

VI. Méthodes quantitatives

Contrairement aux études qualitatives, les méthodes quantitatives mesurent directement l'état ou la variation d'une variable et donnent des résultats numériques directs. Par exemple, la proportion des femmes ayant réalisées au moins quatre consultations prénatales durant la grossesse, la prévalence de l'anémie chez les enfant de moins de 5 ans. Dans l'approche quantitative, les deux qualités essentielles sont la validité de la mesure et la représentativité de l'échantillon choisis pour effectuer la mesure (cf. échantillonnage).

Le but des méthodes quantitatives de collecte de données est de quantifier et de mesurer un phénomène grâce à l'utilisation de questionnaires structurés et au traitement statistique des informations collectées. **Les enquêtes par questionnaire sont la principale méthode quantitative de collecte de données.** Elles consistent à interroger, à l'aide d'un questionnaire, un échantillon d'individus le plus représentatif possible de l'ensemble de la population étudiée. Un des avantages des méthodes quantitatives est de permettre de recueillir, lors d'une seule enquête, des informations comparables d'un individu à l'autre. C'est cette comparabilité des informations qui permet ensuite les dénombrements et, plus généralement, l'analyse statistique (ce que ne permettent pas les méthodes qualitatives de collecte de données). Une enquête doit être relativement simple à effectuer. Les résultats doivent être disponibles suffisamment tôt pour que les données soient utiles à l'élaboration d'une intervention.

Il n'est pas difficile de réaliser une enquête, mais il y a un certain nombre de points critiques qui doivent être effectués correctement afin que les résultats soient valides. La planification, la formation du personnel, leur supervision, l'interaction avec la communauté, les autorités à tous les niveaux et une compréhension basique des concepts épidémiologiques et statistiques sont des éléments indispensables.

Nous avons ici principalement ciblé le propos sur les enquêtes transversales mais la plupart des recommandations méthodologiques et opérationnelles ne sont pas spécifiques à ces types d'enquêtes et sont valables pour d'autres types d'enquêtes (enquêtes sur les connaissances, attitudes et pratiques, enquêtes accès aux soins, enquêtes nutritionnelles, etc.). La principale différence tiendra dans ces cas au contenu des questionnaires utilisés pour réaliser ces enquêtes.

Les enquêtes complexes, qui essaient de répondre à beaucoup de questions en même temps et cherchent à donner une image complète sont difficiles à conduire, analyser et interpréter. Elles sont également coûteuses et nécessitent une expertise très spécialisée. De plus, parce que ces enquêtes sont longues à réaliser, les informations sont souvent périmées le temps que l'enquête soit terminée, et elles ne peuvent pas être reproduites facilement pour évaluer l'évolution de la situation. Il est presque toujours préférable de faire une enquête relativement simple qui répond seulement aux questions les plus pressantes et cruciales, et qui peut être répétée selon l'évolution de la situation. Chaque donnée supplémentaire recueillie, même si elle est relativement simple, empiète sur la qualité et le soin avec lesquels les données critiques sont recueillies, et retarde la fin de l'enquête.

Une enquête est utile dans toutes les phases du cycle de projet (avant le démarrage des activités, pendant la mise en œuvre et en fin de programme).

Une fois la décision de conduire une enquête est prise, il faut organiser à l'avance les modalités précises de la réalisation de l'enquête dans ses moindres détails, en s'efforçant notamment de penser à toutes les situations concrètes qui sont susceptibles de se présenter durant son déroulement: c'est le but du protocole de l'enquête, qui doit être rédigé de façon explicite avant sa mise en œuvre. Cela est très important pour éviter toute improvisation dans la réalisation, toute dérive dans les objectifs, tout flou dans les conditions de recueil des

données, etc. qui sont l'occasion d'autant de biais possibles. Les principaux éléments constitutifs d'un protocole sont les suivants:

- Résumé
- Le rationnel et la justification de l'enquête
- Les objectifs de l'enquête
- La méthodologie de l'enquête
- La gestion, l'organisation, la logistique et le travail de terrain
- Analyse, interprétation et dissémination des résultats
- Protection des données
- Considérations éthiques et légales
- Archivage des données
- Références bibliographiques
- Annexes

6.1. Rationnel et justifications de l'enquête

La décision d'entreprendre une enquête est habituellement prise en concertation avec les autorités, les agences partenaires et les bailleurs de fonds. La première étape consiste à dresser une liste des questions que l'on se pose. Cette liste est établie à la suite de recherches bibliographiques internationales et nationales ainsi que d'expériences et constats des équipes terrain. Il est toujours important de disséminer l'information de quand et où vous projetez entreprendre une enquête afin d'empêcher toute répétition ou duplication inutile par d'autres organisations. Les enquêtes sont d'habitude beaucoup plus instructives si elles sont coordonnées de sorte que les données de plusieurs agences, secteurs géographiques ou groupes de population puissent être examinées ensemble afin de donner une perspective plus large sur la situation.

