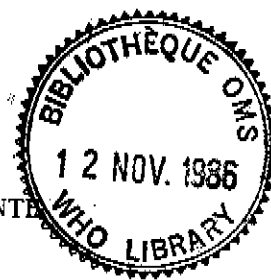


E 1906

45075



WORLD HEALTH ORGANIZATION
ORGANISATION MONDIALE DE LA SANTE



WHO/CDS/LEP/86.1 ✓

ORIGINAL : ANGLAIS

LES SONDAGES DANS LA LUTTE CONTRE LA LEPRE - MANUEL PREPARATOIRE

par

T. K. Sundaresan, Consultant de l'OMS en collaboration avec le Dr H. Sansarriq (ancien Médecin-chef, unité de la Lèpre), et du Dr S. K. Noordeen, Médecin-chef, unité de la Lèpre, Division des Maladies transmissibles, OMS, Genève

TABLE DES MATIERES

	<u>Pages</u>
1. Introduction	1
2. Raisons justifiant les sondages	3
3. Quelques considérations générales sur le choix d'une méthode de sondage	3
4. Techniques de sondage	4
5. Certains concepts essentiels d'échantillonnage	6
6. Emploi d'une table de nombres au hasard et exemples pratiques de sélection d'échantillons	10
7. Expérience provenant des sondages LAT (sondages effectués par l'équipe consultative de la Lèpre de l'OMS)	17
8. Quelques considérations pratiques importantes	18
9. Organisation d'un sondage	20
10. Procédures et critères de diagnostic	23
11. Erreurs non dues à l'échantillonnage	26
12. Sondages périodiques pour évaluation	28
ANNEXE : Formules pour les erreurs types - Sondage en grappes	30

1. INTRODUCTION

Le texte suivant est un extrait des "Directives concernant la lutte contre la lèpre" (OMS, 1980)¹ :

"Les informations concernant la prévalence de la lèpre dans la plupart des pays ne donnent pas une image fidèle de la situation car ni le dépistage ni la déclaration des cas ne se font toujours de façon satisfaisante ... Les sondages aléatoires de l'équipe consultative OMS de la lèpre montrent que même dans les pays où le programme de dépistage est jugé satisfaisant, on peut déceler un nombre de cas nouveaux atteignant les trois quarts du nombre des cas enregistrés ...".

The issue of this document does not constitute formal publication. It should not be reviewed, abstracted, quoted or translated without the agreement of the World Health Organization. Authors alone are responsible for views expressed in signed articles.

Ce document ne constitue pas une publication. Il ne doit faire l'objet d'aucun compte rendu ou résumé ni d'aucune citation ou traduction sans l'autorisation de l'Organisation mondiale de la Santé. Les opinions exprimées dans les articles signés n'engagent que leurs auteurs.

De nombreux pays ont accordé une forte priorité à la lutte contre la lèpre. Il est évident que : 1) des informations fiables doivent être obtenues sur l'étendue de la maladie, en tenant compte de ses manifestations les plus fréquentes, par exemple formes infectieuses, invalidités, etc., pour qu'il soit possible de formuler un plan national de lutte et d'allouer les ressources de façon appropriée, et 2) un système d'information doit être créé, permettant l'évaluation et la surveillance périodiques de tout programme d'intervention pouvant être mis en oeuvre.

Le Comité OMS d'experts de la lèpre a conclu dans son cinquième rapport (Séries de Rapports techniques, N° 607, 1977)² que :

"L'incidence, ou nombre de cas nouveaux se produisant au cours d'une période (généralement un an) par rapport au chiffre de la population, est le seul indice permettant d'évaluer l'efficacité des mesures prises, c'est-à-dire la réduction de la transmission.

L'apparition continue d'infections nouvelles chez les enfants a une importante signification épidémiologique, car elle indique la persistance de la transmission. Il est donc utile d'établir des taux d'incidence distincts pour les enfants et pour les adultes.

Les variations de la prévalence ne reflètent qu'indirectement les effets des programmes sur la situation épidémiologique. Il faut toutefois connaître le nombre total des malades à traiter pour pouvoir planifier et organiser les activités."

