

L'échantillonnage dans les situations de déplacement: un manuel et des exemples pratiques

Juin, 2020

NOUS SOMMES ICI !


Amorcer l'exercice de profilage


Établir le processus de profilage


Concevoir la méthodologie


Mettre en œuvre la collecte de données


Traiter et analyser les données


Valider, rapporter et diffuser

Le Manuel d'échantillonnage du JET

Parties I&II

Introduction

Avec le *JIPS Essential Toolkit* (JET, la boîte à outils essentielle de JIPS), nous nous sommes donnés pour mission de concevoir et diffuser des outils et des guides méthodologiques à la fois accessibles et capables de permettre à nos collaborateurs et partenaires à travers le monde de développer une compréhension commune des situations de déplacement. Parce qu'il donne naissance à une base de données convenue, un exercice de profilage va contribuer à la concrétisation de solutions durables pour les personnes déplacées. Afin de pouvoir partager notre expérience collective dans ce domaine et faire en sorte que personne ne parte de zéro, nous avons décomposé cet exercice en six étapes pour élaborer une approche progressive et accessible. Le JET consiste donc en un ensemble de modèles, de documents d'orientation et de checklists conçus pour guider les personnes pas à pas dans la mise en œuvre d'un exercice de profilage.

S'inscrivant dans la troisième phase du processus qui accompagne les professionnels dans la conception de la méthodologie du profilage, ce manuel sur l'échantillonnage dans les situations de déplacement en constitue l'un des éléments les plus essentiels. En effet, l'objectif qui sous-tend la conception d'une méthodologie est de s'assurer que l'approche choisie, et l'analyse qui en résulte, répondront réellement aux besoins en données du contexte et que les résultats de l'analyse seront cohérents avec les objectifs du profilage. Dans cette perspective, une méthodologie d'échantillonnage bien conçue va veiller à ce que les caractéristiques de l'échantillon soient représentatives ou aussi représentatives que possible de la population globale, en s'appuyant sur une approche comparative. C'est pour cela que le choix et la planification d'une approche d'échantillonnage pertinente vont constituer la pierre angulaire d'une collecte de données solides et fiables.

Néanmoins, ce manuel sur l'échantillonnage dans les situations de déplacement ne propose qu'une présentation sommaire des concepts-clés de l'échantillonnage (comme la population cible, le cadre d'échantillonnage ou le concept de biais). Il a pour but de constituer un document de référence pour aider les professionnels à sélectionner un échantillon lors de situations complexes, comme lors de déplacements prolongés. Ce manuel se veut aussi fonctionnel que possible en proposant des exemples concrets d'échantillonnages directement issus du soutien aux exercices de profilage apporté par JIPS au fil des années. Il inclut également des ressources pratiques essentielles.

La partie I du manuel est consacrée à l'échantillonnage en tant que méthodologie et répond à des questions-clés sur la sélection de l'échantillonnage, la terminologie, les différentes options possibles et la documentation. Elle fournit des définitions et éléments de base sur l'échantillonnage avant d'aborder ensuite les aspects techniques plus détaillés du processus. Ce dernier est décomposé en plusieurs étapes et toutes les options d'échantillonnage sont passées en revue. Elle fournit également une description détaillée de la manière dont le processus doit être mené, de comment décider quelle méthodologie choisir et des erreurs auxquelles on peut s'attendre. Cette partie est donc le document d'orientation auquel se référer pour comprendre les aspects théoriques et techniques de l'échantillonnage. Étant donné l'abondante littérature à ce sujet, elle ne prétend pas en couvrir en profondeur tous les aspects existants mais plutôt à servir de rappel et d'introduction aux concepts-clés.

La partie II constitue un complément pratique à la première partie et présente quelques-uns des exercices de profilage soutenus par JIPS. Les professionnels y trouveront une description détaillée des approches sélectionnées, des raisons pour lesquelles celles-ci ont été choisies et de leurs limites dans six contextes de déplacement différents (Salvador, Grèce, Irak, Kosovo, Somalie et Soudan). Les chercheuses et chercheurs intéressés par un exercice de profilage auront ainsi un aperçu des réalités opérationnelles et

pourront en tirer des hypothèses pour des contextes similaires. En faisant partager notre expérience de terrain, nous espérons encourager et inspirer une communauté de pratiques et montrer la voie vers des standards élevés en matière de traitement responsable des données.

JIPS, c'est quoi?

Ce manuel a été élaboré par l'équipe du Joint IDP Profiling Service (JIPS). JIPS est un service inter-agences fondé en 2009 pour aider les gouvernements et les organisations de l'humanitaire et du développement à concevoir et à mettre en œuvre des exercices de profilage lors des situations de déplacement. Notre mission principale consiste à fournir un soutien, sur place ou à distance, sous la forme d'une assistance technique, d'un renforcement des capacités, de supports de formations ainsi que des outils et des documents d'orientation.

Remerciements

Ce manuel sur l'échantillonnage a été rédigé dans le cadre de la Phase 3 – Concevoir la méthodologie – du JIPS Essential Toolkit. Ce projet est conduit par le Joint IDP Profiling Service et a pour but de renforcer les capacités en matière d'exercice de profilage. Il s'agit d'une initiative collective visant à partager notre expérience de terrain, notre expertise et nos apprentissages avec la communauté au sens large et à s'assurer que personne ne parte de zéro pour la planification et la mise en œuvre d'un exercice de profilage.

Auteurs : Marte Claussen (JIPS), Margharita Lundkvist-Houndoumadi (JIPS), Wilhelmina Welsch (JIPS), Dag Roll-Hansen (JIPS)

L'élaboration de ce manuel a grandement bénéficié des contributions et du soutien de nos partenaires de profilage, notamment Statistics Norway qui, grâce à notre partenariat, a permis le déploiement d'experts-conseils.

Tous nos remerciements et notre gratitude vont également à l'ensemble de l'équipe de JIPS pour leurs relectures et leurs apports tout au long de ce travail, et plus particulièrement à Corina Demottaz dont les remarques et contributions se sont révélées de toute première importance.

Mise en page et conception : Shorouq Manaseer and Mohammed Belo (VisuaLab)

Table des matières

<i>Le Manuel d'échantillonnage du JET Parties I&II</i>	2
Partie I: Manuel sur l'échantillonnage dans les situations de déplacement	5
<i>Introduction</i>	6
<i>Pourquoi est-il essentiel de choisir la bonne méthode d'échantillonnage ?</i>	7
<i>Quels sont les termes les plus importants ?</i>	7
<i>Choisir une méthodologie d'échantillonnage</i>	9
<i>Quelles sont les différentes options méthodologiques ?</i>	13
<i>L'importance de documenter l'approche choisie</i>	24
<i>Pour approfondir, c'est par ici !</i>	25
Partie II: Exemples des échantillonnages mis en œuvre dans différents contextes de déplacement	26
<i>Introduction et remarques générales</i>	27
<i>Tableau récapitulatif</i>	30
<i>Exercices de profilage et méthodologie d'échantillonnage</i>	32
<i>Exercice de profilage au Salvador, 2016-2018</i>	32
<i>Exercice de profilage à Thessalonique, Grèce, 2017-2019</i>	38
<i>Exercice de profilage à Erbil, Kurdistan irakien, 2015-2016</i>	44
<i>Exercice de profilage au Kosovo, 2015-2016</i>	48
<i>Exercice de profilage à Mogadiscio, Somalie, 2015-2016</i>	54
<i>Exercice de profilage au Soudan, 2017-2019</i>	58



J | JIPS
E | Essential
T | Toolkit

Partie I

Manuel sur l'échantillonnage dans les situations de déplacement



Introduction

Lorsque l'on souhaite recueillir des informations sur un groupe de personnes, on peut soit recueillir des informations sur l'intégralité de ses membres ou uniquement un échantillon de ceux-ci. La première alternative est ce qu'on appelle un « recensement ». En théorie, un recensement va permettre de collecter des données plus justes et précises. Cependant, pour les contextes de déplacement, un recensement peut s'avérer particulièrement vulnérable aux biais, d'une part en raison de certaines des spécificités du contexte mais aussi par l'ampleur et la complexité opérationnelle du processus. En outre, la conduite d'un recensement est généralement coûteuse et peut nécessiter des années de planification et de préparation.

La seconde option est ce que l'on appelle une enquête par échantillonnage. Dans le cadre d'une enquête, un échantillon (un sous-ensemble de la population), le plus représentatif possible de l'intégralité du groupe d'intérêt de l'étude, est sélectionné. Si elle est bien conçue, une enquête par échantillonnage est en mesure de fournir des informations suffisamment précises pour produire des résultats fiables. Elle peut également, dans de nombreux contextes, s'avérer moins vulnérable à certains biais¹ et est la plupart du temps moins coûteuse qu'un recensement. Il s'agit donc souvent de l'option privilégiée pour recueillir des informations dans les situations de déplacement ou dans des situations à thématique spécifique.

Le but de ce manuel est de présenter les concepts-clés de l'échantillonnage et de servir de référence pratique pour la sélection d'un échantillon représentatif de la population visée lors d'un contexte de déplacement. Il fournit également, là où cela est pertinent, des exemples des approches d'échantillonnage entreprises lors d'exercices de profilage soutenus par JIPS². Ce manuel n'a pas pour objectif de passer en revue l'intégralité des concepts statistiques théoriques autour de la sélection de l'échantillon en général, mais s'attachera en revanche à indiquer quelles sont les ressources pour ce faire lorsque cela est possible.

Le profilage

Il s'agit d'un exercice collaboratif qui cherche à **fonder une compréhension commune des situations de déplacement** et des circonstances et caractéristiques des personnes touchées. Il s'appuie sur des approches **par méthodes mixtes qui vont souvent impliquer des enquêtes par échantillonnage**. Ces dernières visent à collecter et analyser les données sur les populations déplacées, les communautés qui les accueillent et les autres populations, et à les situer au sein de considérations plus larges sur le contexte économique, politique et social du déplacement. L'objectif général est d'élaborer conjointement une **base factuelle convenue, intégrée et exhaustive à même de contribuer à des interventions humanitaires et de développement plus efficaces ainsi qu'au plaidoyer et au développement des politiques nationales** pour soutenir la concrétisation de solutions durables pour les populations déplacées.

Dans un premier temps, nous examinerons pourquoi, dans le cadre d'une enquête par échantillonnage, il est essentiel de choisir la bonne méthode d'échantillonnage pour aboutir à des résultats fiables. Puis, nous passerons en revue une partie des termes les plus importants pour la suite du manuel et qu'il est important de garder à l'esprit lors des échanges sur le choix de la méthode avec les partenaires ou collaborateurs potentiels. Enfin, nous détaillerons le processus de sélection d'une méthode d'échantillonnage.

- 1 Tels que les biais de sélection (biais d'inclusion ou d'exclusion) qui seront expliqués plus loin dans le document
- 2 Pour plus d'exemples concrets sur les différents échantillonnages utilisés lors d'exercices de profilage soutenus par JIPS, voir la partie II : Exemples des échantillonnages mis en œuvre dans différents contextes de déplacement.

Pourquoi est-il essentiel de choisir la bonne méthode d'échantillonnage ?

Choisir la bonne approche pour sélectionner l'échantillon d'une enquête par échantillonnage constitue une étape essentielle de la conception de la méthodologie globale d'un exercice de profilage. Une approche d'échantillonnage bien conçue vous permettra, avec un certain degré de certitude, de partir du principe que les caractéristiques de l'échantillon vont valoir pour l'ensemble de la population dans laquelle il a été sélectionné et de le généraliser. Il s'agit donc d'un prérequis essentiel à une analyse solide et rigoureuse.

A contrario, une approche d'échantillonnage mal conçue peut sévèrement compromettre la capacité des données collectées à répondre aux questions à l'origine de la recherche. Pour résumer, le travail consacré au choix et à la planification de l'approche d'échantillonnage la plus appropriée va permettre non seulement de réduire le risque d'erreurs mais aussi d'éviter d'éventuels coûts supplémentaires à un stade plus avancé de l'exercice de profilage.

Quels sont les termes les plus importants ?



Population cible

Désigne le groupe de population qui doit être représenté par l'échantillon.

Exemple

"Les réfugiés et demandeuses.eurs d'asile arrivés à Thessalonique après janvier 2015"



Cadre d'échantillonnage

Désigne la liste concrète des individus ou ménages de la population cible à partir de laquelle l'échantillon sera sélectionné. Il peut par exemple s'agir d'une liste d'adresses. Dans l'idéal, il faudrait avoir accès à une liste compilant l'intégralité de la population cible. Cependant, dans les contextes de déplacement cela est rarement le cas et il faut alors trouver d'autres manières de procéder pour parvenir à sélectionner un échantillon représentatif.



Échantillon représentatif

Lors du choix d'une approche d'échantillonnage, l'objectif principal est toujours d'avoir un échantillon représentatif de la population cible. Cela signifie qu'il faut sélectionner un échantillon dont les caractéristiques vont ressembler étroitement à celles de la population cible et qui en sera le reflet le plus fidèle possible. Par exemple, si 50 pourcent de la population cible est composée de femmes, l'échantillon devrait idéalement comprendre 50 pourcent de femmes. Effectuer un tirage aléatoire de l'échantillon permet généralement, et avec un niveau de certitude connu, de produire un échantillon représentatif. Des mesures additionnelles comme l'intégration de strates peuvent être mises en place pour augmenter la probabilité d'obtenir un échantillon représentatif de la population cible. La stratification fera l'objet d'explications plus détaillées plus loin dans ce document.

Niveau de confiance et marge d'erreur

Ces deux variables vont caractériser le niveau de précision des résultats de l'enquête par échantillonnage : si le taux de confiance est de 95 pourcent avec une marge d'erreur de 5 pourcent, cela signifie que, si l'enquête est menée de manière répétitive (en tirant un nouvel échantillon du même cadre d'échantillonnage à chaque fois que l'enquête est répétée), les résultats se situeront à 5 points en pourcentage des valeurs réelles de la population dans 95 pourcent des cas.

Taille de l'échantillon

Se réfère au nombre d'individus ou de ménages de l'échantillon. La taille de l'échantillon doit être déterminée sur la base de certaines variables, telle que le niveau de confiance souhaité, le niveau de marge d'erreur souhaité et le taux de réponse attendu. Le montant du financement influence souvent la taille de l'échantillon envisageable.

Taux de réponse

Désigne la part, en pourcentage, des individus/ménages de l'échantillon qui répondent à l'enquête. Plus le taux de réponse est élevé, plus les résultats seront précis. C'est pourquoi il est crucial d'avoir un taux de réponse élevé pour obtenir des chiffres exacts. Si le taux de réponse est faible, on peut tenter de le corriger en introduisant des pondérations. Il est important de noter qu'il n'est pas possible de résoudre les non-réponses en incorporant simplement des personnes et ménages non inclus initialement dans l'enquête. Cela pourrait introduire un biais dans les résultats et il ne sera plus possible d'affirmer que les résultats de l'enquête sont corrects avec un certain niveau de confiance et de marge d'erreur.

Biais

Fait référence à une situation où les résultats d'une enquête par sondage surestiment ou sous-estiment de manière systématique un paramètre de la population (une valeur qui décrit la population cible). L'apparition de biais peut être due à différents facteurs. Une description de leurs diverses causes se trouve plus loin dans ce guide.





Pondération

Est une variable parfois utilisée pour attribuer une valeur à chacune des observations d'un ensemble de données. Une pondération est attribuée à une variable pour augmenter ou diminuer son importance par rapport aux autres variables de l'ensemble de données. Dans le cadre d'enquêtes par échantillonnage, des pondérations sont parfois utilisées pour que les échantillons soient plus proches de la population cible, soit lorsqu'il s'agit d'échantillons stratifiés non-proportionnés, soit lorsque les taux de non-réponse sont élevés (voir plus loin). Lors d'une enquête, les pondérations indiquent le nombre de personnes de la population cible qu'une observation représente.

Choisir une méthodologie d'échantillonnage

Que faut-il savoir pour choisir la bonne méthodologie d'échantillonnage ?

Plusieurs points clés et facteurs contextuels sont à définir et à prendre en compte pour décider de l'approche d'échantillonnage à sélectionner dans un contexte de déplacement. Voici une liste des plus importants:

-  Déterminer le type d'analyse nécessaire pour répondre aux objectifs
-  Déterminer la population cible
-  Déterminer le cadre de l'échantillon
-  Tenir compte de la disponibilité des ressources

Exemple

Dans une enquête par échantillonnage menée dans le cadre d'un *exercice de profilage au Soudan en 2018*, les ménages de PDI vivant dans deux camps différents ont constitué l'une des deux populations cible de l'enquête. Lors du tirage de l'échantillon dans le cadre de l'enquête, les ménages de PDI vivant dans certaines parties des camps ont été sur-échantillonnés, c.à.d. avaient une plus grande probabilité d'être inclus dans l'échantillon. Ce sur-échantillonnage de certaines zones avait été causée par certaines mesures qui avaient dû être prises sur le terrain. Des pondérations ont donc été utilisées pour corriger la plus grande probabilité de certains ménages d'être inclus dans l'échantillon. Ces pondérations ont diminué l'importance des réponses des ménages sur-représentés dans l'échantillon et augmenté celle de ceux qui ne l'étaient pas.

Décider des résultats attendus de l'analyse

L'une des premières questions qui se pose lors de la planification d'une enquête par échantillonnage pour un exercice de profilage est de déterminer et définir clairement le type d'analyse recherchée par l'enquête pour répondre aux objectifs de l'étude. Avant de choisir la méthode d'échantillonnage, il faut pouvoir répondre à des questions telles que : à quoi l'enquête doit-elle être en mesure de répondre ? À propos de quoi doit-elle être en mesure de fournir des éléments de réponse ?

Le type d'analyse souhaité doit faire l'objet d'une discussion approfondie, car celui-ci s'avérera décisif pour le choix de la méthode d'échantillonnage la plus appropriée. En plus de définir clairement la ou les population(s) cible de l'étude (ce point sera développé plus en détail ci-dessous), il faudra décider d'une éventuelle désagrégation des résultats en fonction des diverses caractéristiques de la population cible et, dans ce cas, décider quelles seront les caractéristiques à désagréger. Cela implique qu'il faudra veiller à avoir des échantillons représentatifs pour chacun des sous-groupes de la population. Par exemple, si vous souhaitez obtenir davantage d'informations sur un sous-groupe donné par localisation ou nationalité, cela pourrait nécessiter la création d'un échantillon stratifié afin de s'assurer que les résultats seront représentatifs pour ce sous-groupe en question. Ce point sera abordé plus en détail ci-dessous.

Il est toutefois important de noter dès à présent que le fait d'introduire de nombreuses strates peut augmenter considérablement la taille de l'échantillon nécessaire à la production de résultats représentatifs pour chacune des strates ainsi que la quantité de ressources requises pour la réalisation de l'enquête. C'est pour cela qu'il convient de porter une grande attention aux types de niveaux d'analyse nécessaires pour répondre aux objectifs de l'exercice.