Entreprendre une enquête est coûteux et prend du temps. De ce fait, avant de commencer une enquête, vous devriez considérer les points suivants:

- Est-ce qu'une autre agence ou organisation n'a pas fait le même travail dans la même zone géographique ? Il faut vérifier ce qui a déjà été écrit sur le sujet. Ce qui est démontré, ce qui interroge. Vérifier les données dont on dispose sur le sujet. Sont-elles disponibles, exploitables, suffisantes ? Y a-t-il des méthodes spécifiques pour aborder cette question ? Qu'est ce qui a été employé jusqu'à présent ? Si par exemple une autre agence a récemment effectué une enquête nutritionnelle dans le même secteur, les résultats de celle-ci devraient être employés plutôt que de répéter une nouvelle enquête.
- Les résultats seront-ils essentiels à la prise de décision ? Si les besoins d'une population sont évidents, l'exécution immédiate de programmes prend la priorité sur le fait d'effectuer une enquête, et l'enquête sera conséquemment reportée.
- Est-il anticipé que les résultats mèneront à une prise d'action ? Il y a peu d'intérêt à faire une enquête si vous savez déjà qu'une intervention ne sera pas possible. Cependant, si l'organisation ayant conduit l'enquête ne peut pas elle-même

implémenter un programme, les résultats doivent tout de même servir à promouvoir une prise d'action par un autre acteur.

- La population affectée est-elle accessible? La stabilité de la population, le type d'habitat (dispersé ou non ?, l'insécurité, les contraintes géographiques (conditions d'accès au terrain) et climatiques peuvent avoir comme conséquence l'accès limité à la population d'intérêt. Si ces contraintes sont extrêmes, une enquête ne peut pas être effectuée.

6.2. Objectifs de l'enquête

Les objectifs permettent de savoir quelles informations on va chercher à collecter et donnent un sens à l'enquête. Des objectifs clairs et précis facilitent une bonne compréhension de la raison pour laquelle l'enquête est entreprise, pour vous, votre équipe, la population locale et les bailleurs de fonds. Par exemples : estimer la prévalence de la malnutrition aiguë chez les enfants âgés de 6 à 59 mois dans la région XYZ ou bien estimer les besoins en information, éducation et communication des mères d'enfants âgés de 6 - 24 mois dans la région Y en matière d'alimentation du nourrisson et du jeune enfant (ANJE), dans le but d'adapter le programme IEC.

6.3. Méthodologie de l'enquête

6.3.1. Définir la région géographique et le groupe cible à enquêter

Dans l'élaboration de l'enquête, la zone géographique et la population à enquêter doivent être soigneusement définies. Le rapport devrait contenir une carte du secteur. Beaucoup d'agences effectuent des enquêtes limitées aux secteurs dans lesquels elles sont présentes ou prévoient déjà d'implémenter des programmes.

Un groupe cible (population à enquêter) est une population précisément définie qui peut ainsi faire l'objet d'une intervention ou d'une étude spécifique. Il peut s'agir d'un groupe spécifique de la population qui partage des caractères communs, par exemple les enfants de moins de 5 ans, les femmes enceintes et allaitantes (ici les questionnaires sont adressés aux individus) ou d'une population plus générale, par exemple sur une région ou un village (questionnaires adressés aux ménages).

6.3.2. Rencontrer les chefs communautaires et les autorités locales

Il est absolument essentiel de rencontrer les chefs communautaires et les autorités locales avant de commencer une enquête. Les réunions devraient couvrir au moins les points

suivants:

- Se mettre d'accord avec la communauté au sujet des objectifs de l'enquête. Si la population ne comprend pas pourquoi vous faites une enquête, elle risque de ne pas coopérer.
- Obtenir une carte de la zone afin de planifier l'enquête. Employez cette carte durant les discussions avec les autorités locales et les chefs communautaires.
- Obtenir des chiffres de population détaillés (en particulier au niveau du village).
- Obtenir des informations sur la sécurité et sur l'accès aux zones (potentielles) d'enquête.

- Obtenir des lettres de permission officielle auprès des autorités locales adressées aux chefs de zone ou de village, déclarant que vous allez les visiter. Les lettres devraient expliquer pourquoi vous conduisez une enquête et demander la coopération de la population.
- Se mettre d'accord sur les dates de l'enquête avec la communauté et les autorités locales.
- Se mettre d'accord sur comment les résultats seront employés. En particulier, discuter de façon réaliste des perspectives, des mesures qui seraient prises et des types de programmes qui sont susceptibles d'être implémentés si la situation s'avère aussi grave que prévue. Ne pas faire de promesses qui ne puissent être tenues.

6.3.3. Méthodes d'échantillonnage

Les méthodes d'échantillonnage consistent à construire un échantillon d'une population mère afin d'en estimer les caractéristiques et opinions. On distingue 2 catégories de méthodes d'échantillonnage : les méthodes probabilistes et les méthodes non probabilistes (dites également empiriques).