Dans de nombreux programmes de lutte, il est devenu nécessaire d'évaluer la situation épidémiologique existante. Le Comité a recommandé qu'une telle évaluation soit entreprise au moyen de sondages permettant de connaître le nombre de cas de lèpre non encore détectés.

Le Comité a recommandé par ailleurs que les renseignements obtenus soient simples et permettent de prendre des décisions; ils devraient pouvoir être recueillis directement par le personnel responsable des activités sanitaires au niveau périphérique, qui n'est pas forcément spécialisé en matière de lèpre.

Il est très tentant de procéder à des investigations par sondage car elles permettent de réaliser des économies et d'obtenir plus rapidement des informations. Cependant, avant d'entreprendre un sondage, il est important de bien en connaître les principes de base et de savoir comment en interpréter les résultats.

Le but de ce manuel est de familiariser les agents de la lutte contre la lèpre à ces principes et de les aider à entreprendre aisément des sondages dans des régions limitées :

- i) pour mesurer la prévalence de la lèpre dans ces régions, le nombre de cas et la proportion des différents types de lèpre;
- ii) pour évaluer l'impact d'un programme en terme de prévalence et
- iii) pour évaluer certains aspects opérationnels des programmes de lutte antiléproux.

Pour toute opération assez importante, comme une enquête nationale, il est préférable de s'adjoindre un statisticien, car il peut être alors opportun d'appliquer des techniques statistiques plus élaborées, et le coût d'un statisticien consultant sera plus que couvert par l'ensemble des économies réalisées.

Les sondages destinés à rechercher le nombre de cas ne sont indiqués que lorsque les taux de prévalence sont au moins de l'ordre de 10 pour 1000. Si les taux attendus sont plus faibles, la taille de l'échantillon devrait être assez grande et des méthodes de surveillance épidémiologique seraient alors plus appropriées. Du fait que, dans tout pays, la lèpre ne sévit pas de façon uniforme, il est possible que la prévalence soit limitée à des régions de forte endémicité. Alors qu'en général, les sondages indiquent les limites à l'intérieur desquelles la prévalence vraie dans la population devrait se situer, des sondages peuvent être parfois utiles quand la prévalence est faible pour savoir quelle peut être la limite supérieure de cette valeur.

Si dans ce manuel, les exemples donnés concernent surtout les estimations de prévalence, les principes généraux de sondage sont applicables de façon générale.

2. RAISONS JUSTIFIANT LES SONDAGES

Quand, dans certains pays, on ne possède pas d'informations sur l'étendue et l'ampleur du problème de la lèpre mais qu'on a de bonnes raisons de croire qu'en fait, cette ampleur est considérable, il peut être extrêmement utile d'effectuer des sondages pour l'évaluer avec plus de précision. De telles informations sont nécessaires dans des buts de planification, et l'examen complet de l'ensemble d'une population doit être écarté car il est impraticable et beaucoup trop coûteux.

Mais il y a une autre raison, peut-être plus importante, de s'adresser aux sondages : pour surveiller les progrès et l'impact d'un programme de lutte en cours. Des sondages périodiques, effectués dans une région où des opérations de lutte ont lieu, permettront, à relativement peu de frais, de mesurer les progrès accomplis, de découvrir les imperfections opérationnelles et, si nécessaire, de s'orienter vers des stratégies de remplacement, d'une meilleure efficacité.

Des sondages à l'échelon national sont utiles pour fournir aux planificateurs les informations nécessaires à la répartition appropriée des ressources. Les sondages périodiques donneront des indications utiles sur la surveillance du programme. Dans les deux cas, un sondage n'est pas un but en lui-même mais seulement un moyen d'obtenir des informations essentielles à certains objectifs. De ce fait, le coût des sondages, et les dépenses qu'ils entraînent, doivent être réduits à l'essentiel, et ne représenter qu'une petite fraction du budget de la lèpre.

