. (2018). Exploring a big data approach to building a list frame for urban agriculture: A pilot study in the city of Baltimore. *Journal of Official Statistics*, 34(2), 323-340.



CONTACTEZ-NOUS À

info@50x2030.org

www.50x2030.org