Méthode probabiliste

La méthode probabiliste vise à tirer au hasard l'échantillon dans la population à analyser, donnant à chaque composante de celle-ci une probabilité connue, non nulle, d'être choisie. La méthode est dite probabiliste, car la représentativité de l'échantillon obtenu est à priori assurée par les lois statistiques de la probabilité. Contrairement à l'usage d'une méthode non probabiliste ou raisonnée, l'échantillonnage probabiliste permet de calculer et d'appliquer des seuils et intervalles de confiances aux résultats obtenus.

Il existe différentes méthodes parmi les plus connues : le sondage ou échantillonnage aléatoire simple, le sondage ou échantillonnage en grappes, le sondage à plusieurs degrés, le sondage stratifié.

Bien qu'une enquête exhaustive soit la meilleure manière d'obtenir les informations les plus fidèles d'une population, très souvent on procèdera à un sondage. Voici quelques raisons d'effectuer un sondage au lieu d'un recensement ou d'une enquête exhaustive:

- Lorsque la population est trop grande puisque cela engendre moins de dépenses coûts (transport, temps, employés, etc.).
- Lorsqu'on ne bénéficie pas de beaucoup de temps.
- Lorsque la population ciblée est difficilement accessible

La méthode d'échantillonnage est choisie en fonction de la manière dont la population (ou les ménages) est distribuée et de la taille de la population à examiner.

- De temps en temps, dans les populations de très petite taille, chaque ménage dans la population peut être visité (enquête exhaustive).
- Quand les domiciles de toute la population à enquêter sont arrangés d'une manière systématique (comme dans un camp de réfugiés), un échantillonnage aléatoire simple ou systématique peut être employé.
- L'échantillonnage par grappes est employé quand les ménages sont distribués d'une manière non structurée qui ne permet pas d'identifier ou de numéroter facilement tous les ménages.

Echantillonnage aléatoire simple ou probabiliste

Quand une liste de chaque ménage ou individu (et si possible l'adresse) est disponible, les individus sont aléatoirement choisis à partir de la liste en utilisant une table de nombres aléatoires. Ainsi chaque individu a la même chance d'être choisi. C'est une situation rare et cette méthode est rarement employée.

Echantillonnage systématique

Parfois appelé échantillonnage par intervalles, l'échantillonnage systématique signifie qu'il existe un écart, un intervalle ou pas de sondage, entre chaque unité sélectionnée qui est incluse dans l'échantillon. Le premier sujet de la liste est choisi au hasard; les suivants de façon systématique. On l'utilise lorsque la population que l'on souhaite étudier est reliée à un site identifié, par exemple la file active d'un centre de santé. On définit dans ce cas un intervalle régulier selon lequel on choisira les personnes à qui proposer l'enquête (par exemple interroger un patient sur cinq). On peut sélectionner au dé, par exemple, la première personne ou le premier ménage interrogé puis on applique l'intervalle choisi. On peut également se servir de tables aléatoires.

Echantillonnage stratifié

On utilise cette méthode lorsqu'on cherche à cibler des sous-groupes spécifiques dans la population à enquêter, par exemple par la classe d'âge, le genre ou encore le fait de résider en zone urbaine ou rurale. On divise alors la population en groupes homogènes (appelés strates), qui sont mutuellement exclusifs, puis on sélectionne à partir de chaque strate des échantillons indépendants. Cela permet d'éviter une sous- ou une surreprésentation de certaines strates à cause du hasard. Chaque sujet de chaque strate est choisi au hasard. La méthode d'échantillonnage stratifié entraîne la sélection d'unités de tous les groupes. L'échantillonnage stratifié nous assure d'obtenir une taille d'échantillon suffisante pour des sous-groupes de la population à laquelle nous nous intéressons. Étant donné que chaque strate devient une population indépendante lorsque l'on stratifie une population, on doit déterminer pour chaque strate la taille de l'échantillon. Cependant, la méthode d'échantillonnage stratifié n'est réalisable que si l'on connaît précisément la proportion de chacun des groupes dans la population, ce qui suppose qu'on dispose d'une liste assez exhaustive des membres de ces groupes (type recensement, enquête sociodémographique).

Echantillonnage en grappes

Il est parfois trop dispendieux de disséminer un échantillon dans l'ensemble de la population. Les coûts de déplacement risquent de devenir élevés lorsque les enquêteurs doivent sonder des gens d'un bout à l'autre de la zone choisie. La technique de l'échantillonnage en grappes permet de réduire les coûts. La forme la plus commune d'échantillonnage en grappes se fait en deux étapes. D'abord, la population entière est divisée, sur papier, en plusieurs petits secteurs géographiques, tels que des villages. Pour chaque village, la taille de population doit être connue ou doit pouvoir être estimée. Des grappes sont aléatoirement sélectionnées parmi ces villages, avec la chance pour chaque village d'être choisi proportionnelle à sa taille. Ceci signifie que chaque personne dans le secteur entier a une chance égale d'être sélectionné. Ceci s'appelle échantillonnage avec la "probabilité proportionnelle à la taille de population". Dans la deuxième étape, les individus sont choisis au hasard dans chaque grappe.