Déterminer la population cible

La population cible est le groupe de population qui doit être représenté par les résultats de l'enquête par échantillonnage. Il est essentiel que la population cible soit clairement définie, tant dans l'espace que dans le temps. Un exemple de population cible bien définie peut être trouvé dans l'exercice de profilage soutenu par JIPS à Thessalonique, Grèce, en 2017 :

☰ Exemple

"Les réfugié.e.s et demandeuses.eurs d'asile arrivés à Thessalonique après janvier 2015"

Cette population cible est définie dans l'espace – Thessalonique – et dans le temps – après janvier 2015 et jusqu'au démarrage de l'exercice. Il a aussi fallu clarifier qui pouvait être qualifié de réfugié.e. ou de demandeuse.re d'asile. Dans ce cas précis, les réfugié.e.s et demandeuse.eurs d'asile ont été définis comme des personnes en possession de cartes de pré-enregistrement ou d'enregistrement en tant que demandeuses.eurs d'asile, de décisions d'octroi d'asile, de permis de résidence ou de demandes d'asile faisant l'objet d'un recours administratif.

L'un des éléments essentiels du profilage réside dans la mise en œuvre d'une approche comparative, qui consiste à comparer la situation des divers groupes de population, par exemple : les personnes déplacées à l'intérieur de leur propre pays, les réfugiés, les personnes revenues à leur lieu d'origine, les migrants économiques et les populations qui les accueillent. Même pour les cas où l'objectif principal du profilage est d'évaluer la situation d'un seul de ces groupes, il est souvent nécessaire d'examiner la situation des différents groupes qui partagent la même localisation (par ex. la même municipalité, région ou pays) pour appréhender si la situation et les caractéristiques de ce groupe de population lui sont spécifiques ou si celles-ci sont partagées par tous.

Si vous décidez de mener une enquête par échantillonnage pour plus d'un groupe de population, il faudra alors déterminer plusieurs populations cible, qui chacune devront être clairement définies, comme précisé ci-dessus. Mais définir les populations cible dans un contexte de déplacement n'est pas toujours tâche facile et il est par exemple possible que les différents partenaires n'arrivent pas à se mettre d'accord sur leur définition. Dans certains cas, il existe des définitions internationalement reconnues sur lesquelles s'appuyer, comme pour les réfugiés³ et les PDI⁴. En revanche, dans d'autres cas, rien d'établi n'existe pour définir certaines des populations cible d'un contexte de déplacement (par ex. comment définir la « population qui accueille ») et il sera nécessaire d'élaborer et de s'accorder sur une définition concrète avec les différents partenaires de chaque étude spécifique.

Même si la tâche de définir les populations cible peut parfois se révéler ardue, il nous semble important d'insister sur **l'importance de ce travail de définition claire des populations cible concernées. Il s'agit d'un prérequis essentiel à la réalisation d'une enquête par échantillonnage** qui en retour vous facilitera la sélection de l'échantillon ainsi que la conduite de l'enquête en général et l'analyse des résultats.

3 Groupe d'experts sur les statistiques relatives aux réfugiés et aux personnes déplacées internes (2018). Recommandations internationales sur les statistiques relatives aux réfugiés. <https://www.jips.org/jips-publication/international-recommendations-on-refugee-statistics/>

4 Groupe d'experts sur les statistiques relatives aux réfugiés et aux personnes déplacées internes (2018). Recommandations internationales sur les statistiques relatives aux personnes déplacées à l'intérieur de leur propre pays https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/International-Migration/2018_1746_EN_08-E.pdf

Élaborer le cadre d'échantillonnage

La prochaine étape du processus de sélection d'un échantillon consiste à élaborer un cadre d'échantillonnage, c'est-à-dire une liste d'individus ou de ménages à partir de laquelle un échantillon pourra être tiré. Le cadre d'échantillonnage devrait idéalement **inclure toutes les personnes ou ménages de la population cible**. Le scénario idéal de toute enquête par échantillonnage est d'avoir accès ou de réussir à dresser une liste exhaustive et actualisée de la population cible. Il s'agit de la meilleure base pour la conception d'une approche d'échantillonnage à même de permettre la sélection d'un échantillon représentatif.

Pour savoir si une telle liste existe ou s'il est possible d'en établir une en combinant des listes en provenance de diverses sources, il faudra aller consulter des données secondaires. Parmi les sources secondaires pertinentes, on trouve par exemple des registres de populations qui identifient les personnes et ménages déplacés, les données d'un recensement récent, des listes de résidents des camps ou de bénéficiaires des programmes liés au déplacement.

Dans les contextes de déplacement, la disponibilité des listes identifiant les personnes ou ménages déplacés va cependant varier d'un contexte à l'autre et il arrive parfois qu'il ne soit pas possible d'établir un cadre d'échantillonnage fiable à partir de données secondaires. Cela peut être dû, par exemple, au fait qu'il s'agisse de données sensibles et délicates à partager. En outre, même si l'on a accès à une liste de la population cible, il se peut qu'on ne sache pas comment celle-ci a été réalisée et il n'est donc pas toujours possible d'évaluer si elle est de qualité suffisante.

S'il n'existe pas de liste de la population cible, il sera nécessaire d'en établir une. Toutefois, selon le contexte du déplacement, il ne sera pas forcément possible d'établir une liste complète et précise de l'intégralité de la population cible. Cela peut être dû à diverses raisons, comme par exemple parce qu'il faudrait avoir accès à des populations difficiles à atteindre et qui vivent dans des lieux dont l'accès est limité, ou parce que ces populations sont mobiles, dispersées à travers de larges zones ou encore parce qu'elles cherchent à demeurer « invisibles ». Il se peut aussi parfois que les fonds manquent tout simplement. L'établissement d'une liste par le biais d'un recensement de la population cible est coûteux et mobilise de nombreuses ressources. Les fonds disponibles vont donc souvent constituer un facteur limitant.

La possibilité d'établir ou d'avoir accès ou non à un cadre d'échantillonnage complet va avoir des conséquences sur la conception de la méthode de sélection de l'échantillon et l'interprétation des résultats. **En d'autres termes, à propos de quoi et de qui sera-t-il vraiment possible de dire quelque chose sur la base des résultats de l'enquête ? Ces différentes implications seront abordées plus loin dans ce manuel.**

Vous trouverez ci-dessous une présentation illustrée des différents scénarios auxquels on peut être confronté lors de l'établissement du cadre d'échantillonnage d'un exercice de profilage :

Le scénario idéal

Une liste actualisée et exhaustive est disponible

Lors d'un exercice de profilage au Kosovo en 2018, l'une des populations cible était les Serbes déplacés et résidant dans des zones spécifiques du Kosovo. Les partenaires impliqués dans l'exercice ont eu accès à une liste de tous les ménages serbes déplacés en provenance de la base de données d'un organisme public serbe.

Il est possible d'établir une liste actualisée et exhaustive en combinant différentes sources

Lors d'un exercice de profilage à Thessalonique en 2017, l'une des populations cible était composée des réfugiés arrivés à Thessalonique après janvier 2015. Les profileurs ont combiné les données issues d'une base de données des Nations Unies sur les réfugiés (base de données d'enregistrement ProGress du HCR) avec une liste des résidents d'un camp de réfugiés, afin d'élaborer un cadre d'échantillonnage pour cette population cible. Mais, les informations contenues dans la base de données se sont finalement avérées obsolètes, ce qui nous amène au scénario suivant.

Un scénario moins idéal, mais qui prévaut souvent

Les listes sont biaisées et/ou non actualisées

Dans les contextes de déplacement, il est rare de trouver des listes d'enregistrement totalement exactes. Cela est souvent dû au fait que la population en question est très mobile, ce qui rend difficile la mise à jour d'une liste d'adresses ou d'autres informations de contact. Lors de l'exercice de profilage à Thessalonique, il est apparu clairement que les profileurs devaient trouver un moyen de mettre à jour le cadre d'échantillonnage existant qui était devenu obsolète. Généralement, la mise à jour d'un cadre d'échantillonnage va impliquer un dénombrement ou la consolidation de sources secondaires telles que les listes des bénéficiaires des programmes ou des listes détenues par les autorités locales.

La nécessité de construire un cadre d'échantillonnage

Un exercice de profilage mené au Salvador en 2016, dont la population cible était constituée des ménages ayant fui des situations de violence entre 2006 et 2016, est un exemple de scénario dans lequel aucune liste n'était disponible ou possible à établir pour la population cible à partir de sources existantes. Les profileurs ont donc dû trouver un autre moyen d'établir un cadre d'échantillonnage et de sélectionner un échantillon représentatif. Un échantillonnage stratifié en grappes a été choisi. Cette approche sera détaillée ci-dessous.

Tenir compte de la disponibilité des ressources

Enfin, l'une des considérations importantes à prendre en compte lors du choix d'une méthode d'échantillonnage va être le temps et les fonds disponibles pour la réalisation de l'enquête. Par exemple, procéder au dénombrement complet d'une zone très vaste peut s'avérer coûteux en raison des coûts d'embauche, de formation et de rémunérations des enquêtrices. En outre, circuler à travers un large périmètre pour accéder à la population cible échantillonnée peut engendrer des frais de déplacements conséquents. Le montant du financement disponible va donc avoir une incidence sur la taille possible de l'échantillon. Des contraintes de temps peuvent également avoir un impact sur l'approche d'échantillonnage envisageable.

Il est donc indispensable de prendre en compte le temps et les fonds disponibles tout au long du processus de choix de l'approche d'échantillonnage. La quantité de ressources disponibles va particulièrement affecter le type d'analyse possible (par rapport aux degrés de désagrégation) et parfois même le nombre de populations cible et leur définition.

Quelles sont les différentes options méthodologiques ?

Cette section présente les différentes approches possibles pour le tirage d'un échantillon, les différents éléments à prendre en compte pour décider de l'approche, les limites de chacune d'elles, et donne des exemples de leur utilisation selon le contexte. Chaque approche sera abordée séparément. Néanmoins, dans de nombreux cas, une combinaison de plusieurs approches peut être utilisée, comme cela sera illustré par les exemples présentés dans la deuxième partie.

Les méthodes d'échantillonnage se divisent en deux catégories : les méthodes probabilistes et les méthodes non probabilistes. Dans les échantillons probabilistes, chacun des membres de la population cible a une probabilité connue d'être sélectionné. Cela signifie qu'il est possible de calculer la probabilité de chacun des membres d'être inclus dans l'échantillon final ainsi **que le degré de représentativité de l'échantillon par rapport à la population cible** (voir la définition de l'intervalle de confiance et de la marge d'erreur).

Lors d'un échantillonnage non probabiliste, les membres sont sélectionnés de manière non aléatoire dans la population cible. Malheureusement, **un échantillonnage non probabiliste ne permet pas de déterminer avec certitude dans quelle mesure les informations tirées de l'échantillon divergent de la population cible : cela demeure tout simplement inconnu**. Il existe néanmoins des moyens de surmonter certaines de ces incertitudes, qui seront abordés plus loin.

Une approche d'échantillonnage probabiliste est toujours l'option à privilégier, bien qu'elle ne soit pas toujours réalisable dans les contextes de déplacement. S'il n'y a pas d'autre choix que de recourir à une approche d'échantillonnage non probabiliste, il faudra s'efforcer de minimiser le biais qui découle d'une telle approche. Nous détaillerons plus tard la manière de procéder pour y parvenir.

L'échantillonnage probabiliste

Lorsque chacun des membres de la population cible a une probabilité connue d'être inclus dans l'échantillon, il est également possible d'appréhender le degré d'incertitude (c.à.d. le niveau de confiance et la marge d'erreur) des résultats, ce qui permet d'évaluer dans quelle mesure les résultats de l'enquête s'appliquent à l'ensemble de la population cible. Il s'agit d'un élément qui est également communicable au public/utilisatrices. eurs potentiels de l'enquête, ce qui permet d'accroître la confiance dans la fiabilité et la crédibilité des résultats. **Pour pouvoir recourir à une approche d'échantillonnage probabiliste, un cadre d'échantillonnage est nécessaire.**

Lorsque vous avez accès à un cadre d'échantillonnage existant

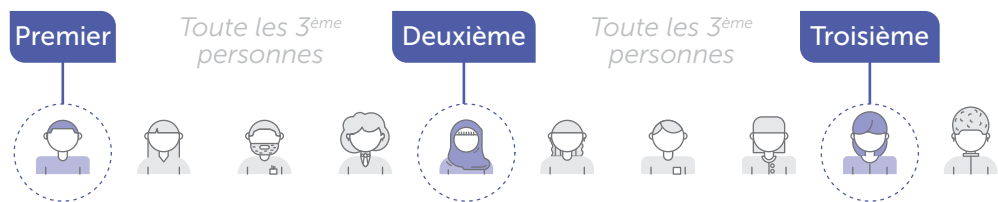
Les approches ci-dessous nécessitent un cadre d'échantillonnage à jour, précis et ayant une bonne couverture de la population cible.

Échantillon aléatoire simple



Avec une approche d'échantillonnage aléatoire simple, on sélectionne au hasard des personnes dans l'intégralité du cadre d'échantillonnage et chacune des personnes du cadre a une probabilité équivalente d'être sélectionnée. Cette approche est considérée comme la forme la plus élémentaire de l'échantillonnage aléatoire et est souvent utilisée en combinaison avec une autre approche, comme par ex. l'échantillonnage stratifié. La sélection d'un échantillon aléatoire simple peut être réalisée à l'aide d'Excel ou d'un autre logiciel informatique en choisissant au hasard des personnes ou des ménages pour l'échantillon.

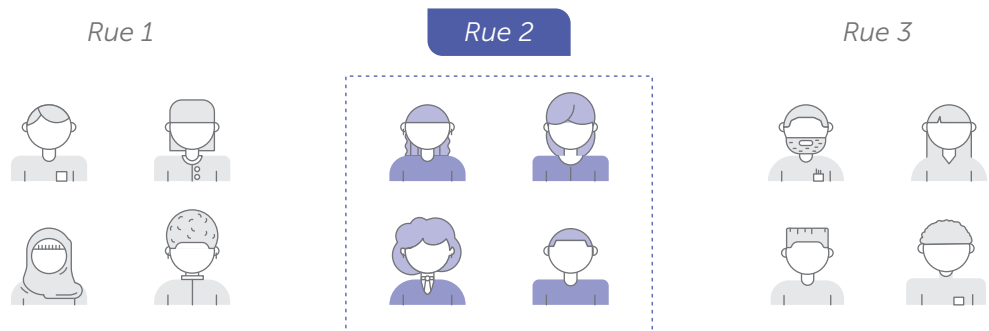
Échantillon systématique



Cette méthode d'échantillonnage est similaire à l'échantillonnage aléatoire simple, dans la mesure où toutes les personnes de la population cible ont la même probabilité d'être sélectionnées. En revanche, avec un échantillonnage systématique, l'échantillon est sélectionné à partir d'une liste selon un intervalle fixe. Supposons que vous disposiez d'une liste papier comportant les adresses de tous les ménages de la population cible et dans laquelle une adresse sur douze est sélectionnée pour l'échantillon. Cela constituerait un moyen simple de sélectionner l'échantillon mais cette méthode devrait en revanche être utilisée avec prudence si la liste des ménages présentait une certaine forme de cyclicité. Admettons que la liste des adresses soit classée par rue. Les douzièmes adresses sélectionnées seraient alors susceptibles d'avoir des points communs, par exemple se rapporter systématiquement à un appartement au rez-de-chaussée ou à une maison d'angle. Ces caractéristiques pourraient à leur tour avoir un impact sur les réponses des répondants de l'échantillon. En conséquence, s'il existe une sorte d'ordre systématique dans la liste, il faudrait de préférence essayer de la réorganiser avant de procéder à l'échantillonnage.

L'échantillonnage systématique est une technique courante mais qui doit être utilisée avec prudence pour s'assurer qu'elle donne des résultats aussi bons que l'échantillonnage aléatoire simple. Un échantillonnage systématique est également souvent associé à une autre approche, notamment à un échantillonnage stratifié.

Échantillon par grappes

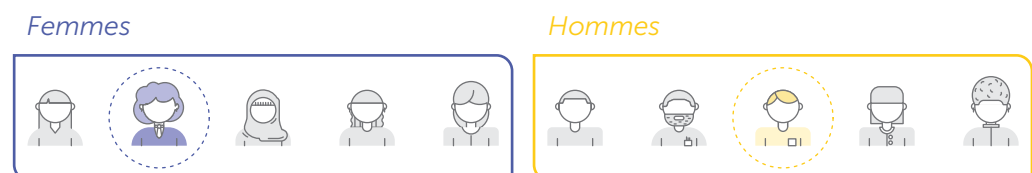


Regrouper par localisation et choisir un groupe

Pour un échantillon par grappes, on va sélectionner des groupes de personnes, comme des zones d'installation, des écoles ou des quartiers etc., plutôt que des personnes. L'avantage de cette méthode est qu'elle est souvent plus rentable et plus facile à réaliser d'un point de vue administratif. Par exemple, pour une enquête par échantillonnage auprès des écoliers d'un pays donné, il serait moins coûteux et plus rapide de sélectionner au hasard des classes dans des écoles et d'interroger tous les élèves d'une même classe, plutôt que d'interviewer au hasard des élèves disséminés dans un grand nombre de classes et d'écoles à travers tout le pays.

L'une des limites de l'échantillonnage par grappes réside dans le fait que les personnes au sein du groupe choisi sont susceptibles de présenter de plus grandes similarités. Cela peut réduire la représentativité de l'échantillon et donc augmenter le degré d'incertitude des résultats par rapport aux résultats obtenus par le biais d'un échantillonnage aléatoire simple ou systématique.

Échantillon stratifié



Dans un échantillon stratifié, la population cible est divisée en sous-groupes, appelés strates, qui ne se chevauchent pas, pour un meilleur contrôle de l'échantillon. Les divisions en strates sont effectuées sur la base d'informations auxiliaires concernant certaines des caractéristiques de la population cible, telles que le lieu de résidence (par exemple, zones rurales/urbaines, régions ou municipalités), le contexte socio-économique ou des variables démographiques. Avec un échantillonnage stratifié, chaque sous-groupe est traité comme une population cible distincte, et un échantillon distinct est donc sélectionné pour chaque strate. La division de la population cible en différentes strates présente souvent plusieurs avantages.

En premier lieu, un échantillon stratifié permet d'améliorer l'efficacité et la représentativité de l'échantillon. Avec un échantillonnage aléatoire simple, on obtient un échantillon représentatif en moyenne, mais on court un plus grand risque que l'un des sous-groupes de l'échantillon soit sur- ou sous-représenté. La stratification d'un échantillon constitue une manière d'éviter les biais et d'obtenir des estimations plus précises sur la population cible. De plus, si tous les sous-groupes identifiés sont homogènes en ce qui concerne les variables étudiées, on pourra obtenir des résultats de la qualité recherchée sur la base d'un échantillon plus restreint qu'avec un échantillonnage non-stratifié.

En second lieu, un échantillonnage stratifié peut permettre de s'assurer que les plus petits sous-groupes de la population cible sont représentés. Considérer les plus petits sous-groupes comme une population cible distincte vous permettra d'intégrer à l'étude un échantillon représentatif de chaque sous-groupe.