3. QUELQUES CONSIDERATIONS GENERALES SUR LE CHOIX D'UNE METHODE DE SONDAGE

Les investigations par sondage donnent les meilleures estimations des valeurs vraies des variables ou caractéristiques (par exemple taux de prévalence) de l'ensemble de la population d'où les échantillons sont tirés. Il ne s'agira cependant que d'"estimations" qui ne correspondront que rarement à la valeur "vraie", inconnue, dans la population. Il est néanmoins possible de dire, avec un niveau élevé de confiance, que la valeur vraie de la population se situe dans certaines limites. Par exemple, si le sondage montre que 2 % de sujets ont la maladie, nous pouvons dire (avec 95 % de confiance) suivant la méthode employée, que la prévalence vraie dans la population se situe entre 1,6 % et 2,4 %. En augmentant la taille de l'échantillon, l'écart (par exemple entre 1,6 et 2,4) peut être réduit autant qu'on le désire. Techniquement, cet écart est connu comme étant "l'intervalle de confiance" et reflète la "précision" de l'estimation obtenue par le sondage. Ainsi dans l'exemple ci-dessus, si l'échantillon avait été très grand, nous aurions pu dire que la valeur vraie de la population se situait entre 1,8 et 2,2. Dans ce cas, la seconde estimation obtenue avec un échantillon plus grand, est plus précise que la première.

Un niveau élevé de précision, souhaitable pour des raisons scientifiques, peut par ailleurs être inutilement coûteux. Par exemple, pour doubler la précision, il peut être nécessaire de quadrupler la taille de l'échantillon. Ainsi, il est important d'envisager soigneusement, et très à l'avance, l'emploi qui doit être fait du résultat du sondage et de décider d'une précision minimum acceptable.

Etroitement liée aux faits mentionnés ci-dessus est la nécessité de fixer les objectifs du sondage de la façon la plus détaillée possible, en spécifiant les catégories de personnes à examiner.

Pour être plus précis, disons qu'il n'existe que peu d'informations sur la prévalence de la maladie et que les planificateurs ont besoin d'une estimation du nombre de cas à l'échelon national pour décider des affectations budgétaires nécessaires. Très souvent, on se trouve dans une situation où les ressources sont rares et il est bon d'avoir une idée de l'ampleur du problème; par exemple, la prévalence est-elle plus proche de 20 pour 1000 ou de 10 pour 1000. Dans ce cas, un sondage rapide portant sur un échantillon de taille modérée peut fournir l'information nécessaire. D'autre part, dans des sondages du type "avant" et "après", qui doivent permettre de déceler les améliorations très nettes résultant des mesures d'intervention, différents niveaux de précision peuvent dicter la taille de l'échantillon et le choix de la méthode.

La lèpre se présente sous de nombreux aspects qui, à des degrés différents, attirent l'attention du public et stimulent l'action gouvernementale. Il existe des invalidités dues à la maladie, des cas actifs avec des lésions manifestes et, à l'opposé, de petites taches presque invisibles qui peuvent n'être décelées qu'après examen complet. Si des informations sur ces différents aspects peuvent être obtenues d'un sondage, le niveau de précision exigé pour chacune de ces variables peut être différent, les programmes d'action correspondants pouvant avoir des implications budgétaires extrêmement variées.

La variation des valeurs dans l'échantillon par rapport à la valeur vraie dans la population est appelée "erreur d'échantillonnage" et on peut y remédier en choisissant une méthode de sondage efficace. Par ailleurs, un certain nombre d'autres facteurs peuvent donner lieu à des divergences par rapport à la valeur vraie de la population. Par exemple, la non-couverture d'une partie de la population de l'échantillon sélectionné, l'examen incomplet des individus, des erreurs de diagnostic, etc. peuvent tous contribuer à des erreurs d'estimation. On les désigne comme des erreurs non dues à l'échantillonnage, et en raison de leur importance, particulièrement dans les sondages sur la lèpre, ils sont traités dans une section spéciale.