L'échantillonnage à plusieurs degrés

L'échantillonnage à plusieurs degrés est une combinaison des méthodes d'échantillonnage aléatoire précédentes, et la méthode choisie peut varier d'un degré à l'autre. Ce type de sélection est celui qui garantit le plus la représentativité de l'enquête par rapport au groupe

cible et la validité des résultats, c'est également une des méthodes les plus complexes. Par exemple si on veut conduire une enquête dans un district donné, on procède d'abord à un échantillonnage en grappes pour sélectionner de façon aléatoire les villages qui seront retenus dans l'enquête (1er degré). Puis, parmi les villages retenus, on sélectionne les ménages qui seront retenus dans l'enquête (2e degré). On peut pour cela utiliser par exemple un échantillonnage systématique, au cours duquel les ménages sont choisis à intervalles réguliers. Le 3e degré de l'échantillonnage consiste à sélectionner les répondants (femmes en âge de procréer, mères d'enfants âgés de 6 à 24 mois, etc.) au questionnaire.

Illustration des subtilités des méthodes probabilistes :

Si on interroge les individus de la population d'un village au hasard (hasard calculé sur la base de la liste de tous les individus habitant le village), on fera un échantillonnage aléatoire simple. Cela suppose qu'on ait une liste des habitants du village.

Si on regroupe la population par lieux d'habitation (rural ou urbain et qu'on interroge au hasard des individus dans chacun des milieux, on fera un échantillonnage aléatoire stratifié.

Si on regroupe la population par villages et qu'on sélectionne au hasard quelques villages à l'intérieur desquels on interroge tous les d'individus, il s'agit d'un échantillonnage en grappe.

Si on regroupe la population par villages et qu'on sélectionne des ménages à l'intérieur desquels on interroge des échantillons d'individus, il s'agit d'un sondage à plusieurs degrés.

Les méthodes non probabilistes

La méthode non probabiliste consiste à sélectionner l'échantillon, non pas à partir d'un tirage aléatoire mais basé sur un choix réfléchi d'individus selon des règles définies. Cette méthode peut être utilisée si on désire tester des questionnaires et aux fins de certaines études préliminaires durant le stade d'élaboration d'une enquête ou lors d'enquêtes sociales où il est impossible ou presque impossible d'effectuer un échantillonnage probabiliste.

Dans le cas de l'échantillonnage non probabiliste, puisqu'on choisit arbitrairement des individus, il n'existe aucune façon d'estimer la probabilité pour une individu quelconque d'être inclus dans l'échantillon. Également, comme la méthode en question ne fournit aucunement l'assurance que chaque individu aura une chance d'être inclus dans l'échantillon, on ne peut estimer la variabilité de l'échantillonnage ni identifier le biais possible. Mais leur utilisation prend peu de temps tout en étant plus économique et plus pratique.

Échantillon de convenance

Les échantillons de convenance sont les échantillons sélectionnés pour des raisons de praticité, d'accessibilité et de coût. Par conséquent celui-ci ne sera pas nécessairement représentatif. Un groupe d'enfants enquêtés dans les centres de santé, aux marchés ou dans d'autres endroits de rassemblement n'est pas un échantillon représentatif. Les enfants qui viennent aux centres de santé ne sont pas les mêmes que ceux qui n'ont pas besoin d'y être présents; pareillement, les enfants en bonne santé peuvent aller au marché, courir autour du village, alors que les enfants malades restent à la maison. La mesure de tels groupes ne devrait jamais s'appeler "enquête" – on l'appelle "prise de mesure en série" et aucune décision ne devrait jamais être basée sur un tel échantillon 'choisi'. De même, les grappes ne peuvent jamais être sélectionnées sur des critères basés sur une plus grande commodité. Si nous devons choisir seulement les ménages dans les villages qui sont près de la route, nous n'obtiendrons jamais un échantillon représentatif. Les villages près des routes ont tendance à avoir un meilleur accès aux transports et peuvent être plus riches que les villages situés à plus grande distance. Dans tous ces exemples, les ménages et les enfants de la population cible n'ont pas une chance égale d'être choisis. Ce n'est pas parce qu'il va au centre de santé, court dans le village, vie près d'une route ou du centre du village qu'un enfant devrait être choisi.