La répartition de la taille de l'échantillon entre les différentes strates peut être proportionnelle ou non proportionnelle à la taille de la population cible de chacune des strates. Si elle est non proportionnelle, la probabilité d'être inclus dans l'échantillon va varier selon les strates. Lorsque la population moyenne est estimée sur la base des données de l'enquête, cette représentation inégale doit être contrebalancée par l'introduction de pondérations. **Chaque pondération doit être égale au nombre de personnes ou ménages représentés par chacune des personnes ou ménages de l'échantillon. Aucune pondération n'est appliquée lors d'analyses qui impliquent une comparaison des strates entre elles. Elles ne sont utilisées que pour le calcul des chiffres qui se rapportent à l'intégralité de l'échantillon.**

Exemple

Lors d'un exercice de profilage mené au Kosovo en 2015, l'une des populations cible était les ménages serbes déplacés entre janvier 1998 et mars 2004. Une liste complète (cadre d'échantillonnage) de la population cible était disponible. La majorité de la population cible vivait dans les municipalités du nord du Kosovo et une minorité dans celles du sud.

Les partenaires du profilage disposaient d'informations indiquant que la situation des PDI du sud était différente de celle des PDI du nord. Il apparaissait donc souhaitable de pouvoir désagréger les résultats par localisation (nord et sud) afin de pouvoir comparer la situation de la population cible des deux zones. Il a été décidé de stratifier la population cible en municipalités du nord et du sud afin de s'assurer que l'échantillon global des personnes déplacées soit représentatif des populations cible de ces deux régions. La taille de l'échantillon a été répartie de manière égale entre les municipalités du nord et du sud, c'est-à-dire de manière non proportionnelle par rapport à la taille de la population cible de chaque strate, ce qui a impliqué l'application de pondérations pour l'analyse des résultats lorsque ceux-ci ont été calculés au niveau national.

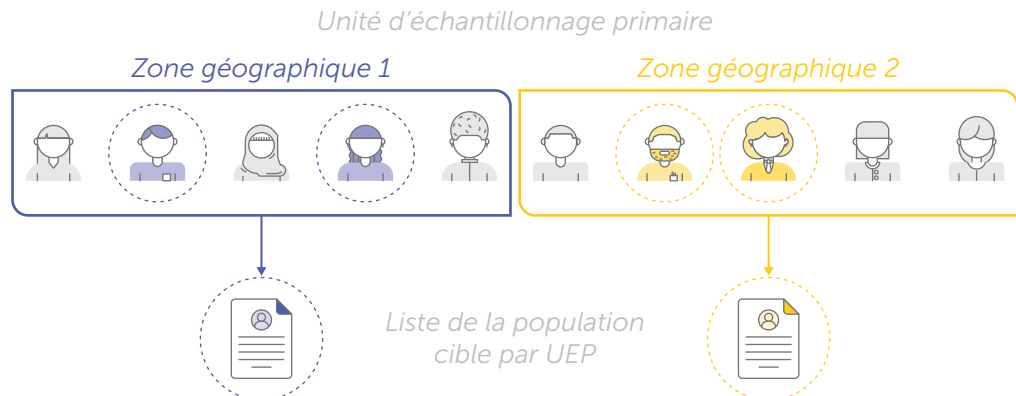
Les profileurs souhaitaient également pouvoir comparer la situation des PDI des zones urbaines et rurales. Ils ont donc stratifié davantage l'échantillon en ajoutant des strates proportionnelles aux tailles des populations cible pour les zones urbaines et rurales. En conséquence, aucune pondération n'a été nécessaire lors de la compilation des résultats pour ces deux strates.

Pour la sélection de l'échantillon, un échantillonnage aléatoire simple a été utilisé pour chacune des strates.

Lorsque vous n'avez pas accès à un cadre d'échantillonnage existant

Dans les contextes de déplacement, ne pas avoir accès à un cadre d'échantillonnage complet et à jour relève de la règle plus que de l'exception. Il est donc souvent nécessaire d'en établir un soi-même. Mais identifier et répertorier toutes les personnes d'une population cible spécifique à une étude est la plupart du temps trop coûteux et exigeant. Cela signifie que l'on ne peut pas s'appuyer uniquement sur les méthodes décrites ci-dessus. L'une des manières de relever ce défi consiste à adopter un échantillonnage à deux phases.

Échantillon aléatoire à deux phases



Cette approche est, comme son nom l'indique, menée en deux phases.

Lors de la première phase, les zones dans lesquelles vit votre population vont être divisées en zones plus petites – appelées unités d'échantillonnage primaire (UEP), qui peuvent être par exemple des municipalités, des quartiers, des blocs ou des secteurs de recensement⁵ – puis un échantillon sera sélectionné dans chaque unité. Le nombre d'UEP à sélectionner va dépendre du nombre de ménages à inclure dans l'échantillon et de la taille des UEP. Lorsque la liste des zones géographiques dans lesquelles vous souhaitez tirer un échantillon est établie, vous pouvez procéder au tirage aléatoire ou systématique de celui-ci.

Si, pour chaque unité, le nombre de personnes appartenant à votre population cible est connu, l'échantillon devra être tiré avec une probabilité proportionnelle au nombre de personnes de la population cible de l'unité. Cela signifie que les zones abritant un nombre élevé de personnes appartenant à la population cible vont avoir une probabilité plus élevée d'être sélectionnées. S'il y a deux fois plus de personnes appartenant à la population cible dans la zone A que dans la zone B, la probabilité qu'un individu ou un ménage de la zone A soit sélectionné devrait être deux fois plus élevée que celle de la zone B.

Lors de la seconde phase, une liste (un cadre d'échantillonnage) de la population cible (par exemple les ménages de personnes déplacées) des UEP sélectionnées est établie, souvent par le biais d'un dénombrement de tous les ménages de chaque unité ou en consultant l'administration locale ou les organisations locales. Un échantillon aléatoire est ensuite tiré dans chaque UEP sur la base des listes établies. Les ménages sont appelés unités d'échantillonnage secondaires (UES). Si les unités d'échantillonnage primaires sont de taille similaire, un même nombre de ménages doit être sélectionné pour chacune d'entre elles. Inversement, si les unités d'échantillonnage primaires sont de tailles différentes, le nombre d'unités d'échantillonnage secondaires sélectionnées pour chaque UEP doit être réparti proportionnellement entre les UEP en fonction du nombre d'UES de chaque UEP.

Un échantillon en deux phases a pour avantage d'aboutir à un échantillon de ménages partageant la même localisation, ce qui réduit les coûts de déplacement des enquêtrices. Cependant, la sélection en deux phases de l'échantillon va accroître l'incertitude autour des résultats et imposer ce qu'on appelle un effet de plan. Cela signifie qu'il sera nécessaire d'avoir un échantillon plus grand que lors d'un échantillonnage aléatoire simple.

⁵ Zone géographique délimitée établie par le bureau national des statistiques en vue d'un recensement.

Exemple

Lors d'un exercice de profilage à Mogadiscio, en Somalie, en 2015, les populations cible étaient les personnes déplacées, les migrants économiques et les communautés d'accueil de Mogadiscio vivant dans des implantations informelles de trois districts de la ville. Une cartographie de toutes les implantations a été réalisée, suivie d'un dénombrement des ménages pour l'ensemble des implantations identifiées afin de les regrouper en différentes populations cible.

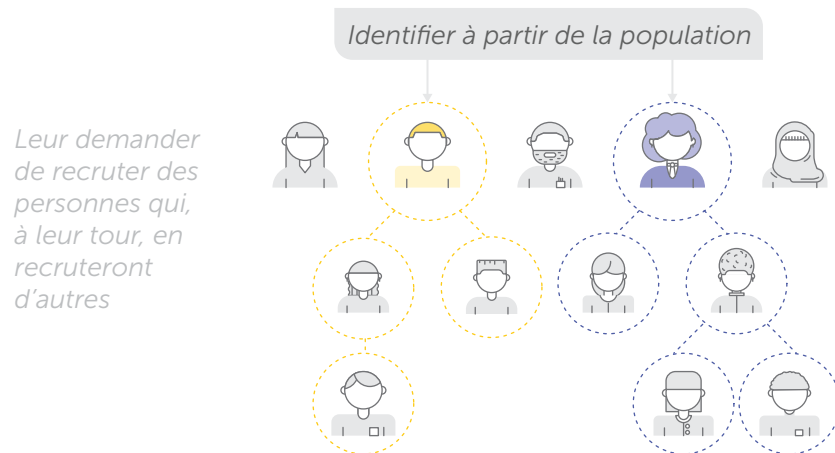
Un échantillonnage à deux phases a ensuite été mené pour la population cible des PDI. Les implantations ont été stratifiées par rapport aux trois districts. Au cours de la première phase, les implantations (unités d'échantillonnage primaires) ont été sélectionnées de manière aléatoire dans chacun des districts, sur la base d'une probabilité proportionnelle à la taille de la population déplacée de chaque implantation. Lors de la deuxième phase, un échantillon aléatoire de ménages (unités d'échantillonnage secondaires) a été constitué pour chacune des localisations sélectionnées sur la base des listes établies lors du recensement. Les ménages sélectionnés ont été interrogés dans le cadre de l'enquête par échantillonnage.

Échantillonnage non probabiliste

Dans les contextes où il n'existe pas de cadre d'échantillonnage, une approche à deux phases est l'option à privilégier. Toutefois, pour les cas où il est difficile de trouver des ménages déplacés malgré la cartographie des populations, ou bien si les ressources manquent pour effectuer un dénombrement complet des unités d'échantillonnage primaires, il est parfois nécessaire de recourir à un échantillonnage non probabiliste. L'utilisation de techniques non probabilistes pour l'établissement de l'échantillon est susceptible d'aboutir à des résultats plus biaisés qu'avec les techniques probabilistes, car certains groupes vont souvent ne pas être représentés dans l'échantillon. En outre, il est difficile d'évaluer si un échantillon non probabiliste est représentatif. En d'autres termes, il est difficile, voire impossible, compte tenu des outils existants, de calculer l'incertitude des estimations faites. Si vous décidez d'utiliser un échantillonnage non probabiliste, il faudra toujours s'assurer d'en indiquer clairement les limites lors de la diffusion des résultats de l'étude.

Toutefois, lorsqu'elle est utilisée avec prudence et accompagnée de mesures visant à augmenter la représentativité, cette technique permet d'aboutir à des résultats relativement fiables, en dépit de résultats plus incertains et de la nécessité d'un échantillon plus grand qu'avec les échantillonnages systématiques ou aléatoires.

Échantillon boule de neige



L'une des formes les plus "élémentaires" et les plus connues d'échantillonnage non probabiliste est l'approche dite « boule de neige ». L'échantillonnage boule de neige s'appuie sur la structure en réseau de la population cible. Cette méthode d'échantillonnage consiste à identifier certaines personnes appartenant à la population cible et à leur demander si elles peuvent recruter ou désigner d'autres personnes appartenant à la même population cible. Les personnes recrutées par les sujets initiaux sont à leur tour invitées à recruter ou à identifier d'autres personnes, ce qui permet à l'échantillon de grandir au fur et à mesure de la répétition du processus. Le processus est poursuivi jusqu'à ce que la taille d'échantillon nécessaire (qui doit être plus importante que dans un simple échantillonnage aléatoire) ou le point de saturation soient atteints. Le point de saturation se définit comme le moment où l'on a identifié avec certitude toutes les personnes de la population cible, et où plus aucune nouvelle identification n'est faite. Il s'agit de la seule manière de vérifier que l'on a bien abouti à un échantillon représentatif (en pratique, un comptage complet) car toute la population cible est incluse. L'échantillonnage boule de neige peut donc être également utilisé pour élaborer le cadre d'échantillonnage dans lequel un échantillon aléatoire simple peut être sélectionné. Cependant, il peut être difficile de déterminer si le point de saturation a réellement été atteint et ce, même si plus aucune nouvelle identification n'est effectuée. Cela peut être dû au fait que, dans les contextes de déplacement, certaines personnes cherchent à rester invisibles, ou parce que des personnes qui n'ont pas de réseau ou un très petit réseau ne vont jamais être identifiées. Il est donc possible d'atteindre le point de saturation dans les cas où les populations cibles sont relativement peu nombreuses, mais cela est rarement le cas lorsque la population cible est conséquente.

Une première limite importante de l'échantillonnage boule de neige réside dans le fait que les répondants sont plus susceptibles de recruter des personnes qui partagent des caractéristiques similaires aux leurs, ce qui diminue la représentativité de l'échantillon. De plus, les personnes dont le réseau est large ont plus de chances d'être incluses dans l'échantillon et d'identifier à leur tour un plus grand nombre de personnes que celles dont le réseau est petit. Il est possible d'augmenter les chances d'aboutir à un échantillon représentatif en s'assurant que le groupe de répondants initial aura des caractéristiques différentes, par exemple du point de vue des variables démographiques et socio-économiques.

Une deuxième limite de cette approche réside dans le fait que, dans certains contextes, il est possible que les répondants ne souhaitent pas identifier d'autres personnes de la population cible parce que cette identification est perçue comme une menace. Ce phénomène s'appelle le « masquage ». L'une des manières de le contourner est de demander aux répondants de recruter eux-mêmes d'autres répondants sans que ces derniers aient à transmettre leurs noms et adresses aux profileurs. De cette façon, les répondants potentiels gardent la possibilité de choisir s'ils souhaitent ou non participer à l'enquête sans avoir été identifiés au préalable.

Échantillon déterminé selon les répondants ou boule de neige probabiliste (RDS)

L'échantillonnage déterminé selon les répondants est similaire à l'échantillonnage boule de neige mais a été élaboré pour compenser le fait que, lors d'un échantillonnage boule de neige, l'échantillon est sélectionné de manière non aléatoire. Cette méthode associe l'échantillonnage boule de neige à un modèle mathématique et vise, d'une part, à générer des échantillons qui sont indépendants de l'échantillon de départ et des répondants initiaux et, d'autre part, à corriger les différences de taille entre les réseaux des personnes⁶.

Ce type d'échantillonnage cherche à limiter le biais lié aux disparités de taille entre les réseaux en mettant une limite au nombre d'identifications possibles par répondant (la limite se situe généralement à trois identifications au maximum). En pratique, cela est réalisé en ne donnant au répondant qu'un nombre limité de formulaires à transmettre à d'autres membres de la population cible. Chaque formulaire a un numéro d'identification unique et comporte également le numéro du formulaire de la personne intermédiaire. De cette manière, on peut savoir qui a été recruté par qui, ce qui permet d'éviter que des personnes identiques soient désignées plusieurs fois. Cette approche cherche également à corriger les disparités de taille entre les différents réseaux en appliquant des pondérations aux données des répondants en fonction de la taille de leur réseau. Les données fournies par des personnes dont le réseau est plus petit font l'objet d'une pondération plus forte que celles fournies par les personnes dont le réseau est plus vaste. Pour ce faire, il faudra prévoir de recueillir des informations sur la taille du réseau de chaque répondant lors de la collecte de données (par exemple en demandant à chaque répondant le nombre de personnes de la population cible qu'elle ou il connaît).

Pour ce type d'échantillonnage, on suggère de prévoir quelques mesures incitatives pour les répondants (par ex. une petite somme d'argent liquide) à la fois pour leur participation et pour le recrutement d'autres répondants. Une petite compensation financière pour chaque recrutement permet d'inciter les personnes à participer à l'enquête et donc d'agrandir l'échantillon.

Le recrutement de nouveaux répondants se poursuit jusqu'à ce que la taille d'échantillon requise soit atteinte, si la taille de la population cible est connue, ou bien jusqu'à ce que l'on soit arrivé à une certaine forme d'équilibre. L'équilibre est atteint lorsque les caractéristiques des répondants initiaux ne vont plus influencer l'échantillon, par exemple lorsque le ratio entre les hommes et les femmes reste stable tout au long des recrutements.

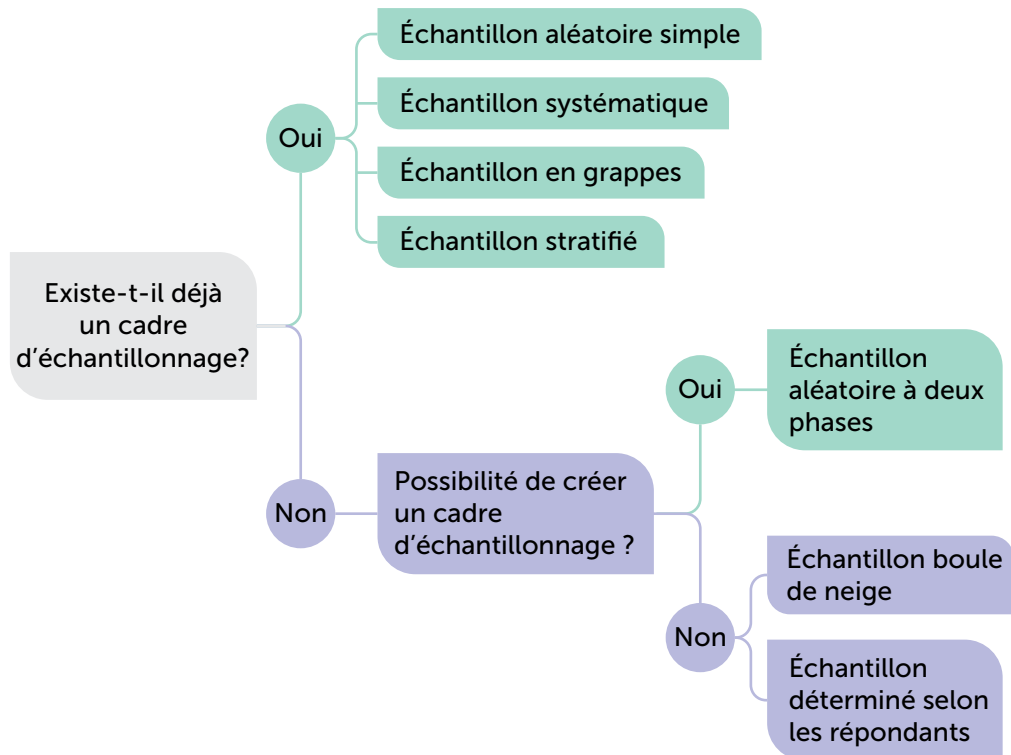
Les limites de l'échantillonnage selon les répondants incluent le fait qu'un certain nombre d'hypothèses doivent être réalisées pour que cette approche puisse aboutir à des échantillons représentatifs. Ce sera le cas par exemple de l'hypothèse selon laquelle les personnes dont le réseau est plus important ont tendance à être sur-sélectionnées dans l'échantillon. En outre, il faut utiliser des outils d'analyse de données spécialisés pour effectuer l'analyse des résultats. Il existe cependant des outils gratuits disponibles en ligne, comme par exemple RDSAT⁷.

⁶ Heckathorn, Douglas D. (1997). Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations. <http://www.respondentdrivensampling.org/>

⁷ <http://www.respondentdrivensampling.org/>

Comment décider quelle est la meilleure méthode d'échantillonnage ?

Pour récapituler la discussion ci-dessus concernant les différentes méthodes d'échantillonnage, nous les avons schématisées sous la forme d'un « arbre de décision » pour vous aider à décider du type d'échantillonnage à utiliser.



Décider de la taille de l'échantillon: combien de personnes interroger ?

La détermination de la taille d'un échantillon purement aléatoire va dépendre du niveau d'incertitude que l'on est prêt.e à accepter pour le contexte. Généralement, cette décision est fonction des coûts afférents et de l'utilisation prévue des résultats. Pour un exercice de profilage, un niveau de confiance de 95% avec une marge d'erreur +/- 5% est suffisant pour l'analyse des grandes tendances d'une population. La taille de la population globale influence également la taille de l'échantillon nécessaire. Cependant, la taille de l'échantillon va peu varier à compter de 20 000 personnes et au-delà.