A cet égard, une distinction doit être faite entre "précision" et "exactitude". Le concept de "précision" est traité dans la section 5. Il faut dire ici que la précision se rapporte à la contiguïté entre la valeur obtenue de l'échantillon et la valeur vraie de la population, et qu'on peut en général résoudre ce problème en choisissant un échantillon de taille voulue. L'exactitude, au contraire, tient compte également des erreurs non dues à l'échantillonnage. Elle dépend à la fois des erreurs dues et non dues à l'échantillonnage. Ce qui importe avant tout c'est que l'échantillon soit vraiment représentatif de la population dont il est tiré. Toute personne appartenant à l'échantillon doit avoir la possibilité d'être représentée dans l'échantillon.

4. TECHNIQUES DE SONDRAGE

Il existe actuellement diverses techniques de sondage dont l'objectif principal est d'offrir un maximum d'efficacité et de facilité opérationnelle. Il peut être utile de mentionner ici quelques-unes de ces techniques particulièrement applicables aux sondages dans la lèpre.

Un des préalables à une investigation par sondage est l'existence ou la construction d'une "base de sondage". Des exemples en sont : des listes de recensement, des listes de ménages, des listes de villages, avec la population dans chacun d'eux. Habituellement, ces listes ne sont pas à jour, mais, si elles ne sont pas trop anciennes, elles peuvent encore être utilisées, au moins dans un premier stade du choix d'un sondage. Ce problème est traité de façon plus détaillée à la section 9 sous "Organisation d'un sondage". Il est mentionné ici parce que la construction d'une base satisfaisante est la première étape du processus d'échantillonnage et que la nature de la base influera les considérations portant sur le choix de la technique la plus appropriée.

4.1 Sondage aléatoire élémentaire

Ceci est le nom donné à une méthode "dans laquelle les unités d'échantillonnage sont sélectionnées indépendamment avec une probabilité égale, l'échantillonnage se faisant sans remplacement et sans restriction".³ Ainsi, si la population est de un million d'habitants et que l'on désire examiner un échantillon représentatif de 10 000, cette méthode permet de s'assurer que chacun des millions d'habitants a la même chance de devenir membre de l'échantillon sélectionné. En pratique, cela supposerait d'établir d'abord la liste du million d'individus de la population et de choisir 10 000 d'entre eux au hasard - en adoptant des méthodes utilisées pour les loteries ou par l'emploi de nombres d'échantillonnages au hasard (voir section 6). Mais cette procédure est extrêmement lourde et irréalisable en pratique dans des sondages de population.

4.2 Sondage stratifié (voir également section 8)

Parfois, il faut obtenir des informations sur la prévalence ou l'incidence de la maladie dans différentes portions ou "strates" de la population. Ceci peut être nécessaire dans des buts administratifs ou pour une meilleure compréhension de l'épidémiologie de la maladie. Les caractéristiques (telles que population urbaine ou rurale, taux d'endémicité, régions montagneuses, groupes spéciaux de population, etc.) sur lesquelles on désire établir la stratification doivent d'abord être déterminées et la population de chaque strate définie. De chaque strate, un échantillon séparé est tiré. Le pourcentage de la population qui constitue l'échantillon n'est pas obligatoirement le même d'une strate à l'autre. On peut également recourir à la stratification pour augmenter l'efficacité d'une procédure de sondage. Par exemple, si la population est stratifiée d'après les différents niveaux d'endémicité (connus ou soupçonnés), il est avantageux de choisir différentes tailles d'échantillons dans les différentes strates et d'obtenir une estimation pour l'ensemble de la population en combinant les estimations de chacune des strates après les avoir pondérées de façon acceptable.

4.3 Sondage en grappes

Tout groupe d'individus, par exemple les membres d'un foyer, les habitants d'un village, etc., peuvent constituer une grappe. Lorsqu'une grappe est définie, par exemple un foyer, on peut alors sélectionner un échantillon de foyers à partir d'une liste complète de ceux d'une communauté et tous les individus appartenant aux foyers sélectionnés sont examinés. Si le village est défini comme étant une grappe, on peut alors sélectionner un échantillon de villages à partir d'une liste des villages et toute la population du village sélectionné sera examinée. Quand les grappes sont "naturellement" définies, par exemple foyers, villages, etc., la taille des grappes, c'est-à-dire le nombre d'individus dans chacune des grappes, ne sera normalement pas égale. Cependant, pour des raisons opérationnelles et statistiques, il peut être utile de créer des grappes de taille à peu près semblable par groupement, par exemple de petits villages peuvent être groupés pour constituer une grappe d'une population totale égale à celle d'un gros village.