Echantillon boule de neige

Dans ce type d'échantillon, appelé aussi échantillon par réseau, les individus sont sélectionnés en fonction de leurs liens avec un « noyau » d'individus. On l'utilise souvent quand la population cible est difficile à atteindre ou à localiser (usagers de drogues, travailleuse du sexe, etc.). Dans ce cas, les enquêteurs créent un réseau d'informateurs faisant partie de la population d'intérêt, qui les dirigent vers d'autres membres du groupe cible et ainsi de suite, jusqu'à ce que le nombre de répondants souhaités pour l'échantillon soit atteint (on parle de progression en cascade). Cette technique permet de constituer une liste assez importante de personnes correspondant aux critères choisis qui n'étaient pas forcément connues de l'enquêteur, ni directement accessibles. Néanmoins, la boule de neige reste une méthode alternative à n'utiliser que lorsqu'un échantillonnage probabiliste est impossible.

Méthode des quotas

La méthode des quotas vise à étudier un échantillon représentatif de la population mère grâce à une sélection de celui-ci réalisée au préalable à partir de critères tels que le sexe, l'âge, le revenu... Ainsi, les caractéristiques des individus notamment leurs opinions et attitudes seront représentatives de la population mère. On donne à chaque enquêteur un plan de travail qui lui impose le respect de certaines proportions au sein de ses individus enquêtés. Par exemple, dans une population composée à 51% de femmes et à 20% d'habitants de villes de plus de 20'000 habitants et un échantillon de 800 individus: l'enquêteur interrogera $800 \times 0,51 \times 0,2 = 82$ femmes vivant en milieu urbain.

Méthode des itinéraires

La méthode des itinéraires indique aux enquêteurs de façon précise le trajet à suivre dans un secteur donné ainsi que les points de passage où ils devront faire escale afin d'interviewer les individus. Elle évite la tentation d'aller au plus simple en interrogeant trop de répondants au même endroit, ce qui peut être source de biais. Mais elle reste difficile à appliquer, surtout si elle est combinée à une méthode des quotas.

6.3.4. Taille de l'échantillon

Lorsque l'on effectue une enquête on s'intéresse à une population mère (population totale) dont on va généralement interroger une petite partie, c'est l'échantillon dont il faut déterminer la taille soigneusement car elle a une grande importance sur la précision des estimations réalisées sur les caractéristiques de la population-mère.

L'une des questions que l'on doit se poser lorsqu'on organise une enquête ou une étude est la suivante : Quelle doit être la taille de l'échantillon ? La réponse dépend des objectifs, de la nature et de l'ampleur de l'enquête ainsi que des résultats escomptés- tous éléments qu'il convient d'examiner soigneusement au stade de la planification.

Pour des raisons économiques, il est nécessaire d'utiliser une taille d'échantillon la plus réduite possible tout en obtenant un taux de confiance et une marge d'erreur suffisants. Il existe différentes façons de calculer ce nombre, en suivant les manuels de calculs statistiques, les protocoles d'échantillonnage ou en ayant recours à un logiciel d'échantillonnage. Quel que soit le processus choisi, la taille de l'échantillon dépend toujours des objectifs de l'enquête et du degré de précision et de représentativité recherché: plus les questions sont détaillées, plus l'échantillon devra être large. Elle dépend également des contraintes budgétaires et logistiques de l'enquête.

Cas d'un échantillon aléatoire simple

Pour un modèle d'enquête fondé sur un échantillon aléatoire simple, on peut calculer la taille d'échantillon requise en appliquant la formule suivante.

$$N = \frac{Z^2 \times P \times (1 - P)}{d^2}$$

Avec :

- N: la taille de l'échantillon.
- Z: valeur correspondant à un niveau de confiance donné (1,96 pour un niveau de confiance de 95 % qui est la valeur généralement utilisée).
- P: Proportion de l'indicateur principal escompté dans la population, exprimé en décimales(0.5 par défaut).
- D: Précision absolue nécessaire, de part et d'autre de la proportion (0.05 ou 0.10 en général).

En général une estimation sommaire de P suffit. A défaut, on retiendra la valeur 0.5 car c'est la valeur la plus sûre puisque la taille de l'échantillon nécessaire est maximale pour P=0.5. Quand la proportion escomptée est donnée sous forme d'un intervalle, on retiendra la borne de l'intervalle de confiance la plus proche de 0.5. Généralement, on ajoute encore 5% à l'échantillon pour tenir compte d'impondérables comme les non-réponses ou les erreurs d'enregistrement.