Des calculateurs d'échantillon en ligne (par ex. Raosoft)⁸ peuvent être utilisés pour déterminer la taille de l'échantillon nécessaire à l'enquête. Vous trouverez ci-dessous un exemple des résultats obtenus pour diverses tailles de population, ainsi que les niveaux de confiance et marges d'erreur.

Exemple

Pour une population de 20000 personnes	Marge d'erreur de 5%	Marge d'erreur de 2%	Marge d'erreur de 1%
Niveau de confiance de 90%	270	1560	5060
Niveau de confiance de 95%	380	2140	6490
Niveau de confiance de 99%	640	3440	9070

Pour une population de 150000 personnes	Marge d'erreur de 5%	Marge d'erreur de 2%	Marge d'erreur de 1%
Niveau de confiance de 90%	270	1670	6470
Niveau de confiance de 95%	380	2360	9030
Niveau de confiance de 99%	660	4040	14940

La taille de l'échantillon doit également prendre en compte la complexité de l'échantillonnage, décrite par l'effet de plan. En bref, si la méthode adoptée est un échantillonnage à deux phases, il faudra un échantillon plus grand en raison de la complexité de cette approche. Il est alors recommandé d'utiliser un effet de plan de 1,8.

Le taux de réponse va également jouer sur la taille de l'échantillon. Il est recommandé de baser l'estimation du taux de réponse sur les enquêtes déjà existantes pour le même contexte ou sur toute autre information utile pour évaluer le taux de réponse auquel s'attendre. Si aucune information n'existe, c'est l'hypothèse d'un taux de réponse de 50% qui est généralement retenue pour le calcul de la taille de l'échantillon.

Il est important de noter que les calculateurs de taille d'échantillon sont généralement basés sur des échantillonnages aléatoires ou systématiques : lors d'un échantillonnage déterminé par les répondants, la taille de l'échantillon doit être augmentée au-delà des recommandations du calculateur. Il peut être difficile de calculer dans quelle proportion l'échantillon nécessite d'être agrandi et cette proportion va très largement dépendre des ressources disponibles.

⁸ <http://www.raosoft.com/samplesize.html>

Quelles sont les erreurs qui peuvent survenir ?

Les résultats d'une enquête par échantillonnage ne vont jamais complètement refléter les caractéristiques réelles de la population cible – des erreurs vont toujours apparaître et, comme nous l'avons vu ci-dessus, certaines approches d'échantillonnage vont produire plus d'erreurs que d'autres. Les erreurs qui surviennent lors des enquêtes par échantillonnage peuvent être divisées en deux catégories : les erreurs d'échantillonnage et les erreurs non dues à l'échantillonnage⁹.

Les *erreurs d'échantillonnage* résultent de la conception de l'échantillonnage et du fait que les résultats sont basés sur un échantillon de la population et non sur la population toute entière. L'*erreur d'estimation* est la différence entre les valeurs trouvées dans l'échantillon et les valeurs réelles de la population cible (voir la définition de la *marge d'erreur* dans la section consacrée à la terminologie). Un autre type d'erreur d'échantillonnage est appelé erreur de spécification et a lieu lorsque la probabilité réelle d'être inclus dans l'échantillon diffère de la probabilité prévue lors la conception de l'échantillonnage. Il y aura une erreur de spécification si, par exemple, la même personne ou le même ménage est inclus deux fois ou plus dans le cadre de l'échantillonnage.

Les *erreurs non dues à l'échantillonnage* surviennent lors du recueil des réponses aux questions du questionnaire. Les erreurs non dues à l'échantillonnage concernent les situations où des personnes ne répondant pas aux critères de la population cible ont malgré tout été incluses (*erreur d'inclusion*) et où des personnes répondant aux critères de la population cible ont, elles, été exclues (*erreur d'exclusion*). Dans des contextes de déplacement, les profileuse.r.s sont souvent confrontés à des erreurs d'inclusion et d'exclusion et doivent être conscients de l'existence de ce risque. Un autre type d'erreurs non due à l'échantillonnage est l'*erreur de mesure*, qui se produit si la ou le répondant.e comprend mal la question, refuse de donner la véritable réponse ou si les enquêteurs se trompent lors de l'enregistrement des réponses. Une erreur de mesure peut également se produire si les enquêtrices.eurs posent des questions orientées ou si la.le répondant.e éprouve des difficultés à se rappeler de sa réponse, par ex. pour des questions sur des périodes de remémorations. Un dernier type d'erreurs non dues à l'échantillonnage est la *non-réponse* : lorsque les personnes de l'échantillon refusent tout simplement de répondre au questionnaire ou bien ne répondent pas à toutes les questions du questionnaire.

9 Bethlehem, J. (2010). Selection bias in web surveys. https://pure.uva.nl/ws/files/1040458/92124_330481.pdf

L'importance de documenter l'approche choisie

La manière dont les résultats de l'enquête d'échantillonnage devront être interprétés et utilisés tient dans une grande mesure à la conception de l'échantillonnage, comme nous l'avons vu plus haut. C'est pourquoi il est extrêmement important que l'approche d'échantillonnage, ses limites ainsi que la manière dont l'échantillon a été sélectionné soient présentés en toute transparence aux partenaires et au grand public en général. Lors du profilage d'une situation de déplacement, les choses ne se passent souvent pas comme prévu, notamment concernant la sélection de l'échantillon. Cela signifie qu'il ne sera parfois pas possible de mettre en œuvre l'approche d'échantillonnage envisagée initialement. Si cela est le cas, il est important de garder à l'esprit que la communication publique autour de l'enquête devra refléter l'approche adoptée dans les faits pour la sélection de l'échantillon. Ainsi, même si un échantillonnage à deux phases avec dénombrements et tirage aléatoire d'unités d'échantillonnage était planifié initialement, mais qu'en raison d'événements inattendus, c'est un échantillonnage boule de neige qui a finalement été réalisé, c'est cette dernière approche qui devra être communiquée au public.

Outre une totale transparence vis-à-vis des partenaires et du grand public, il est fortement recommandé de documenter méthodiquement l'approche d'échantillonnage présentée ainsi que l'intégralité des étapes suivies pour réaliser l'échantillon de l'enquête, pour référence future en interne. Cela implique de compiler les difficultés rencontrées tout au long du processus et la manière dont celles-ci ont été résolues. La documentation des différentes étapes de la sélection d'un échantillon pourra s'avérer utile à vous-mêmes ou à vos collaborateurs pour l'élaboration des approches d'échantillonnage de futurs processus de collecte de données.

Pour approfondir, c'est par ici !

Le département des affaires économiques et sociales de la Commission statistique des Nations Unies a publié un manuel sur l'échantillonnage des enquêtes sur les ménages intitulé « Guide pratique pour la conception d'enquête sur les ménages ». Il est disponible en russe, espagnol, français, arabe et anglais :

https://unstats.un.org/unsd/publication/seriesf/Seriesf_98f.pdf

De nombreuses informations utiles pourront également être trouvées dans le manuel intitulé « Enquêtes sur les ménages dans les pays en développement et les pays en transition ». Ce manuel présente les bonnes pratiques concernant plusieurs aspects de la conduite d'enquêtes sur les ménages dans les pays en développement et en transition, notamment concernant la conception de l'échantillon, la mise en œuvre de l'enquête, les coûts et l'analyse des données de l'enquête. Cette publication est disponible en arabe, chinois, anglais, français, russe et espagnol.

https://unstats.un.org/unsd/publication/seriesf/Seriesf_96f.pdf

Le « Integrated Framework for Household Survey » élaboré par le HCR est une boîte à outils sur l'échantillonnage destinée à faciliter la conception, la collecte et l'analyse, entre autres sujets :

<https://unhcr.github.io/Integrated-framework-household-survey/Sampling.html>

Le HCR fournit également des recommandations au travers de la « Standardised Expanded Nutrition Survey (SENS) » qui peuvent être utiles pour la mise en œuvre d'enquêtes sur les ménages parmi les populations de PDI et de réfugiés :

<http://sens.unhcr.org/>

L'UNICEF propose plusieurs outils au travers de son « Multi-Indicator Cluster Survey (MICS) », notamment un outil pour la cartographie et le listage des ménages, qui peut être utile si une approche d'échantillonnage à deux phases et avec dénombrement est envisagée :

<http://mics.unicef.org/tools?round=mics5>

Voir également les manuels et formations de l'USAID intitulé « Demographic and Health Surveys (DHS) » :

<http://www.dhsprogram.com/publications/publication-dhsm4-dhs-questionnaires-and-manuals.cfm>

<http://www.dhsprogram.com/What-We-Do/Student-Resources.cfm>

La publication d'Eurostat sur « Survey Sampling Reference Guidelines » donne une bonne vue d'ensemble des méthodes d'échantillonnage probabilistes :

<https://ec.europa.eu/eurostat/documents/3859598/5901961/KS-RA-08-003-EN.PDF/833f7740-0589-47e1-99a5-c14878a2c1a8>

Voir également ce site internet pour en savoir plus sur l'échantillonnage déterminé selon les répondants :

<http://www.respondentdrivensampling.org/>

Partie II

Exemples des échantillonnages mis en œuvre dans différents contextes de déplacement



Introduction et remarques générales

L'objectif de cette partie est de passer en revue et consolider les différentes méthodes d'échantillonnage mises en œuvre dans différents exercices de profilage soutenus par JIPS au cours des dernières années. Ce manuel fournit une présentation détaillée des différentes approches utilisées et des raisons pour lesquelles elles ont été choisies tout en identifiant leurs limites éventuelles.

La réalité opérationnelle des contextes de déplacement se caractérise par un certain nombre de contraintes qui vont peser sur le choix d'une méthode d'échantillonnage. Cela peut parfois conduire à la sélection d'une méthode d'échantillonnage positionnée en deuxième choix voire même au recours à des méthodes d'échantillonnage non probabilistes. Ce type de méthode permet malgré tout de produire des résultats pertinents pour l'enquête d'échantillonnage, du moment que le processus de sélection de l'échantillon est entièrement transparent et minutieusement documenté dans la partie du rapport de profilage consacrée à la description de la méthodologie.

Cette partie vise également à fournir des exemples de la manière dont les approches d'échantillonnage ont réussi à atteindre un degré de qualité suffisant malgré les limites posées par différents contextes. Plus spécifiquement, elle propose un passage en revue des méthodologies d'échantillonnage adoptées lors de processus de profilage menés au Salvador, en Grèce, en Irak, au Kosovo, en Somalie et au Soudan. Notre objectif est que cette partie demeure évolutive et intègre des exemples complémentaires au fur et à mesure des exercices de profilage réalisés. Ce faisant, nous espérons susciter et encourager une communauté de pratiques qui valorise la transparence et montre la voie en matière de traitement responsable des données.

Les principales conclusions à retenir sont les suivantes :

L'importance de documenter la méthodologie d'échantillonnage et ses limites :

Un manque global de documentation de la méthodologie d'échantillonnage va **limiter la valeur et l'utilité des résultats de l'enquête** pour toute personne/organisation autre que les partenaires eux-mêmes. Lorsque la méthodologie et ses limites sont clairement communiquées, les données générées sont non seulement plus simples à comprendre et à utiliser, mais aussi à partager. Cette démarche est la garantie d'un processus de gestion et de traitement des données responsable.

Un manque de documentation et de transparence au sujet des limites de l'échantillonnage peut entraîner une **mauvaise interprétation des résultats** (c.à.d. de la mesure dans laquelle ils peuvent être généralisés). Cela entretient aussi l'idée courante que les approches intégralement probabilistes sont la norme, même dans des contextes de déplacement, tandis que toutes les études empiriques montrent qu'elles relèvent plus de l'exception que de la règle.

Les limites de l'échantillonnage et les objectifs de l'exercice doivent être considérés conjointement et leur interdépendance doit être communiquée clairement afin d'anticiper les attentes quant à l'analyse possible :

Les objectifs spécifiques de l'exercice et l'analyse attendue vont définir les impératifs de l'approche d'échantillonnage. Il convient donc de réfléchir soigneusement et le plus concrètement possible à leur formulation. Dans certains contextes, l'approche d'échantillonnage sera limitée par certaines réalités opérationnelles. Si cela est le cas, les objectifs devront être revus pour s'assurer qu'ils reflètent bien l'analyse envisageable.

La conception de l'approche d'échantillonnage nécessite un examen complet des données de population disponibles, une définition des populations cible et une compréhension approfondie du type d'analyse qui répondra le mieux aux objectifs :

Pour créer un cadre d'échantillonnage, il faut disposer d'informations sur la taille de la population cible et sa répartition sur le territoire. Cependant, dans la plupart des cas et pour des raisons diverses, aucun registre à jour n'est accessible. Dans la pratique, d'autres sources peuvent être disponibles et suffisantes pour établir le cadre d'échantillonnage. Cela signifie que plusieurs étapes préparatoires sont nécessaires avant que l'enquête ne puisse être réalisée (par ex. un échantillonnage à plusieurs phases). Souvent, un dénombrement s'avèrera nécessaire.

Le profilage des situations de déplacement va typiquement s'articuler autour d'un échantillonnage des populations déplacées et non déplacées pour permettre une analyse comparative. Ces populations cible devront faire l'objet de définitions claires, opérationnelles et convenues. En complément, l'analyse peut procéder à une comparaison de certains des sous-groupes, par ex. par localisation géographique, type de logement etc. Cela nécessitera l'ajout de cadres d'échantillonnage et méthodes spécifiques. L'identification des sous-groupes pertinents pour l'échantillonnage requiert une très bonne appréhension du contexte de déplacement et des programmes et politiques auxquels les résultats vont contribuer.

Dans de nombreuses configurations, il sera nécessaire de mettre en place des méthodes d'échantillonnage spécifiques pour les différents groupes cible, étant donné que la disponibilité d'un cadre d'échantillonnage va varier d'un groupe à l'autre. Cela peut limiter la comparabilité des résultats des différents groupes cible de l'enquête d'échantillonnage.

Les réalités opérationnelles de la collecte de données peuvent avoir un impact sur la conception finale de l'échantillonnage ; il est primordial que ces changements éventuels soient communiqués en totale transparence lors de la description de l'échantillonnage final :

Les difficultés couramment rencontrées lors de la collecte de données vont souvent concerner des informations personnelles obsolètes (adresses, numéros de téléphone etc.). Cela va peser sur la validité de l'approche d'échantillonnage choisie et doit être pris en compte pour que la taille de l'échantillon requise puisse être atteinte. Certaines dynamiques de la réalité opérationnelle sont également susceptibles d'avoir un effet sur l'approche d'échantillonnage et l'exercice dans son ensemble : il s'agit par exemple de l'évolution de la situation en matière de sécurité, de l'accès possible aux populations et aux zones à évaluer, de mouvements au sein de la population cible, de groupes de population qui ne souhaitent pas être identifiés, etc. Par conséquent, il est souvent nécessaire d'ajuster la conception de l'échantillonnage et parfois même de la modifier intégralement. Il est crucial que ces changements soient répertoriés dans la description de l'échantillonnage initial.

Avant de choisir une méthode d'échantillonnage, 3 éléments doivent (presque¹) toujours avoir été déterminés : **(1)** les objectifs de la recherche, **(2)** la ou les population(s) cible que l'on veut étudier et **(3)** le cadre d'échantillonnage existant ou, s'il n'y en a aucun, la stratégie retenue pour son élaboration. Ces trois points seront donc intégrés à la description de chacun des six exemples d'exercices de profilage en amont de la présentation des différentes approches d'échantillonnage. À la fin de chaque exemple, les principales limites et difficultés liées à l'approche choisie seront également identifiées.

¹ L'exception à cette règle concerne les échantillonnages boule de neige et déterminé par les répondants lorsqu'aucun cadre d'échantillonnage fiable n'est disponible et que le point (3) n'est donc rempli.

Un tableau récapitulatif permettant une vue d'ensemble des aspects clés des méthodes d'échantillonnage retenues pour les différents exercices sera présenté dans la section suivante. Ce tableau peut servir de référence lorsque l'on recherche des exemples présentant des caractéristiques différentes. Le reste du document donne un aperçu des approches d'échantillonnage mises en œuvre dans les différents exercices. Pour une explication des différents termes et méthodologies d'échantillonnage, veuillez-vous référer à la Partie I de ce Manuel sur l'échantillonnage dans les situations de déplacement ².

Profilage

Une méthode/un système collaboratif de collecte de données qui cherche à établir une **compréhension partagée des situations de déplacement** et des circonstances et caractéristiques des personnes touchées. Il s'appuie sur des **méthodes mixtes, qui impliquent souvent des enquêtes par échantillonnage** pour collecter et analyser les données sur les populations déplacées, les communautés qui les accueillent ou les autres populations. Il va également situer ces données au sein du contexte économique, politique et social plus large du déplacement. Son objectif général est d'élaborer une **base de données exhaustive et convenue pour éclairer et améliorer les interventions humanitaires et du développement, le plaidoyer et le développement des politiques nationales** qui soutiennent la mise en place de solutions durables pour les populations déplacées

Tableau récapitulatif

	Population(s) cible	Cadre	Méthodologie d'échantillonnage
Salvador, 2016-2018	Population déplacée au Salvador en raison de la violence entre 2006 et 2016	Aucune liste complète de la population déplacée n'était disponible. L'ensemble de la population du Salvador a servi de cadre.	Échantillon aléatoire stratifié
	Population non déplacée au Salvador en raison de la violence entre 2006 et 2016	L'ensemble de la population du Salvador a servi de cadre.	Échantillon stratifié à deux phases
Thessalonique, Grèce, 2017-2019	Réfugiés et demandeurs d'asile arrivés à Thessalonique après janvier 2015	Construit en combinant deux sources différentes mais obsolètes	Recensement et échantillon boule de neige (tentative d'échantillon aléatoire stratifié)
	Personnes sans document des services d'asile	Pas de liste disponible, estimation du nombre total de ménage disponible	Échantillon boule de neige
Erbil, Kurdistan irakien, 2015-2016	Réfugiés syriens résidant dans des zones spécifiques de la province d'Erbil	Liste disponible, sans certitude si complète	Échantillon aléatoire simple
	Population des PDI déplacée depuis décembre 2013 résidant dans des zones spécifiques de la province d'Erbil	Liste disponible, sans certitude si complète	Échantillon stratifié à deux phases
	Communauté d'accueil résidant dans des zones spécifiques de la province d'Erbil	Liste disponible, sans certitude si complète	Échantillon stratifié à deux phases

	Population(s) cible	Cadre	Méthodologie d'échantillonnage
Kosovo, 2015-2016	PDI albanaises déplacées entre janvier 1998 et fin mars 2004, résidant dans des municipalités spécifiques	Construit en combinant des sources obsolètes	Échantillon boule de neige (tentative d'échantillon aléatoire stratifié)
	PDI serbes déplacées entre janvier 1998 et fin mars 2004, résidant dans des logements privés et municipalités spécifiques	Liste disponible	Échantillon aléatoire simple
	PDI serbes déplacées entre janvier 1998 et fin mars 2004, résidant en centres collectifs et dans des municipalités spécifiques	Liste disponible	Recensement
	PDI Roms/Ashkalies/ Égyptiennes (RAE)	Compilée sur la base de deux sources différentes mais obsolètes et incomplètes	Recensement et échantillon boule de neige
Mogadiscio, Somalie, 2015-2016	PDI résidant dans des implantations informelles à Mogadiscio	Aucune liste disponible, créée par cartographie et dénombrement	Échantillon stratifié à deux phases
	Migrants économiques résidant dans des implantations non informelles à Mogadiscio	Aucune liste disponible, créée par cartographie et dénombrement	Échantillon aléatoire simple
	Communautés d'accueil résidant dans des implantations informelles à Mogadiscio	Aucune liste disponible, créée par cartographie et dénombrement	Échantillon aléatoire simple
Soudan, 2017-2019	PDI résidant dans les 2 camps d'Abu Shouk et El Fasher	Aucune liste disponible, créée par cartographie et dénombrement	Échantillon en grappes à deux phases
	Population non déplacée résidant dans les zones urbaines et péri-urbaines d'El Fasher	Aucune liste disponible, créée par cartographie et dénombrement	Échantillon en grappes à deux phases

Exercices de profilage et méthodologie d'échantillonnage

Exercice de profilage au Salvador, 2016-2018



Objectifs

Les objectifs du profilage étaient de déterminer le nombre de familles poussées au déplacement par la violence des dernières années, et de mieux comprendre leurs conditions de vie et besoins spécifiques³. Plus précisément, l'étude visait à :

- (1) Obtenir des estimations statistiques du **nombre de familles et d'individus** contraints de se déplacer à l'intérieur du pays en raison de violences criminelles, avec des estimations représentatives au niveau national ;
- (2) Identifier les causes et **impact spécifiques** de cette mobilité intérieure engendrée par la violence criminelle ;
- (3) Avoir une description du **profil et de l'état** de la population contrainte de se déplacer à l'intérieur de son propre pays en raison de la violence criminelle.