Quand une maladie transmissible comme la lèpre est étudiée, il est vraisemblable que les individus d'une grappe seront plus semblables entre eux, en ce qui concerne le risque de la maladie, que les individus d'une autre grappe. Une telle similitude à l'intérieur des grappes est appelée la "corrélation intra-grappe". Quand cette corrélation existe, les estimations provenant de l'échantillon auront une variation plus grande et de ce fait seront moins précises (voir le chapitre suivant sur "certains concepts essentiels du sondage"). La taille de l'échantillon, calculée théoriquement sur la base d'un simple échantillon aléatoire d'individus, devra être augmentée pour tenir compte de cette corrélation intra-grappe.

Le sondage en grappes est le plus largement utilisé pour estimer certaines caractéristiques de la maladie, comme la prévalence. Toute augmentation éventuellement nécessaire de la taille de l'échantillon est généralement contre-balançée par des inconvénients opérationnels. De toute façon, en ce qui concerne les sondages sur la lèpre, la sélection aléatoire d'individus à partir d'une population tout entière peut être considérée comme impraticable.

4.4 Sondage à plusieurs degrés

Quand le sondage porte sur une vaste population, comme celle du pays, le sondage est généralement effectué par degrés. Par exemple, si le pays est divisé en districts, un échantillon de districts peut d'abord être sélectionné. Puis de chaque district sélectionné, on peut sélectionner un échantillon de sous-district. A partir de chaque sous-district, on peut sélectionner un échantillon de villages. Une telle méthode est appelée sondage à plusieurs degrés, la sélection de districts, de sous-districts, de villages représentant les différents degrés. A chaque degré, la probabilité de sélection d'une sous-unité peut être rendue proportionnelle à sa population; ainsi, un sous-district assez vaste aura une chance relativement plus grande d'être sélectionné qu'un plus petit.

Il faut se rappeler que les résultats de l'analyse (c'est-à-dire les inférences du sondage) ne s'appliquent dans tous les cas qu'à la collectivité d'où l'échantillon a été tiré. Par exemple, si l'on décide d'exclure de l'échantillon certains groupes de population, comme les nomades, les résultats du sondage ne sont applicables qu'à l'ensemble de la population, à l'exclusion des nomades.

5. CERTAINS CONCEPTS ESSENTIELS D'ECHANTILLONNAGE

5.1 L'erreur type

Aussi complète que soit la couverture, aussi représentatif que soit l'échantillon et aussi parfaits que soient les outils diagnostiques, il reste que lors d'investigations par sondages, on examine une partie de la population et que le résultat obtenu (tel que le taux de prévalence) est rarement identique à celui que l'on obtiendrait en examinant la population tout entière. Par ailleurs, si l'on prend un certain nombre d'échantillons, leurs valeurs ont tendance à se grouper autour de la valeur vraie de la population, parfois un peu au-dessous, parfois un peu au-dessus. Cette variation par rapport à la valeur vraie de la population est mesurable et elle s'exprime par l'erreur type. Les formules suivantes s'appliquent au sondage aléatoire élémentaire et sont données en raison de leur simplicité. Des formules pour le sondage en grappes sont données dans l'annexe. Si :

- 1) n est la taille de l'échantillon;
- 2) $p\%$ est le taux observé de prévalence de la lèpre;
- 3) $q\%$ ($= 100 - p$) est le pourcentage d'individus sans lèpre;

l'erreur type (e.t) de p est alors $\sqrt{\frac{pxq}{n}}$ *

Les principes statistiques nous apprennent également que la prévalence vraie dans la population ($p\%$) a toutes les chances d'être $p \pm 1,96$ (e.t). La probabilité que la prévalence vraie se situe en dehors de ces limites est de 5 %.