Cas d'un échantillonnage en grappe ou à plusieurs degrés

Dans la pratique effective d'une enquête sur le terrain, il est peu probable que la méthode de choix consiste dans un échantillonnage aléatoire simple. Si l'on se sert d'une autre méthode, il faut en principe augmenter la taille de l'échantillon à cause du « biais d'échantillonnage » (biais conceptuel ou design effect). Ceci est dû au fait que les sujets dans une même grappe sont généralement plus semblables entre eux qu'aux membres des autres grappes. Ceci a comme conséquence la diminution de la précision. L'imprécision de l'échantillonnage en

grappes est compensée en prenant la taille calculée pour un échantillonnage aléatoire simple et en le multipliant par un facteur appelé le « design effect ». On peut estimer ce biais à 2. Autrement dit, pour obtenir la même précision, il faut examiner deux fois plus de sujets qu'avec un échantillonnage par sondage aléatoire simple.

$$N = \frac{2 \times Z^2 \times P \times (1 - P)}{d^2}$$

Avec :

- N: la taille de l'échantillon.
- 2: Effet de grappe ou design effect (en général il est de 1.5 ou 2)
- Z: valeur correspondant à un niveau de confiance donné (1,96 pour un niveau de confiance de 95 % qui est la valeur généralement utilisée).
- P: Proportion de l'indicateur principal escompté dans la population, exprimé en décimales(0.5 par défaut).
- D: Précision absolue nécessaire, de part et d'autre de la proportion (0.05 ou 0.10 en général).

Exemple: Nous souhaiterions estimer la proportion de femmes enceintes dans une population qui bénéficiait de soins prénataux au cours du premier trimestre de la grossesse. Combien faut-il inclure de femmes dans l'échantillon pour que la proportion soit connue avec une erreur maximale de ± 5 points de pourcentage par rapport à la valeur exacte, pour un niveau de confiance de 95% et en admettant que la proportion réelle des bénéficiaires de soins prénataux ne dépasse pas 40%.

Première étape : Calcul de la taille de l'échantillon de base

$$N = \frac{1.96^2 \times 0.4 \times (1 - 0.4)}{(0.05)^2} = 368.80 \approx 369$$

Deuxième étape : Effet du plan d'échantillonnage

L'enquête va reposer sur un échantillonnage en grappes (sélection représentative de villages), et non pas sur un échantillon aléatoire simple. Pour corriger la différence, on multiplie la taille de l'échantillon par l'effet du plan d'échantillonnage (design effect).

$$N = 369 \times 2 = 738$$

Troisième étape : Effet des impondérables

On ajoute encore 5% à l'échantillon pour tenir compte d'impondérables comme les non-réponses ou les erreurs d'enregistrement.

$$N = 738 + (738 \times 0.05) = 774.9$$

Quatrième étape : Distribution des sujets observés

Pour conclure, on arrondit le chiffre obtenu au nombre le plus proche du nombre de grappes (30 villages) à étudier.

Taille d'échantillon finale: $N = 780$ femmes

On divise ensuite la taille d'échantillon finale (N) par le nombre de grappes (30) pour déterminer le nombre de sujets à observer par grappe

Nombre de sujets par grappe : $780/30 = 26$ femmes par grappe.

La taille des grappes est choisie sur la base du nombre de ménages à couvrir en un jour qu'il est raisonnable d'attendre d'une équipe; sans prendre de raccourcis et en enquêtant chaque grappe convenablement tout en laissant assez de temps pour retourner aux ménages où les enfants étaient absents. On divise ensuite la taille d'échantillon finale (N) par le nombre de grappes pour déterminer le nombre de sujets à observer par grappe. Par exemple, dans le cas d'une enquête nutritionnelle chez les moins de 5 ans, si on prévoit que les équipes peuvent seulement mesurer 20 enfants par jour, et qu'on ait prévu d'avoir 30 grappes alors la meilleure stratégie est d'augmenter ce nombre. L'effet de grappe est plus petit avec un plus grand nombre de grappes. Ainsi, même si il y a plus de grappes, un nombre d'enfants requis moins important est possible. Par exemple, 40 ou 45 grappes de 20 enfants est plus efficace que 30 grappes de 30 enfants. Puisque le nombre de grappes réduit le design effect rapidement, en dessous d'environ 26 grappes les résultats deviennent incertains. C'est pourquoi, vous devriez toujours projeter d'avoir au moins 30 grappes.

Conversion de la taille de l'échantillon exprimée en nombre de ménages

La taille finale de l'échantillon peut être exprimée de deux façons : soit par un nombre fixe de ménages, soit par quota, c'est à-dire en nombre fixe de personnes. La recommandation générale est d'utiliser la méthode de l'échantillonnage en nombre fixe de ménages comme cible à atteindre pour deux raisons principales :