Pour obtenir les informations recherchées, l'étude a adopté une approche fondée sur des méthodes mixtes en s'appuyant sur des techniques de collecte de données qualitatives et quantitatives, et notamment sur une enquête d'échantillonnage, dont il est ici question.

Population cible

L'exercice de profilage comportait deux populations cible :

- (1) La principale population cible était composée de ménages⁴ **dont au moins l'un des membres avait été contraint au déplacement** en raison de la violence au Salvador entre 2006 et 2016.
- (2) En complément, une autre population cible a été définie pour permettre une comparaison entre la situation des ménages de personnes déplacées et celle des ménages de personnes non déplacées. La population cible de comparaison a été déterminée comme les ménages du Salvador dont aucun des membres n'avait été déplacé en raison de la violence entre 2006 et 2016.

³ Étude de profilage sur la mobilité intérieure due à la violence au Salvador (2018).
<https://www.jips.org/uploads/2018/03/El-Salvador-profiling-report-EN.pdf>

⁴ Défini dans l'étude comme une ou plusieurs personnes qui vivent ensemble dans la même maison et partagent les dépenses pour subvenir à leurs besoins alimentaires.

Cadre d'échantillonnage

Aucune liste des personnes déplacées ou des ménages n'était à la disposition des partenaires du profilage. Ce manque était à l'origine de l'exercice, puisque l'un de ses objectifs était de fournir une estimation du nombre des familles déplacées en raison de la violence. Des contraintes de temps et de ressources ont, en revanche, rendu impossible la constitution d'une liste complète de ces ménages déplacés. Il a donc fallu rechercher des solutions alternatives pour pouvoir établir un cadre d'échantillonnage permettant la sélection d'un échantillon représentatif et une estimation du nombre des PDI.

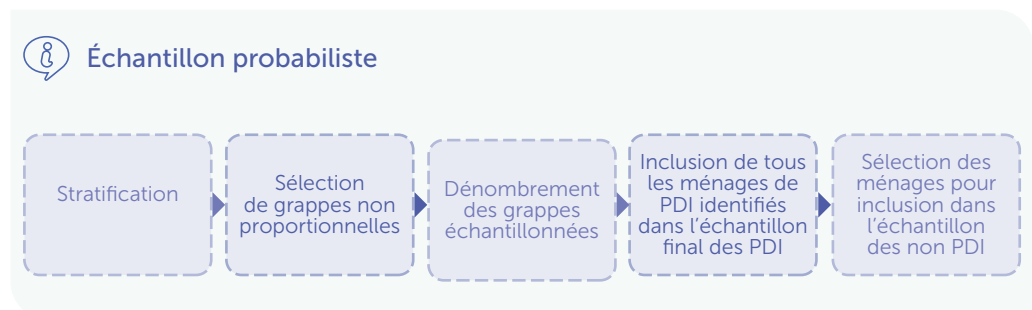
Unités d'échantillonnage et d'analyse



Méthode d'échantillonnage

Pour cet exercice de profilage, les objectifs consistaient à la fois à pouvoir donner une estimation du nombre de ménages déplacés en raison de la violence au niveau national (objectif **(1)**), et à fournir des informations sur les caractéristiques et les conditions de vie de ce groupe (objectifs **(2)** et **(3)**). Comme il n'existait pas de liste de la principale population cible et qu'il n'était pas possible de procéder à un dénombrement complet de l'ensemble du pays, les partenaires chargés du profilage ont dû élaborer une approche d'échantillonnage qui leur permettrait d'estimer la totalité de la population déplacée au niveau national sur la base d'un échantillon. En outre, les partenaires devaient s'assurer que l'échantillon serait représentatif de la population cible, afin de pouvoir atteindre l'objectif d'identification des caractéristiques et conditions de vie des ménages déplacés.

Pour la population cible déplacée, une méthode d'échantillonnage aléatoire en grappes a été sélectionnée. Pour la population cible non déplacée, c'est une méthode d'échantillonnage à deux phases qui a été adoptée. Le processus de sélection des deux échantillons peut être récapitulé comme suit :



⁵ Zone géographiquement définie et établie dans le but de procéder à un recensement. De telles zones peuvent servir à diviser géographiquement des zones plus vastes (par ex., des municipalités) afin de sélectionner un échantillon pour le profilage.

⁶ Défini dans l'étude comme une ou plusieurs personnes qui vivent ensemble dans la même maison et partagent les dépenses pour subvenir à leurs besoins alimentaires.

Dans un échantillonnage stratifié, on divise la population cible en différentes strates (groupes) et on sélectionne un échantillon dans chacune des strates. Un échantillonnage stratifié non proportionnel signifie que l'échantillon tiré d'une strate particulière n'est pas proportionnel à la taille de cette strate, mais à une autre de ses caractéristiques. Dans un échantillonnage en grappes, des groupes de personnes sont sélectionnés plutôt que des personnes. Si aucun cadre d'échantillonnage n'est disponible, il est courant de sélectionner un échantillon de grappes pour le dénombrement et l'identification des membres des populations cible. Vous trouverez ci-dessous une description détaillée de la manière dont l'échantillonnage a été effectué.

Étape **1** : Stratification

L'une des manières de procéder pour le dénombrement aurait été de sélectionner aléatoirement un certain nombre de municipalités du Salvador. Cependant, cette approche n'aurait probablement pas été en mesure de fournir des informations suffisantes pour l'estimation du nombre des ménages déplacés du pays et la taille de l'échantillon aurait été trop petite pour répondre aux objectifs cités ci-dessus, à savoir mieux comprendre les différents besoins et situations des ménages déplacés. Pour compenser, des strates ont été introduites afin de sur-échantillonner les zones qui abritaient, d'après des sources concordantes, une plus grande proportion de ménages déplacés.

Au départ, les partenaires du profilage ont rassemblé des informations sur la répartition supposée des ménages déplacés dans le pays afin d'éclairer la stratification. Grâce au passage en revue des sources existantes, il a été déterminé que 20 municipalités salvadoriennes abritaient un plus grand nombre de ménages déplacés que les autres. Les sources examinées incluaient des statistiques démographiques et migratoires, ainsi que des statistiques sur les faits de violence par municipalité.

Une liste de secteurs de recensement⁷ pour chacune des municipalités du pays a été mise à la disposition des partenaires du profilage par la Direction générale des statistiques et des recensements du Salvador (El Salvadoran General Directorate of Statistics and Censuses) et permis une subdivision plus poussée des municipalités. Les secteurs de recensement ont servi de grappes et d'unités d'échantillonnage primaires. Les entretiens et les ateliers participatifs organisés avec des informateurs clés des 20 municipalités prioritaires ont permis d'établir une liste de 648 des 4 427 secteurs de recensement supposés présenter une plus forte concentration de ménages déplacés.

Sur la base de ces informations, les secteurs de recensement ont été divisés en trois strates :

- (1) 648 secteurs priorités au sein des 20 municipalités établies comme prioritaires
- (2) 3 779 secteurs de recensement au sein des 20 municipalités prioritaires
- (3) 7 996 secteurs restants dans le pays

⁷ Zones géographiques plus petites que les municipalités fournies par la Direction générale des statistiques et des recensements du Salvador (El Salvadoran General Directorate of Statistics and Censuses).

Étape 2 : Décider de la taille de l'échantillon

Au sein de chacune des trois strates, une sélection aléatoire des segments de recensement a été réalisée. La taille de l'échantillon des secteurs de recensement a été déterminée sur la base de critères techniques de précision (95% de confiance) pour les estimations d'incidence faible⁸, avec un taux de non-réponse supposé élevé⁹, et en intégrant l'effet de plan¹⁰ conséquent causé par la stratification. La taille de l'échantillon a été fixée à 542 secteurs de recensement, ce qui correspond à un objectif estimé de 40 650 ménages.

Étape 3 : Sélectionner un échantillon en grappes (secteurs de recensement)

L'échantillon a été réparti de manière non proportionnelle entre les strates afin de garantir un nombre minimum de membres de la population cible. Une probabilité d'être sélectionné plus élevée a ainsi été attribuée à l'échantillon des secteurs de recensement des deux strates prioritaires comparé aux secteurs de recensement des autres strates. Les secteurs ont ensuite été tirés aléatoirement pour chacune des strates. Le tableau ci-dessous montre la répartition des secteurs de recensement échantillonnés dans les trois strates.

Strate	Nb de secteurs au Salvador	Répartition des secteurs au Salvador	Nb de secteurs dans l'échantillon	Répartition des secteurs dans l'échantillon
1	648	5%	184	34%
2	3779	30%	125	23%
3	7996	64%	233	43%
Total	12423	100%	542	100%

En raison de la répartition non proportionnelle des secteurs de recensement dans les strates, des pondérations ont dû être appliquées lors de l'analyse des résultats à l'échelle nationale, de façon à compenser la représentation inégale des secteurs dans l'échantillon. Ces pondérations sont appelées pondérations d'échantillonnage ou pondérations de probabilité inverse et sont généralement calculées en déterminant le nombre de ménages que chacun des ménages inclus dans l'échantillon représente (pondération d'échantillonnage = nombre total des ménages des strates/nombre total des ménages dans l'échantillon des strates). Pour obtenir le nombre de ménages au sein de chaque strate, le nombre de personnes de chaque strate peut être divisé par la taille moyenne des ménages. Dans ce profilage, les partenaires n'ont pas gardé la trace du calcul des pondérations, mais on peut supposer qu'ils se sont appuyés sur cette méthode très répandue pour calculer les pondérations d'échantillonnage.

⁸ Dans une étude similaire au Honduras, la proportion moyenne des ménages déplacés était de 4%.

⁹ Le taux de non-réponse de l'étude menée au Honduras s'est élevé à environ 50%.

¹⁰ Implique une plus grande incertitude des résultats en raison de la complexité de la méthode d'échantillonnage.

Étape 4 : Dénombrement des grappes sélectionnées (secteurs de recensement)

À la suite de la sélection d'un échantillon de 542 secteurs de recensement, un dénombrement de tous les ménages a été effectué dans 501 des 542 secteurs de recensement échantillonnés. Les enquêteurs n'ont pas pu accéder aux 41 secteurs de recensement restants en raison de restrictions liées à la sécurité. Les partenaires du profilage ont également dû faire face à un certain nombre de non-réponses lors du dénombrement mais dans une mesure plus faible que prévu (37 %). Le dénombrement a permis d'identifier les ménages comme déplacés ou non déplacés. Sur la base du nombre de ménages déplacés dans chacun des secteurs, les chiffres ont pu être extrapolés pour la globalité du pays¹¹, afin de remplir l'objectif (1). Les résultats du dénombrement ont également permis de constituer un échantillon des ménages déplacés ainsi qu'un cadre d'échantillonnage pour la sélection des ménages non déplacés en vue de l'enquête d'échantillonnage planifiée en réponse aux deux autres objectifs.

Étape 5 : Échantillon résultant des ménages PDI

Le dénombrement a permis d'identifier 466 ménages déplacés en raison de la violence, soit 1 811 personnes. Ces ménages ont tous été inclus dans l'échantillon et invités à participer à l'enquête-ménage. 431 de ces ménages, soit l'équivalent de 1 668 personnes, ont accepté et répondu à l'enquête.

Étape 6 : Sélectionner un échantillon de ménages non déplacés

Un échantillon pour la population cible de comparaison a également été tiré au sein des secteurs de recensement échantillonnés par le biais d'un échantillonnage aléatoire simple. À chaque fois que trois enquêtes étaient administrées à un ménage déplacé, une enquête était administrée à un ménage non déplacé sélectionné aléatoirement dans le même secteur de recensement et sur la base des listes du dénombrement. Comme la sélection de l'échantillon des ménages non déplacés a été effectuée en deux phases (c.à.d. en tirant d'abord un échantillon de segments de recensement puis un échantillon de ménages), cette approche peut être appelée méthode d'échantillonnage « à deux phases ».

¹¹ Pour des explications sur la manière dont l'extrapolation a été effectuée, voir l'annexe 1 du rapport de profilage (uniquement en espagnol): <https://www.jips.org/jips-publication/profiling-report-el-salvador-2018/>

Difficultés et limites

Le profilage au Salvador illustre un scénario qui prévaut souvent dans les contextes de déplacement : l'absence d'une liste des populations concernées pouvant servir de cadre d'échantillonnage. Cette absence a un impact sur la conception de la méthode d'échantillonnage qui va souvent devoir être plus complexe, ce qui va ensuite accroître l'incertitude autour des résultats. Cela a été le cas au Salvador où les partenaires ont choisi d'utiliser à la fois un échantillonnage stratifié en grappes et un échantillonnage à deux phases.

Un autre problème couramment rencontré lors des enquêtes d'échantillonnage est la **non-réponse**, c.à.d. ne pas réussir à entrer en contact avec les personnes de l'échantillon ou se heurter à leur refus de répondre au questionnaire. Comme nous l'avons vu plus haut, ce processus de profilage a été confronté à un taux de non-réponse important lors du dénombrement, bien que ce taux ait été moins important que prévu. En outre, les partenaires chargés du profilage ont été confrontés à des contraintes liées à la sécurité qui les ont empêchés de dénombrer tous les secteurs de recensement de l'échantillon. Ces deux éléments sont susceptibles d'avoir engendré des biais dans les résultats.

Enfin, la **répartition entre les sexes** n'a pas été prise en compte lors de l'administration de l'enquête aux ménages déplacés et non déplacés échantillonnés. Cela peut également avoir introduit un biais dans les résultats.

Exercice de profilage à Thessalonique, Grèce, 2017-2019



Objectifs

Cet exercice de profilage avait pour objectif de fournir à la municipalité de Thessalonique et aux autres acteurs concernés une base de données factuelles pouvant servir à la conception d'initiatives de sensibilisation, de politiques, d'offres de services et de programmes d'intégration mieux adaptés aux besoins des personnes touchées par le déplacement dans la zone métropolitaine de ville.

Les objectifs spécifiques convenus par le profilage étaient les suivants :

- (1) Établir un **profil démographique** des réfugiés, des demandeurs d'asile et des arrivées spontanées désagrégé par sexe, âge et autres critères de diversité pertinents ;
- (2) **Analyser les moyens et capacités, vulnérabilités et mécanismes d'adaptation** des populations cible ;
- (3) Évaluer le **degré d'intégration** des populations cible, particulièrement en ce qui concerne l'accès aux services ;
- (4) Produire un ensemble d'indicateurs permettant de mesurer leur degré d'intégration.

Population cible

L'exercice de profilage avait deux populations cible principales :

- (1) **Les réfugié.e.s et demandeuse.eur.s d'asile arrivés à Thessalonique après janvier 2015** et hébergés dans des logements fournis par le programme d'hébergement ESTIA¹² dans les structures ouvertes (Open Reception Facilities, ORF) de Diavata¹³, ou qui sont hébergés par leurs propres moyens. Cette population cible comprenait les personnes disposant d'une carte de pré-enregistrement et d'enregistrement en tant que demandeuse.eur d'asile, d'une décision d'octroi d'asile, d'un permis de séjour ou d'une demande d'asile faisant l'objet d'un recours administratif.

¹² Le dispositif Emergency Support to Integration & Accommodation (ESTIA) fournit des solutions d'hébergement aux réfugiés et demandeuse.eur.s d'asile par le biais d'appartements en zone urbaine. Il est géré par le HCR en coordination avec 12 organisations partenaires dans le nord et le centre de la Grèce.

¹³ Lieux d'hébergement provisoires créés pour loger les réfugié.e.s et les demandeuse.eur.s d'asile bloqués du côté grec de la frontière entre la Grèce et la Macédoine du Nord après la fermeture de la frontière au début de 2016. Ces sites n'étaient pas adaptés à un hébergement de longue durée.

(2) Ressortissant.e.s de pays tiers non enregistrés auprès des services d'asile grecs. Il s'agissait notamment de personnes figurant sur des registres de police, de personnes dont les documents délivrés par l'État grec avaient expiré ou qui n'avait pas reçu de documents de la part de l'État. L'étude a intégré les personnes qui résidaient officieusement à l'ORF de Diavata, les personnes hébergées par des particuliers à Thessalonique ainsi que les personnes sans-abri.

Plus précisément, le périmètre géographique du profilage comportait 6 arrondissements de Thessalonique et les municipalités environnantes de Kalamaria, Neapolis-Sikeon, Pavlou Mela, Kordeliou-Evosmos, Ampelokipon-Menemenis, Delta, Thermis and Oreokastro. Les résident.e.s du centre d'accueil ouvert de Diavata ont également été inclus.

Les réfugié.e.s et les demandeuse.eur.s d'asile sans adresse de résidence valide n'ont pas été inclus dans la population cible de l'étude, en raison du manque d'informations disponibles sur ce groupe. En outre, les réfugié.e.s et les demandeuse.eur.s d'asile de nationalité turque ont été exclus de l'enquête parce que leurs adresses exactes n'étaient pas accessibles. Des groupes de discussion ont cependant été organisés avec la population turque.

Cadre d'échantillonnage

Réfugié.e.s et demandeuse.eur.s d'asile : un cadre d'échantillonnage a été construit en combinant deux sources différentes – la base de données ProGres du HCR (une liste complète des personnes hébergées via le dispositif d'hébergement urbain ESTIA et les personnes logées par leurs propres moyens, incluant leurs informations de contact) et une liste de tous les résidents du site de l'ORF de Diavata.

Ressortissant.e.s de pays tiers non enregistrés auprès des services d'asile : étant donné l'absence de liste ou registre complets pour ces personnes, les partenaires ont dû rechercher d'autres moyens de concevoir une approche d'échantillonnage adéquate. Différents fournisseurs de service disposaient d'informations sur cette population cible mais, pour des raisons liées à la protection, les partenaires du profilage n'ont pas pu y avoir accès. Cependant, en comparant les informations agrégées de plusieurs fournisseurs de services, on a estimé qu'environ 200 ménages appartenaient à cette population cible.