Quand nous disons "toutes les chances", cela signifie que si nous généralisons d'un échantillon à la population comme ci-dessus, c'est-à-dire quand nous disons que la prévalence vraie est de $p \pm 1,96$ e.t, nous aurons finalement raison dans 95 % des cas. En pratique, nous pouvons dire qu'il y a 95 % de certitude que la prévalence vraie dans la population soit $p \pm 1,96$ e.t. Pour cette raison, les deux valeurs $p + 1,96$ e.t et $p - 1,96$ e.t sont les "limites de confiance" à 95 %. L'intervalle de confiance peut être large ou étroit suivant la valeur de l'erreur type, qui détermine donc la précision de l'estimation.

* La formule exacte est $e.t = \sqrt{1-f} \sqrt{\frac{pq}{n}}$ où f = fraction de l'échantillon qui est n/N où n est la taille de l'échantillon et N la population. Cependant, le plus souvent, n/N est petit et le facteur de multiplication $\sqrt{1-f}$ sera presque égal à 1 et peut être négligé.

A titre d'exemple, supposons qu'il y ait deux échantillons de taille différente, donnant tous deux une estimation de 2 % de prévalence. Dans un cas, nous pouvons dire que la prévalence vraie se situe entre 2,5 % et 1,5 % et, dans l'autre cas, l'échantillon le plus grand, nous pourrions être à même de dire que la prévalence vraie est entre 2,2 % et 1,8 %. Dans ce dernier cas, l'estimation est dite être "plus précise" que la première.

Manifestement, une estimation qui donne un très large intervalle de confiance n'a guère d'utilité. Mais d'autre part, si l'on veut que l'intervalle soit très étroit, c'est-à-dire si l'on veut une estimation extrêmement précise, cela peut entraîner une augmentation énorme de la taille de l'échantillon et de ce fait être très coûteux. Il faut donc trouver un certain compromis et rechercher, a priori, un niveau acceptable de précision. Ceci sera davantage discuté dans la prochaine section sur la taille de l'échantillon.

5.2 Le coefficient de variation

De ce qui précède, on peut voir que la variable déterminant la précision est l'erreur type. Maintenant, nous pouvons avoir deux sondages donnant des taux différents de prévalence mais avec à peu près la même erreur type. Par exemple, un sondage peut donner un taux de prévalence de 2 % avec une erreur type de 0,2 % et un autre, un taux de prévalence de 4 % mais également avec la même erreur type, c'est-à-dire 0,2 %. Les intervalles de confiance à 95 % dans les deux cas seront de 1,6 à 2,4, et de 3,6 à 4,4. Il est évident que dans le second cas, l'estimation paraît être plus précise parce que dans le premier les limites de confiance sont $2\% \pm 20\%$ de 2, et dans le second cas ils sont de $4\% \pm 10\%$ de 4. Ainsi, avec des taux variables de prévalence il est utile de songer à un indice qui lie l'erreur type à l'estimation de la prévalence. Un tel indice est fourni par le coefficient de variation, qui se définit comme le rapport entre l'erreur type et l'estimation, exprimé en pourcentage.

5.3 Coefficient de variation (c.v) = $\frac{e.t}{p} \times 100$

Ainsi, dans l'exemple ci-dessus, la prévalence de 2 % est associée à un coefficient de variation de $\frac{0,2}{2} \times 100$ ou 10 %, et dans le second cas la prévalence de 4 % est associée à un coefficient de variation de $\frac{0,2}{4} \times 100$ ou 5 %.