- Il est recommandé d'utiliser la méthode de l'échantillonnage aléatoire simple ou systématique pour choisir les ménages au sein des grappes (au second degré d'échantillonnage), ces deux méthodes étant plus fiables en termes de représentativité de l'échantillon et introduisent moins de biais. Ces deux méthodes sont basées sur la sélection de ménages soit à partir d'une liste (simple) soit par intervalle d'échantillonnage (systématique). Par conséquent, lorsqu'on utilise l'une de ces techniques d'échantillonnage, il est plus logique d'avoir un nombre fixe de ménages comme cible à atteindre pour chaque grappe. En d'autres termes, puisqu'il n'est possible d'estimer que le nombre approximatif d'enfants admissibles par ménage avant la collecte des données lors d'une enquête nutritionnelle (ce qui ne reflète pas nécessairement le nombre réel d'enfants des ménages sélectionnés), il sera impossible de connaître à l'avance le nombre de ménages à sélectionner et qui comporteront le nombre exact d'enfants.
- Lorsque les équipes doivent atteindre un nombre cible de personnes à enquêter (femmes en âge de procréer, enfants de 6-59 mois, etc.), elles peuvent avoir tendance à sauter les ménages sans enfants par exemple. Or, d'autres indicateurs recueillis au cours de la même enquête (tels que mortalité, eau et hygiène, sécurité alimentaire, etc.) sont mesurés au niveau du ménage. Ne mesurer ces indicateurs que chez les ménages qui ont des enfants, en excluant tous les autres, introduira un biais important.

Pour calculer la taille des ménages à enquêter lors d'une enquête SMART, on utilise la formule suivante :

$$N_{mn} = \frac{N_{enf}}{(Taille_{mn} \times \% \text{ des } <5 \text{ ans} \times 0.9)}$$

Avec :

N_{mn} : taille de l'échantillon en termes de ménages

N_{enf} : taille de l'échantillon en termes d'enfants

$Taille_{mn}$: taille moyenne d'un ménage

% des <5 ans : proportion d'enfants de moins de cinq ans dans la population

0.9 : fraction des enfants de 6-59 mois dans la catégorie des moins de cinq ans.

Exemple : Si le nombre d'enfants requis est de 401, la taille moyenne d'un ménage de 5, le proportion des enfants de moins de cinq ans de 15%, alors la taille de l'échantillon en nombre de ménage sera:

$$N_{mn} = \frac{401}{(5 \times 0.15 \times 0.9)} = 594.07 \sim 595 \text{ ménages}$$

En tenant compte des non répondants (3% des ménages par exemple), on obtient une taille d'échantillon en termes de ménages de **613 ménages à enquêter**.

Exactitude, précision et erreur d'échantillonnage

Quand on sélectionne un échantillon au sein de la population totale, il est convenu que seule une estimation de l'indicateur d'intérêt et non la valeur réelle de la population sera obtenue. Par conséquent, il faut considérer :

- à quel point l'estimation est susceptible de se rapprocher de la valeur réelle de la population (niveau d'exactitude);
- à quel degré les résultats seraient similaires si plusieurs enquêtes étaient menées en utilisant la même méthodologie (niveau de précision)

Exactitude

L'exactitude mesure la validité de l'estimation; un manque d'exactitude est lié à la présence de biais et reflète la différence entre l'estimation issue de l'échantillon et la valeur réelle issue de la population; cela peut être dû à l'erreur de mesure, à la sélection d'un échantillon non représentatif, ou à des facteurs autres que la taille de l'échantillon. **On ne peut pas contrôler les biais une fois qu'ils sont présents**. Il est cependant possible de réduire leur occurrence en standardisant les procédures d'enquête, en fournissant une formation adéquate à chaque membre de l'équipe d'enquête, et en supervisant étroitement tout le travail d'enquête.

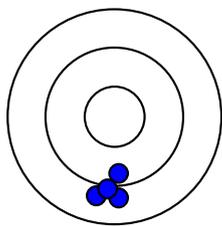
La précision

La précision mesure la consistance (c'est-à-dire la fidélité, ou encore la reproductibilité) des résultats et est reliée à l'erreur (ou fluctuation) d'échantillonnage. L'erreur d'échantillonnage

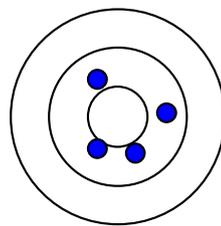
est la différence entre l'estimation issue de l'échantillon et la valeur obtenue pour la population, due au fait de mesurer un échantillon plutôt que l'ensemble de cette population; on peut contrôler l'erreur d'échantillonnage en manipulant la taille de l'échantillon. En substance, l'erreur d'échantillonnage est la mesure de l'incertitude due au fait que notre estimation provient de la mesure d'une partie relativement faible de la population-cible et ne pourra donc jamais correspondre exactement à la valeur réelle qui aurait été obtenue pour la population, même si l'échantillon est sélectionné aléatoirement et de manière représentative. Plus la proportion de la population-cible mesurée est grande, plus cette incertitude s'amenuise. On peut donc dire que plus la taille de l'échantillon est grande, plus la précision des mesures augmente. Une taille d'échantillon plus grande accroît la précision des résultats, mais ne garantit pas l'absence de biais. Lorsque la taille de l'échantillon est très importante, contrôler la qualité devient plus difficile à effectuer en raison du grand nombre d'équipes à former et superviser, ce qui augmente le risque d'introduire des biais.

Prévenir la présence de biais est essentiel; ne tentez jamais d'atteindre une précision plus élevée sans tenir compte du risque d'introduire des biais. Il vaut mieux avoir un échantillon plus petit et donc des résultats moins précis, mais aussi moins biaisés.