Unités d'échantillonnage et d'analyse



Méthode d'échantillonnage

Les objectifs du profilage – mieux comprendre les caractéristiques des populations cible et leurs situations – impliquaient que les résultats de l'enquête de profilage devaient idéalement s'appliquer aux deux populations cible. Cela signifiait qu'une sélection probabiliste de l'échantillon aurait été l'approche à privilégier pour garantir un échantillon représentatif avec un niveau de confiance et une marge d'erreur connus. Mais, comme indiqué ci-dessus, cela n'a pas été possible puisque que les partenaires du profilage ne disposaient d'un cadre d'échantillonnage que pour l'une des deux populations cible.

L'existence d'une liste complète de la population cible des réfugié.e.s et demandeuse.eur.s d'asile signifiait qu'un tirage probabiliste de l'échantillon était en théorie possible, tandis que l'absence d'un registre pour les personnes ne disposant d'aucun document des services d'asile impliquait l'utilisation d'une stratégie d'échantillonnage non probabiliste. Pour le cas présent, celle-ci a pris la forme d'un échantillonnage boule de neige.

Si cette configuration n'était pas idéale et signifiait que les résultats des enquêtes d'échantillonnage ne pourraient pas être directement comparés, les partenaires du profilage n'avaient accès à aucune alternative réalisable. Bien que le degré de représentativité n'ait pas pu être évalué et les résultats généralisés à l'ensemble de la population cible¹⁴, cette approche a tout de même permis d'aboutir à la production de données pertinentes pour un groupe de population sur lequel très peu d'informations étaient disponibles.

Les deux sections ci-dessous proposent une description plus détaillée des deux approches d'échantillonnage choisies pour les deux populations cible.

Réfugié.e.s et demandeuse.eur.s d'asile

La méthodologie choisie initialement pour cette population cible consistait en un échantillonnage aléatoire stratifié. Néanmoins, lors de la mise en œuvre de l'exercice, cette approche a dû être modifiée pour des raisons que nous détaillerons ci-dessous. L'approche prévue initialement peut être résumée de la manière suivante :



Étape 1 : Stratification

La population cible a été divisée en trois strates en fonction de leurs différents types d'hébergement. Ces trois strates avaient été choisies parce qu'on s'attendait à ce que les situations des personnes varient significativement en fonction de leur type d'hébergement. Il était donc souhaitable de pouvoir comparer la situation des personnes selon ce critère. Un échantillon suffisamment important a ainsi été tiré pour chacun des trois types d'hébergement afin de pouvoir compiler des résultats représentatifs pour chaque groupe.

La population cible a été groupée au sein des strates suivantes :

- (1) Les réfugié.e.s et demandeuse.eur.s d'asile à qui le dispositif d'hébergement urbain ESTIA avait fourni des appartements (1,280 ménages).
- (2) Les réfugié.e.s et demandeuse.eur.s d'asile logés par leurs propres moyens à Thessalonique, c.à.d. qui louaient un appartement par eux-mêmes, ou étaient hébergés par des amis, de la famille ou des bénévoles (482 ménages).
- (3) Les réfugié.e.s et demandeuse.eur.s d'asile officiellement enregistrés comme résidents de la structure d'accueil ouverte (Open Reception Facilities, ORF) de Diavata (148 ménages).

¹⁴ À moins que le point de saturation ne soit atteint, c'est-à-dire le moment où toutes les personnes de la population cible ont été identifiées et où plus aucune désignation n'est faite.

Étape 2 : Décider de la taille de l'échantillon

Pour la strate (3), il a été décidé d'effectuer un comptage complet de tous les ménages, en raison du relativement petit nombre des ménages résidant dans le camp. Pour les strates (1) et (2), la taille des échantillons a été calculée sur la base d'un niveau de confiance de 95 % et une marge d'erreur souhaitée de 5 %. En outre, avec un taux de réponse attendu de 20%, la taille de l'échantillon a été ajustée pour chaque strate afin de refléter la petite taille de la population cible. En conséquence, la taille de l'échantillon de la strate (1) était de 230 ménages et celui de la strate (2) de 190 ménages.

Étape 3 : Sélectionner l'échantillon

Comme indiqué ci-dessus, toutes les personnes de la strate (3) ont été sélectionnées pour être incluses dans l'échantillon. Les questionnaires ont été distribués à tous les ménages de la structure et personne n'a refusé de répondre à l'enquête, ce qui a donné lieu à un taux de réponse de 100 % pour cette strate. Il convient toutefois de noter que la population de cette strate était relativement peu nombreuse par rapport à celles des autres strates et que, pour certaines variables, le nombre d'observations était très faible, ce qui a rendu l'analyse de ces résultats plus incertaine.

Après la détermination de la taille des échantillons des strates (1) et (2), un échantillon aléatoire a été tiré dans chacune de ces deux strates. La probabilité d'être inclus dans l'enquête différait selon les strates : dans la strate (1) 18 % de la population cible était incluse dans l'échantillon, dans la strate (2) 39 % et dans la strate (3) 100 %. Pour estimer une moyenne de population pour l'ensemble de l'échantillon, cette représentation inégale a dû être corrigée par l'utilisation de pondérations. Ces pondérations, appelées également pondérations de probabilité inverse, sont égales au nombre de ménages que chaque ménage de l'échantillon représente. Des pondérations ont uniquement été utilisées lors du calcul des chiffres pour l'ensemble des trois strates (par ex. pour l'estimation du revenu moyen de l'ensemble de la population cible) et aucune pondération n'a été appliquée lorsqu'il s'agissait d'estimer une moyenne d'une seule des trois strates (par ex., le revenu moyen de la population dans l'ORF de Diavata).

Lors du contact des personnes des échantillons des strates (1) et (2) par téléphone, il est rapidement apparu que les numéros de la majorité des personnes de l'échantillon n'étaient plus valables en raison de changements fréquents de carte SIM. Les ressources étant limitées, il n'était pas non plus possible de rendre visite aux ménages à l'improviste, ce qui signifiait que ces personnes ne pouvaient tout simplement pas être jointes. Le taux de non-réponse allait donc être trop important pour que l'échantillon soit représentatif. Les partenaires du profilage ont alors décidé de renoncer à un échantillonnage aléatoire pour ces deux strates et d'interviewer à la place toutes les personnes joignables de la strate, en espérant obtenir un taux de réponse acceptable.

Un recensement a finalement été réalisé dans les deux strates et associé à un échantillonnage boule de neige où les répondants orientaient les recenseuses vers d'autres ménages de la population cible. Le taux de réponse pour la strate (1) a été de 34,2 % (pour 1,280 ménages) et pour la strate (2) de 41,2 % (pour 482 ménages).

Étape 4 : Analyse de la représentativité et pondérations post-stratification

Les taux de réponse relativement élevés dans les strates (1) et (2) ont introduit un risque de biais dans les résultats, puisque les personnes ayant répondu au questionnaire étaient susceptibles de ne pas suffisamment refléter les caractéristiques de la population cible globale.

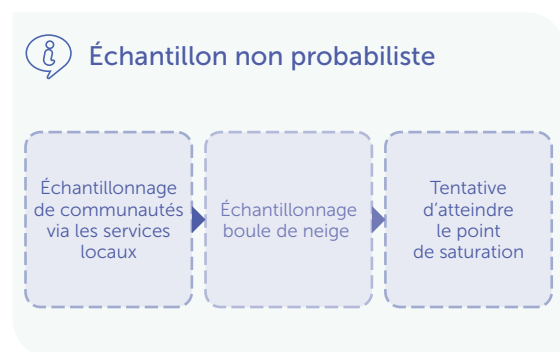
Une analyse des caractéristiques démographiques clés (âge, nationalité, et répartition par sexe) a été effectuée pour les répondant.e.s et comparée à celles de la population cible (accessible dans la base de données ProGres) pour évaluer la représentativité

des répondant.e.s. La répartition des caractéristiques des répondant.e.s s'est avérée finalement très proche de celle de la population cible. Sur cette base, il a été conclu que les résultats de l'enquête pouvaient être considérés comme représentatifs.

La réalisation d'un recensement a signifié qu'il n'a pas été nécessaire d'appliquer des pondérations à ce stade. En revanche, pour tenir compte des taux différents de réponse dans les trois strates, des pondérations post-stratification ont été calculées et appliquées à toutes les tables où les réponses des répondant.e.s étaient considérées ensemble, c'est-à-dire lorsque la variable concernée n'avait pas été désagrégée par strate.

Personnes sans documents des services d'asile

Une approche d'échantillonnage non probabiliste a été utilisée pour cette population cible sous la forme d'un échantillonnage boule de neige. Un échantillonnage boule de neige implique de sélectionner quelques personnes à interviewer dans la population cible et de leur demander d'identifier d'autres membres de la population cible pour les enquêtes. Les identifications des nouveaux répondants potentiels par les répondant.e.s interviewés se font par « vagues » jusqu'à ce que les enquêtes aient obtenu le nombre de répondant.e.s souhaité pour l'échantillon. L'approche adoptée peut être résumée de la manière suivante :



Étape 1 : Identifier les répondants

Des fournisseurs de services locaux qui apportaient une assistance aux personnes sans documents d'asile ont été engagés pour contacter des répondant.e.s potentiels et fixer des rendez-vous pour les interviews. En outre, il était demandé aux personnes contactées de désigner d'autres personnes de la population cible. La population cible a été estimée à approximativement 200 personnes. L'objectif était d'atteindre un nombre équivalent de ménages dans l'espoir d'atteindre le point de saturation, c'est-à-dire le moment où un décompte complet de la population a été obtenu.

Étape 2 : Conduire les interviews

Le nombre de ménages interrogés a été plus élevé que prévu. L'équipe de recensement a interrogé 451 personnes, représentant 227 ménages. L'échantillon recouvrait des situations d'hébergement diverses : 222 personnes étaient sans-abri (dans la rue, dans des bâtiments inachevés ou abandonnés, ou dans des refuges pour sans-abris), 161 personnes vivaient dans l'ORF de Diavata sans être enregistrés en tant que résident.e.s et 27 vivaient dans des logements loués à Thessalonique.

Ce nombre élevé de répondant.e.s peut, entre autres, s'expliquer par un afflux d'arrivées à Thessalonique au courant du mois où les entretiens ont été menés. Les résultats de l'enquête allaient dans le sens de cette théorie, puisque plus de la moitié des répondant.e.s se trouvaient à Thessalonique depuis moins d'un mois au moment de l'entretien. Ce grand nombre d'arrivées récentes a rendu l'estimation de la population cible totale plus incertaine. En outre, un grand nombre de personnes contactées ont refusé d'être interviewées. Par conséquent, il n'a pas été possible d'évaluer la représentativité des personnes interviewées pour la population cible.

Difficultés et limites

Les partenaires chargés du profilage ont dû faire face à plusieurs difficultés qui ont limité la possibilité d'une enquête au moyen de l'approche d'échantillonnage privilégiée, tant au stade de la planification initiale que lors de la mise en œuvre de l'exercice.

En premier lieu, les difficultés rencontrées lors de la prise de contact avec les membres de l'échantillon des demandeuse.eur.s d'asile et des réfugié.e.s n'ont pas permis de baser l'enquête sur un échantillon sélectionné aléatoirement. Le **taux élevé de non-réponse** dans la population cible finalement incluse a conduit à une incertitude sur les résultats. L'analyse de la représentativité de l'échantillon a toutefois montré que les caractéristiques des personnes interrogées semblaient être représentatives des populations cible, et il a donc été supposé que les résultats pouvaient être extrapolés à l'ensemble de la population cible. Néanmoins, il faut garder à l'esprit qu'il est possible que certaines autres caractéristiques par lesquelles les répondants se distinguent de la population cible n'aient pas été identifiées et évaluées, ce qui créerait une incertitude plus grande quant à la représentativité des résultats.

Deuxièmement, **l'absence d'un cadre d'échantillonnage** a conduit à un échantillon non probabiliste pour les ressortissants des pays tiers non enregistrés auprès des services d'asile. Pendant la phase de collecte des données, le refus de la part d'un grand nombre de membres de la population cible de participer à l'enquête a encore davantage compliqué l'exercice. Dans l'ensemble, cela a abouti à ce que les résultats soient incertains et au fait qu'il était impossible d'affirmer que ceux-ci étaient représentatifs de l'ensemble de la population cible.

Exercice de profilage à Erbil, Kurdistan irakien, 2015-2016



Objectifs

La province d'Erbil, qui accueille 2,01 millions de personnes (chiffres de 2014), abrite la capitale du Kurdistan irakien. Comme le reste du Kurdistan, elle était, au moment de l'exercice de profilage, fortement touchée par des vagues de déplacement causées par les conflits en Syrie et dans le reste de l'Irak, ainsi que par une crise financière omniprésente qui affectait les secteurs public et privé de l'économie.

Si beaucoup d'informations sur les PDI et réfugié.e.s vivant dans les camps étaient disponibles, peu d'éléments existaient sur ceux résidant hors des camps. En outre, la plupart des stratégies visant à atténuer les effets du déplacement s'étaient jusqu'alors consacrées aux besoins des populations déplacées et réfugiées, sans que les communautés d'accueil bénéficient de la même attention.

Les objectifs spécifiques de l'étude étaient les suivants :

- (1) Fournir des **profils démographiques** désagrégés par sexe, âge et statut de déplacement (c'est-à-dire réfugié.e.s, personnes déplacées et communautés d'accueil) dans les zones ciblées ;
- (2) Fournir des profils des zones urbaines comportant une forte concentration de populations déplacées hors des camps ;
- (3) **Analyser les capacités et moyens, vulnérabilités et mécanismes d'adaptation** de la population de ces zones ;
- (4) **Analyser les relations** entre les populations déplacées et les populations qui les accueillent ;
- (5) Analyser la résilience des zones urbaines par rapport à la disponibilité ou aux limitations des services ;
- (6) Produire un ensemble de données à la disposition du gouvernement régional du Kurdistan et de la communauté humanitaire/du développement.

Les objectifs (1), (3) et (4) ont été atteints par le biais d'une enquête-ménage par échantillonnage.

Population cible

Trois populations cible ont été définies :

- (1) Réfugié.e.s syriens;
- (2) Population des PDI déplacées depuis décembre 2013 ;
- (3) Communauté accueillante (composée des personnes qui n'entrent dans aucune des catégories ci-dessus, y compris les migrants économiques).

Tous résidaient dans les zones géographiques suivantes :

- Partie centrale du district d'Erbil, composé de la ville de Hawler (Nawandy Hawler) et de Ainkawa;
- District périphérique d'Erbil, composé du noyau urbain immédiat entourant le centre du district et relié directement à lui : Baharka, Bnaslaw (Dashty Hawler Centre), Daratu, Kasnazan, Khabat Centre (Nawandy Khabat), et Rizgari ;
- Villes qui constituaient des centres urbains à l'intérieur de la province : Harir, Koya Centre, Shaqlawa Centre et, conjointement, Soran Centre (Nawandy Soran) et Diana.

Cadre d'échantillonnage

Un cadre d'échantillonnage distinct a été établi pour chacune des populations cible sur la base de différentes sources de données secondaires :

- (4) Réfugié.e.s syriens : données de la base de données ProGres du HCR pour l'enregistrement des réfugié.e.s;
- (5) Population des PDI déplacées depuis décembre 2013 : données en provenance de la première phase de l'Enregistrement complet des personnes déplacées (Comprehensive Registration of Displaced People CRDP,) conduit par le Bureau régional des statistiques du Kurdistan (Kurdistan Regional Statistics Office KRSO,) en juin 2015 ;
- (6) Communauté accueillante : données datant de 2009, collectée dans le cadre de la préparation d'un recensement qui était planifié mais n'a jamais été réalisé.

Unités d'échantillonnage et d'analyse



Méthode d'échantillonnage

L'exercice avait pour objectifs de décrire les caractéristiques et situations des différentes populations cible et d'analyser les relations entre ces groupes. Ces objectifs nécessitaient de sélectionner des échantillons représentatifs de ces populations. Cela a abouti à privilégier le recours à une approche d'échantillonnage probabiliste. Les partenaires du profilage souhaitaient également que les différents résultats soient comparables entre les différentes zones géographiques et les différentes populations cible de chaque zone. Cela impliquait que les partenaires en charge de l'établissement des profils devaient concevoir une approche permettant d'obtenir des échantillons suffisamment importants pour chacune des populations cible des trois zones. Un cadre d'échantillonnage spécifique a été conçu pour chacune des populations cible. Le KRSO a conçu la méthodologie d'échantillonnage (en partie en collaboration avec JIPS) et procédé à la sélection des échantillons et à la collecte des données. Une brève description des approches adoptées pour chaque population cible est présentée ci-dessous.

Réfugié.e.s syriens

Un échantillon aléatoire simple a été tiré à partir du cadre d'échantillonnage. Celui-ci était relativement dépassée et/ou déficient, sans possibilité d'être mis à jour. Des entretiens ont néanmoins pu être réalisés avec les ménages de l'échantillon. Aucune information sur le taux de réponse n'est disponible. Il convient de noter qu'un taux de non-réponse élevé serait susceptible d'introduire un biais dans les résultats et donc de réduire la représentativité de l'échantillon simple.

PDI

Une approche stratifiée à deux phases a été utilisée pour le tirage de l'échantillon de la population des PDI.

Phase 1 : Les districts présentant la plus forte concentration de personnes déplacées ont été sélectionnés dans les trois strates. Les informations sur la concentration de population déplacée de ces districts étaient à la disposition du KRSO grâce à la première phase de l'Enregistrement complet des personnes déplacées (Comprehensive Registration of Displaced People). La répartition de l'échantillon entre les districts devait être proportionnelle à la taille de la population déplacée de chaque district. Dans chacun des districts, un certain nombre de quartiers ont été choisis et dénombrés. JIPS ne sait pas si les quartiers ont été choisis aléatoirement ou s'ils ont été sélectionnés parce qu'ils présentaient la plus forte concentration de personnes déplacées.

Phase 2 : Un échantillon des ménages déplacés a été tiré pour chaque quartier, sur la base des informations du dénombrement.

Communauté accueillante

Comme pour les PDI, une approche stratifiée à deux phases a été utilisée pour tirer l'échantillon des communautés accueillantes mais l'unité primaire d'échantillonnage de la première phase a été constituée des blocs d'habitations et non des districts. JIPS ne dispose pas d'informations sur les raisons ayant motivé le choix d'une unité d'échantillonnage différente et sur la manière dont les blocs ont été sélectionnés (aléatoirement ou non).

Difficultés et limites

La méthodologie d'échantillonnage est globalement insuffisamment documentée dans le rapport de profilage. Cela limite l'utilité des résultats de l'enquête pour toute autre personne ou organisation que les partenaires du profilage eux-mêmes.

L'échantillon des réfugiés syriens a été tiré à partir d'un cadre d'échantillonnage obsolète et il est donc difficile d'estimer dans quelle mesure les informations obtenues sur les ménages échantillonnés sont réellement représentatives des groupes de population de chaque strate.