5.4 Taille de l'échantillon

Quand un sondage est planifié, la première question doit être : "Quelle taille doit avoir l'échantillon ?". Pour répondre à cette question de façon satisfaisante, il faut posséder des informations sur les points suivants :

1. Quelle est l'ampleur attendue de la variable qui va être estimée ? Par exemple, s'il s'agit d'un sondage pour estimer la prévalence, nous devons avoir une idée de l'ampleur de cette prévalence; est-elle d'environ 1 à 2 %, ou de 10 à 15 %, etc. A ce stade, seul une série approximative de valeurs est nécessaire, pour servir de point de départ. Généralement, cette information peut être basée sur des dossiers antérieurs, sur des appréciations bien fondées, etc.
2. Le deuxième point à être précisé est l'ampleur des variations tolérables. Par exemple, est-il suffisant d'avoir un coefficient de variation de 20 % ou doit-il être plus faible, etc. Pour donner une réponse valable à cette question, il est important d'avoir à l'esprit l'emploi qui sera fait du résultat du sondage. Pour planifier des stratégies sommaires à l'échelon national, un degré élevé de précision peut ne pas être essentiel alors que pour évaluer l'impact d'un programme sur de petites communautés, c'est l'inverse qui peut être vrai. Il faut se rappeler que le premier objectif d'un sondage est d'obtenir une estimation suffisamment proche de la valeur vraie. Or il ne s'agit pas d'un exercice de recherche de cas. Exiger un degré élevé de précision peut nécessiter un échantillon énorme et les dépenses inhérentes peuvent ne pas être justifiées par le but du sondage lui-même.
3. Le troisième point à envisager est la technique des sondages réalisables sur le plan opérationnel. Par exemple, le sondage en grappes l'est-il ? Si oui, quelle taille doivent avoir les grappes ? Une stratification est-elle indiquée pour des raisons administratives ou d'efficacité statistique, etc.

Bien que le sondage aléatoire élémentaire soit rarement utilisé dans des enquêtes de population sur la lèpre, il reste la base de la plupart des autres techniques de sondage. En pratique, il est habituel de mesurer la taille de l'échantillon en supposant un sondage aléatoire élémentaire et de la multiplier ensuite par un facteur approprié.

Dans le cas d'un sondage aléatoire élémentaire, la formule générale pour déterminer la taille n de l'échantillon est :

$$n = (z^2 pq) / d^2 \text{ où :}$$

p = le pourcentage attendu de la maladie. Comme on l'a dit plus haut, c'est un chiffre approximatif basé sur des expériences précédentes. Si celles-ci laissent envisager que la prévalence se situe entre deux valeurs, disons 2 % et 4 %, la valeur la plus basse doit être retenue. Ceci, parce que la taille d'un échantillon déterminée d'après la valeur la plus basse sera suffisamment grande pour couvrir des estimations qui seraient éventuellement plus élevées, mais que d'autre part, si la valeur la plus élevée est choisie, et si le résultat réel du sondage donne un chiffre inférieur, l'estimation ne sera pas assez précise.

$q = 100 - p$ (c'est-à-dire le pourcentage d'individus sans la maladie)

z = une valeur qui pourrait être obtenue à partir de tables statistiques pour correspondre au niveau de probabilité des limites de confiance (95 %, 99 %, etc. - voir section 5). Dans la plupart des cas, le niveau de probabilité à 95 % est suffisant et la valeur correspondante de z est 1,96.

d = l'expression de la précision exigée. C'est la différence entre l'estimation réelle et sa limite de confiance (ou demi-intervalle de confiance). En pratique, il est habituel de spécifier que cette différence ne devrait pas dépasser de x % (disons 20 %) l'estimation de p ; dans ce cas, d sera pris comme $xp/100$ (0,2 p dans l'exemple).

L'exemple suivant peut rendre l'emploi de la formule plus clair.

Supposons que l'on désire estimer la prévalence de toutes les formes de lèpre dans une communauté; à partir des expériences précédentes, on pense que la prévalence est au moins de 2 % mais pas plus de 4 %; de plus l'estimation est demandée avec une précision qui devrait conduire à une limite de confiance de 95 %, sans dépasser de 20 % la valeur de toute prévalence pouvant être obtenue quand le sondage est effectué; ceci signifie que si le sondage donne, quand il est effectué, une estimation de 3 %, la limite de confiance devrait se situer entre $3 \pm 20\%$ de 3 (=0,6) - c'est-à-dire que nous devrions pouvoir dire que la prévalence vraie dans la population se situe entre 2,4 % et 3,6 %.