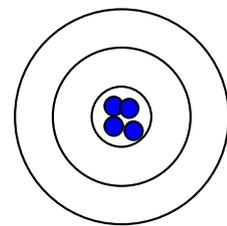
Figure 4 : Exactitude et précision



Des résultats précis mais
inexact.



Des résultats imprécis mais
exact.



Des résultats à la fois précis
et exacts.

Erreur d'échantillonnage

Comme l'échantillon est toujours basé sur une sélection, les résultats statistiques pourront au mieux approcher la réalité. Les statisticiens calculent donc ce qu'on appelle l'erreur d'échantillonnage (p. ex. un «écart-type» ou un «intervalle de confiance»). Ils rendent ainsi compte de la précision de la méthode d'échantillonnage utilisée. Par convention, un intervalle de confiance (IC) de 95% est choisi. Cela signifie qu'en l'absence de biais, une fois l'estimation obtenue, vous pouvez être sûr à 95% (niveau de confiance) que la valeur réelle de la population de votre indicateur se situe dans les limites de l'intervalle de confiance calculé. Plus l'échantillon est représentatif et les mesures exactes, plus vous pouvez être certain de l'absence de biais, et du fait que la vraie valeur de votre indicateur pour la population se situe à l'intérieur de l'intervalle de confiance.

La gestion, l'organisation, la logistique et le travail de terrain

Analyse, interprétation et dissémination des résultats

Une fois la phase de terrain terminée et la totalité des données recueillies, commence l'analyse de celles-ci (parfois, surtout si l'enquête doit durer longtemps, on procède à certaines analyses avant que l'enquête soit terminée).

Ordinateur et logiciels spécialisés

Il n'est plus concevable aujourd'hui d'analyser une enquête sans recours à l'ordinateur et aux logiciels qui donnent accès à des bibliothèques de programmes spécialisés pour la statistique et l'épidémiologie.

Analyses préliminaires

Avant de procéder aux calculs visant à répondre aux questions posées dans le protocole, certaines analyses préliminaires doivent toujours être effectuées.

Etude de la validité des données

Tout d'abord, il est fondamental d'étudier la validité des données recueillies : recherche des données manquantes ou mal remplies dans les questionnaires, des sujets initialement sélectionnés et n'étant pas entrés dans l'étude (ou perdus de vue), étude de la variabilité des données en fonction des différents enquêteurs et/ou du statut des sujets (cas ou témoins, âge des sujets, etc.). Une telle analyse peut permettre de détecter certains biais, et si possible de les corriger, ou du moins d'en tenir compte dans l'analyse et l'interprétation définitive des résultats.

Dépouillement à plat

Ensuite, il est toujours nécessaire de procéder à un premier dépouillement à plat de l'enquête : on étudie isolément la répartition de chaque variable (en cherchant notamment des valeurs aberrantes qui peuvent provenir d'erreurs), on trace des histogrammes et des courbes simples, on fait des tableaux croisés (ventilation des valeurs d'une variable en fonction de celles d'une ou de plusieurs autres), etc. De telles analyses préalables où chaque variable est considérée isolément, ou n'impliquant que deux ou trois variables simultanément, permettent d'avoir une bonne connaissance des données, de détecter d'éventuelles erreurs, de décrire les caractéristiques essentielles de l'enquête (nombre et répartition des sujets en fonction de certaines caractéristiques : âge, sexe, etc.), et de choisir au mieux les méthodes d'analyse ultérieures.

Analyse multidimensionnelle

Enfin, on peut procéder aux analyses statistiques destinées à répondre aux questions de l'enquête. Parfois, les méthodes sont simples et peuvent pratiquement faire l'objet de calculs manuels ; mais, la plupart du temps, on est amené à réaliser des calculs plus ou moins complexes, essentiellement du fait qu'il faut tenir compte simultanément de plusieurs variables (standardisation sur plusieurs facteurs simultanément ; calcul d'un risque relatif en prenant en compte plusieurs facteurs de confusion : âge, consommation de tabac et d'alcool, etc.) : c'est le domaine de la statistique multidimensionnelle.

Présentation des résultats

Ce paragraphe, qui rappelle les principales informations qu'il faut fournir lorsqu'on présente les résultats d'une enquête concerne également le lecteur d'une publication qui peut s'attendre à y trouver les éléments qui lui permettront un jugement critique.

Informations à fournir

La présentation des résultats d'une enquête peut se faire de différentes façons (non exclusives) : rédaction d'articles scientifiques, de rapports de recherche, de communications orales ou des posters à des colloques... Mais quelle que soit la forme matérielle de la présentation, il est nécessaire de toujours fournir certaines informations qui permettront au lecteur (ou à l'auditeur) de porter un jugement critique sur le travail effectué et les conclusions qui peuvent en être tirées (voir annexe).