En outre, parce que les districts ont été choisis sur la base de la concentration de la population des PDI au sein de chaque district, il n'a pas été possible de saisir quoi que ce soit au sujet des PDI et communautés d'accueil qui vivaient dans des districts à plus faible concentration. Il se peut que ces ménages aient été confrontés à des situations très différentes de celles des ménages des districts à forte concentration. De plus, s'il s'avérait que les ménages déplacés constituent une part non négligeable du total de la population des PDI de chaque strate, l'échantillon sélectionné ne serait alors plus représentatif des PDI et des communautés d'accueil de ces districts, mais uniquement des districts à forte concentration.

Exercice de profilage au Kosovo, 2015-2016



Objectifs

La population qui a été poussée au déplacement par le conflit de 1998/99 et des émeutes de 2004 se trouve depuis dans une situation de déplacement prolongé. Les estimations effectuées au moment du profilage ont montré que la majorité des personnes déplacées à l'intérieur du Kosovo résidait dans des logements privés et une petite minorité dans des centres collectifs. L'exercice de profilage avait pour but de contribuer à la mise en œuvre d'une politique pour les solutions durables et notamment à la stratégie en matière de solutions durables du Ministère des communautés et des retours (Ministry for Communities and Return), en fournissant une analyse factuelle de la situation de déplacement, conformément au Cadre de l'IASC sur les solutions durables pour les personnes déplacées à l'intérieur de leur propre pays.

Ses objectifs spécifiques étaient les suivants :

- (1) Établir un **profil démographique** de la population déplacée au Kosovo, désagrégé par âge, sexe, localisation et critères de diversité ;
- (2) Procéder à une analyse **exhaustive de la situation de déplacement** concernant: l'intégration sociale et économique des personnes déplacées, les difficultés et vulnérabilités liées au déplacement, les ressources et moyens, la jouissance des droits et enfin les intentions et projets des PDI pour l'avenir.
- (3) Renforcer la capacité **des institutions à concevoir et à plaider en faveur d'une programmation commune** en soutien de solutions durables pour les PDI par l'identification des priorités spécifiques de ces groupes et une meilleure coordination de l'analyse humanitaire et du développement.

Population cible

L'étude a porté sur quatre principales populations cible des localités mentionnées ci-dessous. Toutes ont été contraintes de fuir leur lieu de résidence permanent entre janvier 1998 et fin mars 2004 en raison du conflit armé, de la situation de violence généralisée et des violations des droits de l'homme ou de leurs conséquences, mais sont restées au Kosovo. Le périmètre géographique des populations cible recouvrait les municipalités avec la plus forte concentration de personnes déplacées. Les personnes d'ethnicité autre n'ont pas été incluses dans l'étude.

- (1) **PDI albanaises** à Prishtinë/Pristina, au sud de Mitrovica et Vushtrri/Vucitrn,
- (2) **PDI serbes dans des logements privés** à Zveqan/Zvecan, Leposaviq/Leposavić, nord de Mitrovica, Zubin Potok, Graçanicë/Graçanica, Fushë Kosovë/Kosovo Polje, Lipjan/Lipljane, Shtërpçë/Strpce, Gjiilan/Gnjilane, Kamënicë/Kamenica, Viti/Vitina,

- (3) **PDI serbes en centres collectifs** (tous les centres collectifs du Kosovo)
- (4) **PDI Roms/Ashkalies/Égyptiennes (RAE)** à Fush K./K. Polje, Gracanicë/Gracanica, Ferizaj/Urosevac, Pejë/Peć, Gjakovë/Dakova et dans d'autres municipalités où des PDI RAE ont pu être localisées.

Un groupe de comparaison a également été défini pour comparer les résultats sur les populations déplacées à la situation des populations non déplacées au Kosovo. Il a été décidé que le groupe de comparaison serait la population générale du Kosovo, et que des données secondaires seraient utilisées pour rendre compte de cette population cible.

Cadre d'échantillonnage

PDI albanaises : une liste de 1 167 ménages et 5 879 individus a été établie. La source a été le recensement effectué en 2011 par l'Agence kosovare de la statistique (Kosovo Agency of Statistics, KAS). Un ensemble de questions figurant dans le recensement a été identifié comme pouvant servir à déterminer si la personne appartenait à la population albanaise déplacée et, sur cette base, une extraction des PDI albanaises a pu être effectuée. Pour vérifier cette liste, des sources additionnelles ont été recoupées et confrontées à l'extraction : une liste des PDI albanaises qui avaient déposé des demandes de compensations – principalement en lien avec les déplacements causés par le conflit – pour la perte de leurs biens auprès de l'Agence kosovare de la propriété (Kosovo Property Agency), une liste des PDI albanaises qui s'étaient adressées directement au DRC et au Bureau administratif de Mitrovica Nord (MNAO, Mitrovica North Administrative Office) à Mitrovicë, ainsi que les leaders de la communautés albanaises qui avaient établi des contacts avec le HCR et le DRC.

PDI serbes dans des logements privés : une liste complète des 3 872 ménages et 16 383 individus était disponible. La source était la base de données du Commissariat serbe aux réfugiés et aux migrations (SCRM, Serbian Commissariat for Refugees and Migrations), qui regroupait des données actualisées sur les noms des ménages déplacés, leurs tailles, adresses et informations de contact. Le cadre d'échantillonnage fourni par le SCRM a été recoupée avec des informations en provenance du HCR.

PDI serbes résidant en centres collectifs : une liste complète des 140 ménages était disponible. Les sources étaient la base de données du Commissariat serbe aux réfugiés et aux migrations (SCRM) et la base de données consolidée du HCR.

PDI Roms/Ashkalies/ Égyptiennes (RAE) : une liste de 137 ménages et de 638 individus a été élaborée en combinant une base de données du HCR et les listes d'autres organisations apportant de l'aide à cette population. La base de données du HCR n'était pas récente et n'avait pas été actualisée. Or, on s'attendait à ce que ces ménages soient particulièrement mobiles du fait de leur vulnérabilité économique, ce qui risquait d'amoindrir la précision du cadre d'échantillonnage.

Unités d'échantillonnage et d'analyse



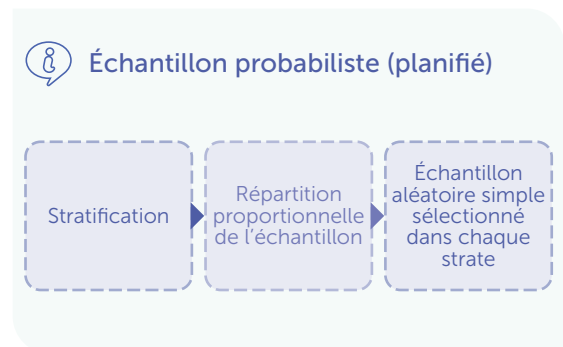
Méthode d'échantillonnage

Les objectifs de l'exercice étaient de générer une meilleure compréhension des caractéristiques et situations des populations cible, ce qui nécessitait que les résultats de l'enquête par échantillonnage soient représentatifs de chacune des populations cible. À cette fin, une approche d'échantillonnage probabiliste aurait été l'option privilégiée. Cependant, la disponibilité limitée de cadres d'échantillonnage actualisés a finalement nécessité le recours à des approches non probabilistes pour certaines des populations cible.

Les méthodes d'échantillonnage adoptées pour les quatre populations cible ont ainsi été élaborées sur mesure et conçues pour que les différents résultats soient comparables et non pour donner une image de la population dans son ensemble. Vous trouverez ci-dessous une description des différentes approches adoptées pour chacune des populations cible.

PDI albanaises

Initialement, un échantillonnage aléatoire stratifié avait été planifié pour cette population cible. Toutefois, l'exercice avançant, les partenaires du profilage ont dû renoncer à cette approche en raison des difficultés exposées ci-dessous. En dernier recours, ils ont fait appel à un échantillonnage boule de neige. L'approche d'échantillonnage prévue initialement peut être résumée comme suit :



Étape 1 : Stratification

La population cible a été divisée en trois strates basées sur des zones administratives :

- (1) Municipalité de Prishtinë/Pristina
- (2) Municipalité du sud de Mitrovica
- (3) Municipalité de Vushtrri/Vucitrn

Au sein des trois strates ci-dessus, la population cible a été divisée en deux strates supplémentaires :

- (1) Zones urbaines
- (2) Zones rurales

Étape 2 : Décider de la taille de l'échantillon

Un échantillon de 510 ménages a été calculé sur la base d'un niveau de confiance de 95 % et d'une marge d'erreur souhaités de 5 % avec un taux de non-réponse attendu de 15 %.

Étape 3 : Sélectionner l'échantillon

La taille de l'échantillon a été répartie proportionnellement entre toutes les strates en fonction de la taille de la population des PDI albanaises de chacune des strates, en s'appuyant sur le recensement de 2011 de l'Agence kosovare de la statistique (KAS, Kosovo Agency of Statistics). Un échantillon aléatoire a ensuite été tiré pour chacune des strates sur la base de l'extraction du recensement.

Après le lancement du travail de terrain et lors de la prise de contact avec les ménages de l'échantillon pour les interviews, il s'est avéré que les adresses indiquées n'abritaient quasiment aucune famille de PDI albanaises. Dans une petite partie des cas, les familles qui résidaient à ces adresses lors du recensement avaient depuis déménagé. Mais, pour la majorité des cas, la raison de leur absence à ces adresses n'a pas pu être élucidée.

Cela a abouti à l'abandon du plan d'échantillonnage initial et les partenaires du profilage ont décidé qu'ils n'avaient pas d'autre choix que de recourir à un échantillonnage boule de neige, sans pour autant renoncer à une taille d'échantillon équivalente et aux strates initialement prévues. Les ménages de l'échantillon de départ qui résidaient aux adresses indiquées ont servi de points de référence et invités à identifier d'autres ménages de PDI albanaises.

Les partenaires du profilage ont plus ou moins réussi à atteindre la taille d'échantillon globale souhaitée, puisque 502 ménages au total ont été interrogés. La répartition finale des ménages interrogés au sein des trois strates de municipalités ne correspondait cependant pas tout à fait à la répartition prévue (comme le montre le tableau ci-dessous). En raison de l'approche non probabiliste finalement adoptée, la représentativité de l'échantillon pour l'ensemble de la population cible est demeurée incertaine.

Municipalité	Répartition initiale des ménages	Répartition finale des ménages
Prishtinë/Pristina	114	45
Sud de Mitrovica	315	435
Vushtrri/Vucitrn	81	12
Autre	0	10
Total	510	502

PDI serbes dans des logements privés

Un échantillonnage aléatoire stratifié a été choisi pour la sélection de l'échantillon de cette population cible. L'approche adoptée peut être résumée de la manière suivante:

Étape 1 : Stratification

La population cible a été stratifiée en deux strates en fonction de leur localisation géographique :

- (1) Les municipalités du nord (qui ont accueilli la plus grande part de population serbe déplacée)
- (2) Les municipalités du sud (qui ont accueilli la plus petite part de population serbe déplacée)

Au sein des deux strates ci-dessus, la population cible a été divisée en deux strates supplémentaires :

- (1) Zones urbaines
- (2) Zones rurales

Échantillon probabiliste



Étape 2 : Décider de la taille de l'échantillon

Un échantillon de 617 ménages a été calculé sur la base d'un niveau de confiance de 95 % et d'une marge d'erreur souhaités de 5 % avec un taux de non-réponse attendu de 15 %.

Étape 3 : Sélectionner l'échantillon

La taille de l'échantillon a été répartie à parts égales entre les strates des municipalités du nord et du sud. Cette répartition a rendu nécessaire l'utilisation de pondérations lors de l'analyse des données afin d'intégrer les différences de taille entre les deux populations des strates. Ces pondérations étaient égales au nombre total de ménages représentés par chacun des ménages échantillonnés dans les deux strates. Les pondérations ont été appliquées à chaque fois que les résultats ont présentés globalement, c'est-à-dire quand les résultats des deux strates étaient agrégés. Lorsque les résultats des strates ont été présentés séparément, aucune pondération n'a été appliquée.

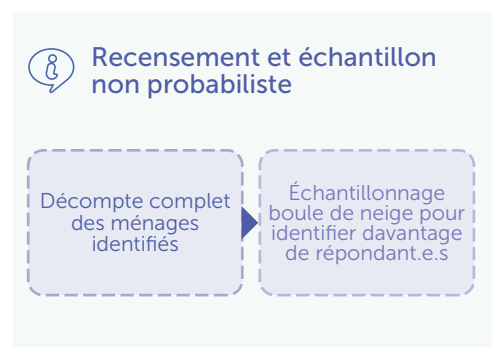
La taille de l'échantillon a été répartie proportionnellement entre les zones urbaines et rurales en fonction de la taille de la population serbe déplacée dans chacune des strates des municipalités du nord et du sud. L'échantillon a été tiré du cadre de l'échantillonnage dans toutes les strates et des interviews ont été menées pour les ménages de l'échantillon. Un total de 567 ménages a répondu à l'enquête, ce qui représente un taux de réponse élevé de 92 %.

PDI serbes en centres collectifs

La petite taille de cette population cible (140 ménages) a conduit à la décision d'inclure l'intégralité des ménages dans l'enquête, c'est-à-dire de procéder à un recensement de ce groupe. La liste des personnes appartenant à la population cible était à jour, et 135 des ménages ont répondu à l'enquête, ce qui correspond à un taux de réponse élevé de 96 %.

PDI Roms, Ashkalies et Égyptiennes (RAE)

Comme pour les PDI serbes en centre collectif, les partenaires du profilage ont décidé de procéder à un recensement complet de tous les ménages RAE identifiés, étant donné leur petit nombre. Au total, 137 ménages avaient été identifiés, mais la liste qui récapitulait les membres de la population cible était un peu dépassée et il était probable qu'une partie des ménages ne pourrait pas être localisée. Dans ce contexte, une approche d'échantillonnage boule de neige a été utilisée pour compléter la liste des ménages de la base de données du HCR. L'approche adoptée peut être décrite comme suit :



Les ménages de la liste ont été contactés par les recenseurs et interviewés. Il leur a aussi été demandé de désigner d'autres membres de la population cible de leur connaissance. Cette approche a permis d'interroger 123 ménages, soit 491 personnes. Les ménages interrogés se trouvaient principalement à Mitrovicë/Mitrovica sud (39 ménages), Gračanica/Graçanicë (26 ménages) et, dans une moindre mesure à Obiliq/ Obilić (11 ménages) et à Fushë Kosovë / Kosovo Polje (13 ménages). L'approche boule de neige a permis aux recenseuse.eur.s d'identifier que le point de saturation était quasiment atteint, ce qui signifiait que tous les ménages des PDI RAE des municipalités visées avaient été identifiés et interviewés. Il a donc pu être conclu que les résultats pouvaient être considérés comme représentatifs de l'ensemble de la population cible.

Difficultés et limites

L'approche d'échantillonnage initiale a dû être repensée en raison du manque de précision du cadre d'échantillonnage fourni par le recensement, ce qui a rendu nécessaire le recours à une approche alternative. L'approche d'échantillonnage non probabiliste adoptée dans un deuxième temps pour identifier les PDI albanaises n'a pas atteint le point de saturation (502 ménages sur un total estimé de 1 167 ont été interrogés), ce qui signifie que la représentativité de l'échantillon n'a pas pu être déterminée avec certitude. En outre, la majorité des ménages de l'échantillon a été identifiée dans la municipalité de Mitrovice/Mitrovica, et un nombre plus restreint à Prishtine/Pristina. Comme il était probable qu'un certain nombre de ménages de PDI soient arrivés à Prishtine/Pristina au cours des dernières années, il a été conclu que seul l'échantillon identifié par boule de neige pouvait être considéré comme représentatif pour la municipalité de Mitrovice/Mitrovica.

Peu de données secondaires étaient disponibles pour la population globale du Kosovo. Trois sources de données secondaires ont été utilisées : le recensement de population et de logement (Population and Housing Census) de 2011, les enquêtes en grappes à indicateurs multiples de l'UNICEF en 2013-2014 et l'enquête Mosaic du PNUD en 2015. Comparer les résultats de l'enquête sur les personnes déplacées menée en 2016 et ceux du recensement de 2011 est loin d'être idéal. Comme les données sont représentatives de deux périodes différentes, il aurait été problématique de comparer les résultats des deux sources, car les caractéristiques pertinentes de l'ensemble de la population du Kosovo étaient susceptibles d'avoir changé de manière significative entre 2011 et 2016.

Exercice de profilage à Mogadiscio, Somalie, 2015-2016



Objectifs

Deux décennies de conflits armés couplées à des sécheresses et inondations massives et récurrentes ont poussé un très grand nombre de somaliens à quitter leurs foyers. Mogadiscio accueille la plus grande concentration de personnes en situation de déplacement prolongé. La plupart de ces PDI vivent dans les implantations informelles de la ville et de sa périphérie.

L'objectif global de l'exercice de profilage était de fournir une base factuelle pour contribuer à la planification de solutions durables pour les PDI par le biais d'actions de plaidoyer et de réponses humanitaires et de développement conjointes. Plus spécifiquement, l'exercice visait à :

- (1) Fournir une **estimation désagrégée des chiffres** de la population de personnes déplacées vivant dans des implantations informelles ;
- (2) Identifier les déplacements intérieurs en analysant **l'histoire migratoire des ménages** ;
- (3) **Analyser la situation actuelle des différentes populations** cible, y compris leur situation socio-économique, leurs conditions de vie, leur accès aux services de base et leurs préoccupations en matière de protection ;
- (4) **Comprendre la résilience des familles**, définie comme les compétences, les capacités, les besoins particuliers et les mécanismes d'adaptation des populations cible qui vont contribuer aux décisions qu'elles prennent pour leur avenir.

Population cible

Trois populations cible ont été incluses dans l'enquête d'échantillonnage de l'exercice de profilage :

- (1) **PDI** : personnes d'origine somalienne qui ont été déplacées de leur lieu d'origine en direction ou à Mogadiscio principalement en raison d'un conflit, d'une catastrophe, de l'insécurité ou de la violation de leurs droits.
- (2) **Migrants économiques** : personnes d'origine somalienne qui ont quitté leur lieu d'origine et sont venues à Mogadiscio à la recherche de moyens de subsistance et pour d'autres raisons économiques.
- (3) **Communautés d'accueil** : personnes originaires de Mogadiscio qui n'ont jamais quitté leur lieu d'origine et vivent dans des implantations informelles à Mogadiscio.

Les populations cible suivantes ont également été incluses dans l'exercice de profilage, mais exclues de l'enquête par échantillonnage. Les informations sur ces groupes ont été obtenues par le biais de collectes de données qualitatives :

- (4) **Rapatrié.e.s** : personnes d'origine somalienne ayant fui la Somalie principalement en raison d'un conflit ou d'une catastrophe et qui depuis sont retournées en Somalie.
- (5) **Réfugié.e.s** : personnes originaires d'autres pays qui se sont réfugiées en Somalie et à Mogadiscio à la recherche d'une protection internationale.

Cadre d'échantillonnage

Le contexte partiel de l'exercice de profilage était que les partenaires du profilage ne disposaient d'aucune liste complète identifiant les PDI, les migrants économiques et les communautés d'accueil vivant dans les implantations informelles de Mogadiscio. Il a donc été décidé que des cadres d'échantillonnage devraient être élaborés par le biais d'une cartographie des implantations informelles et d'un dénombrement des ménages y résidant.

Unités d'échantillonnage et d'analyse

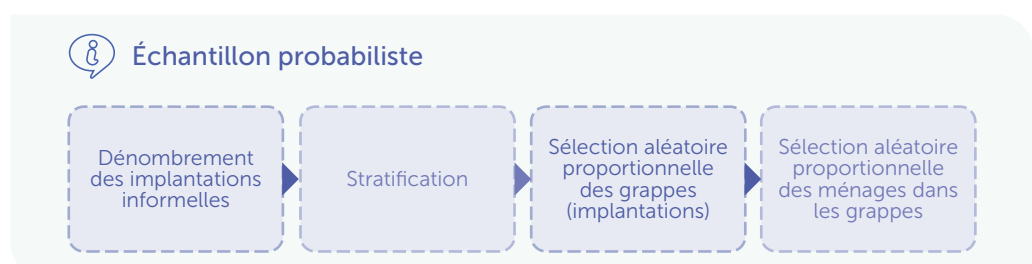


Méthode d'échantillonnage

Conformément aux objectifs de l'exercice, les partenaires chargés du profilage devaient adapter leur approche d'échantillonnage de manière à pouvoir estimer le nombre total de personnes déplacées vivant dans des camps à Mogadiscio et aboutir à des informations sur la situation et les caractéristiques des PDI, des migrants économiques et des communautés d'accueil pouvant être extrapolées à l'ensemble des trois populations cible. Les partenaires voulaient également pouvoir comparer les trois populations cible – les personnes déplacées, les migrants économiques et les communautés d'accueil – et identifier la localisation géographique de la population cible des PDI. Vous trouverez ci-dessous une description des approches adoptées pour chacune des populations cible.

PDI

Une approche d'échantillonnage en grappe stratifiée à deux phases a été retenue pour cette population cible, où les implantations faisaient office de grappes (c'est-à-dire d'unités d'échantillonnage primaires) et les ménages d'unités d'échantillonnage secondaires. Cette approche peut être résumée comme suit :



Étape 1 : Cartographie et dénombrement

Les partenaires du profilage ont en premier lieu procédé à une cartographie de toutes les implantations informelles de Mogadiscio où les PDI étaient supposées vivre. Cette étude documentaire s'est consacrée à l'identification des implantations à Mogadiscio en répertoriant les informations existantes pour les implantations désignées, leurs coordonnées GPS et les estimations de population. La cartographie a mis en lumière de nombreux manques d'information, notamment concernant les deux districts (Daynille et Kaxda) connus pour ou supposés abriter la plus grande population de PDI et le plus grand nombre d'implantations informelles. Il a donc été décidé qu'un dénombrement serait mis en place pour identifier et procéder à un comptage de toutes les personnes vivant dans ces implantations. Cela allait permettre d'obtenir des estimations de la population des IDP des implantations (pour remplir l'objectif (1)), et servir de cadre d'échantillonnage pour les trois populations cible de l'enquête.

Le dénombrement visait à effectuer un recensement complet des ménages de toutes les implantations identifiées lors de la cartographie. Les partenaires du profilage ont, en outre, collaboré avec les commissaires de district pour identifier d'autres implantations éventuelles dans les districts à dénombrer. Pour les districts de Daynille et de Kaxda, l'imagerie spatiale de Google Earth a été utilisée pour quadriller la zone et la décomposer en zones de dénombrement gérables.

Le processus de dénombrement a permis d'identifier les implantations informelles vides et de les exclure plus tard de la procédure d'échantillonnage. Pour chaque ménage dénombré, un code unique a été peint au pochoir sur la structure extérieure de l'abri pour éviter que les ménages ne soient décomptés deux fois et faciliter l'identification des ménages ciblés pour l'enquête. Les questions du dénombrement visaient à identifier les membres des différentes populations cible. Le dénombrement a identifié 68 796 ménages PDI, représentant 399 292 individus.

Étape 2 : Stratification

Pour être en mesure d'analyser les situations et les différentes caractéristiques des ménages de PDI, il fallait sélectionner un échantillon représentatif de cette population cible pour que les résultats puissent être considérés comme valables pour l'ensemble de la population cible. Les parties prenantes souhaitaient également pouvoir comparer les résultats des PDI des districts de Daynille et Kaxda (qui abritaient le nombre le plus important de PDI) avec ceux des autres districts.

La population cible a donc été divisée en trois strates sur la base de la localisation géographique de la population des PDI :

- (1) District de Daynille (accueillant 35 % des ménages PDI de Mogadiscio)
- (2) District de Kaxda (accueillant 20% des ménages PDI de Mogadiscio)
- (3) Le reste des districts de Mogadiscio

Étape 3 : Décider de la taille de l'échantillon

La taille de l'échantillon devait être représentative avec un niveau de confiance de 95 % et une marge d'erreur de 5 % pour l'ensemble de l'échantillon, et une marge d'erreur de 10 % lorsque l'échantillon était stratifié. La taille de l'échantillon a été fixée à 612 ménages.

Étape 4 : Sélectionner un échantillon d'implantations

À la suite de la cartographie, du dénombrement des implantations et de la stratification des zones géographiques, une sélection aléatoire des implantations a été effectuée avec une probabilité proportionnelle à la taille de la population des PDI de chaque strate.

Étape 5 : Sélectionner un échantillon de ménages

À la suite de la sélection des implantations, des ménages ont été tirés aléatoirement au sein des implantations sélectionnées pour être intégrés à l'échantillon final.

Migrants économiques

Le dénombrement des implantations a donné lieu à l'établissement d'une base de données pouvant servir de cadre d'échantillonnage pour cette population cible. Un échantillon aléatoire simple a été tiré parmi les ménages identifiés comme des migrants économiques lors du dénombrement.

Cet échantillon était représentatif avec un taux de confiance à 95 % et une marge d'erreur de 5 %. La taille de l'échantillon a été fixée à 638 ménages.



Communauté accueillante

Comme pour les migrants économiques, les résultats du dénombrement ont pu servir de cadre d'échantillonnage pour la population cible. Un échantillon aléatoire simple a été tiré parmi les ménages identifiés comme appartenant à la communauté d'accueil.

Cet échantillon était représentatif avec un taux de confiance à 95 % et une marge d'erreur de 5 %. La taille de l'échantillon a été fixée à 638 ménages.

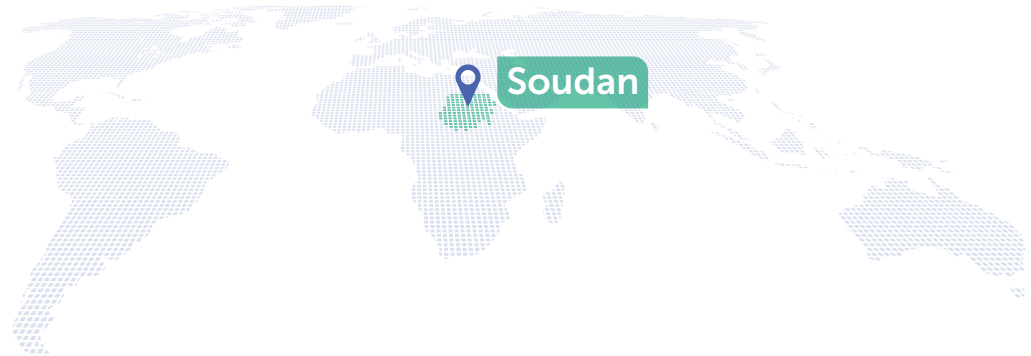


Difficultés et limites

L'instabilité de la situation à Mogadiscio, notamment en raison des expulsions subies par les ménages déplacés, a posé des difficultés significatives aux partenaires du profilage. Une accélération des expulsions en mars et avril 2015 a provoqué l'arrêt du processus pour cette période. **Afin d'atténuer les effets de ces changements de situation brutaux, le délai entre le dénombrement et l'enquête d'échantillonnage a été réduit.** Malgré cela, les expulsions ont continué à poser problème tout au long du processus et de la phase de collecte de données, bien que de manière moins prégnante. En revanche, comme les expulsions ont principalement engendré des déplacements à l'intérieur de la ville même, c.à.d. d'une implantation de PDI à une autre, elles n'ont eu qu'un impact limité sur les chiffres globaux et l'identification des difficultés principales rencontrées par la population de Mogadiscio. Par conséquent, le profilage a tout de même permis d'aboutir à une analyse exhaustive de l'ampleur et des grandes tendances de la situation de déplacement, ainsi qu'à une analyse thématique pour éclairer la planification de solutions durables.

Deux districts – Heliwa and Abdulaziz – ont été exclus de l'enquête-ménage pour des raisons de sécurité et les résultats du profilage ne peuvent donc pas être généralisés à ces deux zones. Tous deux ont cependant été inclus dans la cartographie des installations et dans la zone de dénombrement.

Exercice de profilage au Soudan, 2017-2019



Objectifs

Quinze ans après le violent conflit ayant éclaté au Darfour en 2003, la majorité des PDI de la région vivait toujours dans des camps qui étaient presque devenus des installations permanentes et continuaient à dépendre de l'aide humanitaire. Cette situation de déplacement prolongé étant loin d'être résolue, il est devenu crucial de trouver des solutions durables au déplacement pour les PDI et les réfugiés rapatriés.

L'exercice de profilage avait pour but de fournir des données exhaustives permettant de donner une image précise de la situation à laquelle étaient confrontées les personnes déplacées vivant dans les camps à la périphérie de El Fasher, Darfour du Nord. Les objectifs du profilage étaient les suivants :

- (1) **Fournir un profil complet des PDI** résidant dans les camps d'Abu Shouk et d'El Salaam ainsi que dans les zones de la ville voisine d'El Fasher. Ce profil comprenait des données démographiques de base, des informations socio-économiques et sur les moyens de subsistance, un éclairage sur les besoins des personnes déplacées, leurs vulnérabilités, leurs mécanismes d'adaptation, capacités et moyens. Ce profil prenait également en compte leurs intentions et projets d'installation ainsi que les facteurs encourageant ou entravant un éventuel nouveau déplacement, un retour ou une intégration locale.
- (2) **Proposer une analyse comparative** des populations déplacées et non déplacées afin d'identifier les vulnérabilités spécifiques des personnes déplacées liées à leur déplacement. Cette analyse visait à permettre une meilleure compréhension des obstacles et des possibilités en matière de solutions durables.
- (3) **Contribuer aux réponses programmatiques** du gouvernement soudanais et des acteurs de l'humanitaire et du développement et renforcer la résilience des communautés d'accueil conformément au Cadre de l'IASC.
- (4) **Piloter un processus conjoint de profilage du déplacement et de planification de solutions durables** dans le but de pouvoir étendre ce modèle à d'autres contextes de déplacement au Soudan. À ce titre, le pilote devait permettre le développement d'outils de collecte et de méthodologies convenus qui pourraient être réutilisés après les adaptations nécessaires.

Population cible

Les populations cible de l'exercice étaient les suivantes :

- (1) La population des PDI vivant dans les deux camps de Abu Shouk et El Salam ;
- (2) La population non déplacée vivant à la périphérie d'El Fasher (une zone urbaine située à la sortie d'El Fasher et formant un rayon de 3 km autour des deux camps et dans la ville d'El Fasher même (les quartiers urbains de la ville d'El Fasher).

Cadre d'échantillonnage

Aucun registre ou autre type de liste de la population cible n'était disponible pour les partenaires du profilage. Il a donc été décidé qu'un cadre d'échantillonnage devrait être élaboré pour le processus de sélection de l'échantillon.

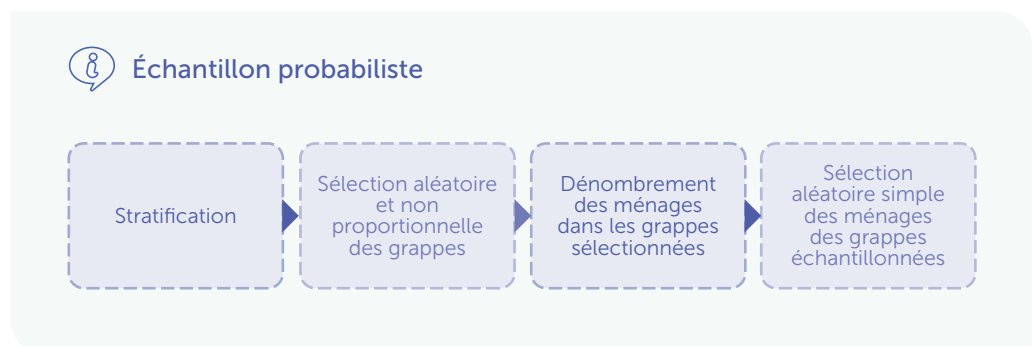
Unités d'échantillonnage et d'analyse



Méthode d'échantillonnage

Les objectifs du profilage étaient de fournir une analyse des situations et des caractéristiques des personnes déplacées par rapport à celles de leurs voisins non déplacés. Les partenaires souhaitaient également pouvoir comparer les personnes déplacées des deux différents camps et les personnes non déplacées vivant dans des environnements péri-urbains et urbains, car on s'attendait à ce que ces groupes soient confrontés à des situations différentes selon leur localisation. L'approche d'échantillonnage adoptée devait donc produire un échantillon représentatif pour tous ces sous-groupes.

En l'absence d'un cadre d'échantillonnage déjà existant, les partenaires du profilage ont décidé de procéder à un échantillonnage en grappes stratifié à deux phases. Cette approche peut être décrite de la manière suivante :



Étape 1 : Stratification

La population cible a été divisée en quatre strates en fonction des différents groupes à comparer :

- PDI du camp de Abu Shouk
- PDI du camp de El Salam
- Population non déplacée dans les zones péri-urbaines de El Fasher
- Population non déplacée dans les zones urbaines de El Fasher

Ces quatre strates ont à leur tour été divisées en grappes plus petites qui, d'après le quadrillage effectué à partir de la cartographie de la zone concernée, étaient supposées avoir des tailles de population équivalentes.

Étape 2 : Décider de la taille de l'échantillon

Dans chacune des strates, un échantillon aléatoire de grappes a été sélectionné et dénombré. Puis, les ménages qui allaient être inclus dans l'échantillon final ont été sélectionnés au sein des grappes échantillonnées. La taille de l'échantillon a été fixée à 3 000 ménages, qui devaient représenter 252 grappes. L'échantillon des ménages et des grappes planifié a été réparti entre les strates de la manière suivante :

Type de population	Strates	Nombre de ménages	Nombre de grappes
PDI	Camp de Abu Shouk	996	84
	Camp de El Salam	996	84
Communauté d'accueil	Voisine de El Fasher	504	42
	Non voisine de El Fasher	504	42
Total		3,000	525

Étape 3 : Sélectionner un échantillon de grappes

Un échantillon simple de grappes a été tiré pour chacune des strates. Aucun chiffre de population fiable n'existait pour les grappes et celles-ci ne pouvaient donc pas être sélectionnées avec une probabilité proportionnelle à leur taille. Cela impliquait que des pondérations d'échantillonnage devraient être appliquées lors de l'agrégation des résultats des différentes strates à une moyenne de la population globale.

Étape 4 : Dénombrement des grappes sélectionnées

Lors de l'étape suivante, tous les ménages de chacune des grappes sélectionnées ont été dénombrés. L'établissement de cette liste a abouti à **un nombre de grappes dénombrées bien inférieur à ce qui avait été prévu initialement. En effet, certaines des grappes sélectionnées se sont avérées impossibles à atteindre pour les recenseurs, principalement dans le camp d'El Salam.** De plus, les grappes qui abritaient moins de 12 ménages ont été retirées de l'échantillon. Certains des ménages dénombrés qui avaient été listés deux fois ou vivaient hors du périmètre des grappes ont également été retirés de l'échantillon.

Étape 5 : Sélectionner un échantillon de ménages

Le dénombrement des ménages au sein des grappes sélectionnées a fourni un cadre d'échantillonnage pour la sélection par échantillonnage aléatoire simple de l'échantillon final des ménages à interviewer. Pour sélectionner l'échantillon, les données du dénombrement ont été limitées aux ménages déplacés des strates des deux camps et aux ménages non déplacés des deux strates d'El Fasher. Le nombre de ménages à sélectionner devait initialement être réparti de manière égale entre les grappes sélectionnées, avec 12 ménages sélectionnés dans chacune d'entre elles. Cependant, en raison du nombre moins élevé de grappes dénombrées, la sélection pour les entretiens de 12 ménages par grappe aurait conduit à une taille d'échantillon nettement inférieure à celle des 3 000 ménages prévus. Par conséquent, certaines grappes ont été aléatoirement sélectionnées dans chacune des strates pour être sur-échantillonnées afin de compenser la baisse du nombre de grappes. Les grappes sélectionnées pour être sur-échantillonnées devaient faire l'objet de 24 interviews. Ainsi, seules les grappes comptant plus de 24 ménages répertoriés étaient éligibles pour être sélectionnées pour le sur-échantillonnage.

Des pondérations d'échantillonnage ont été calculées et appliquées aux grappes en fonction de leur taille et du nombre de ménages sélectionnés dans chaque grappe (12 ou 24). Les résultats pondérés n'ont pas pu être confrontés aux données de population en raison d'un manque d'informations actualisées et fiables. Néanmoins, en ce qui concerne les camps, les tailles de population obtenues grâce aux pondérations se rapprochaient des dernières estimations de population de l'Organisation internationale des migrations (OIM).

Le tableau ci-dessous montre la répartition finale et effective des grappes répertoriées, des ménages interrogés et des individus inclus dans l'échantillon interviewé de chacune des quatre strates :

Type de population	Strate	Nombre de grappes listées	Nombre de ménages interviewés	Nombre d'individus interviewés
PDI	Camp de Abu Shouk	82	996	5,849
	Camp de El Salam	50	986	5,960
Communauté d'accueil	Voisines de El Fasher	40	509	3,303
	Non voisine de El Fasher	41	511	3,376
Total		213	3,002	18,533

Difficultés et limites

L'échantillon dans les deux strates d'El Fasher (urbaine et péri-urbaine) ne comprenait que des ménages non déplacés. Il faut donc garder à l'esprit que les résultats en provenance de ces strates ne représentent pas la situation de tous les résidents de ces quartiers urbains et péri-urbains, étant donné que certains des ménages déplacés de ces strates n'ont pas été inclus dans l'échantillon. Plus précisément, le dénombrement a indiqué qu'environ 28 % des ménages péri-urbains et 21 % des ménages urbains étaient des déplacés internes. L'analyse comparative s'est donc concentrée sur les différences relatives au statut de déplacement entre ces groupes de population et ne peut pas être utilisée pour comparer les zones en tant que telles.

Les objectifs de l'exercice, à savoir apporter une contribution aux programmes d'intégration locale et régionale, auraient bénéficié d'une comparaison supplémentaire par zone géographique (centre-ville vs zones péri-urbaines vs camps). Une telle approche aurait soit nécessité un échantillon plus important (pour inclure les deux groupes cibles – c.à.d. les personnes déplacées et non déplacées – dans les strates péri-urbaine et urbaine), soit une réduction de chaque échantillon tiré (afin ne pas mobiliser plus de ressources globalement). Ceci aurait conduit à des échantillons moins robustes sur le plan statistique mais aurait permis de conférer à l'analyse une pertinence programmatique. Il s'agit de considérations que les partenaires doivent avoir à l'esprit dès le démarrage et qui doivent être priorisées conjointement.