Pour décider de la taille de l'échantillon, nous prendrons la plus faible des deux estimations soupçonnées (c'est-à-dire 2 %). La précision spécifiée devrait conduire dans ce cas à une valeur de $d = 20\%$ de 2 (c'est-à-dire 0,4).

En substituant ces valeurs à la formule :

$$\begin{aligned} n &= 1,96^2 \cdot 2 \cdot (100 - 2) / 0,4^2 \\ &= 4705,96 \text{ ou } 4706 \text{ individus.} \end{aligned}$$

TABLE XXXIII. NOMBRES AU HASARD (IV)⁴

10 27 53 96 23	71 50 54 36 23	54 31 04 82 98	04 14 12 15 09	26 78 25 47 47
28 41 50 61 83	64 85 27 20 18	83 36 36 05 56	39 71 65 09 62	94 76 62 11 89
34 21 42 57 02	59 19 18 97 48	80 30 03 30 98	05 24 67 70 07	84 97 50 87 46
61 81 77 23 23	82 82 11 54 08	53 28 70 58 96	44 07 39 55 43	42 34 43 39 28
61 15 18 13 54	16 86 20 26 88	90 74 80 55 09	14 53 90 51 17	52 01 63 01 59
91 76 21 64 64	44 91 13 32 97	75 31 62 66 54	84 80 32 75 77	56 08 25 70 29
00 97 79 08 06	37 30 28 59 85	53 56 68 53 40	01 74 39 59 73	30 19 99 85 48
36 46 18 34 94	75 20 80 27 77	78 91 69 16 00	08 43 18 73 68	67 69 61 34 25
88 98 99 60 50	65 95 79 42 94	93 62 40 89 96	43 56 47 71 66	46 76 29 67 02
04 37 59 87 21	05 02 03 24 17	47 97 81 56 51	92 34 86 01 82	55 51 33 12 91
63 62 06 34 41	94 21 78 55 09	72 76 43 16 94	29 95 81 83 83	79 88 01 97 30
78 47 23 53 90	34 41 92 45 71	09 23 70 70 07	12 38 92 79 43	14 85 11 47 23
87 68 62 15 43	53 14 36 59 25	54 47 33 70 15	59 24 48 40 35	50 03 42 99 36
47 60 92 10 77	88 59 53 11 52	66 25 69 07 04	48 68 64 71 06	61 65 70 22 12
56 88 87 59 41	65 28 04 67 53	95 79 88 37 31	50 41 06 94 76	81 83 17 16 33
02 57 45 86 67	73 43 07 34 48	44 26 87 93 29	77 09 61 67 84	06 69 44 77 75
31 54 14 13 17	48 62 11 90 60	68 12 93 64 28	46 24 79 16 76	14 60 25 51 01
28 50 16 43 36	28 97 85 58 99	67 22 52 76 23	24 70 36 54 54	59 28 61 71 96
63 29 62 66 50	02 63 45 52 38	67 63 47 54 75	83 24 78 43 20	92 63 13 47 48
45 65 58 26 51	76 96 59 38 72	86 57 45 71 46	44 67 76 14 55	44 88 01 62 12
39 65 36 63 70	77 45 85 50 51	74 13 39 35 22	30 53 36 02 95	49 34 88 73 61
73 71 98 16 04	29 18 94 51 23	76 51 94 84 86	79 93 96 38 63	08 58 25 58 94
72 20 56 20 11	72 65 71 08 86	79 57 95 13 91	97 48 72 66 48	09 71 17 24 89
75 17 26 99 76	89 37 20 70 01	77 31 61 95 46	26 97 05 73 51	53 33 18 72 87
37 48 60 82 29	81 30 15 39 14	48 38 75 93 29	06 87 37 78 48	45 56 00 84 47
68 08 02 80 72	83 71 46 30 49	89 17 95 88 29	02 39 56 03 46	97 74 06 56 17
14 23 98 61 67	70 52 85 01 50	01 84 02 78 43	10 62 98 19 41	18 83 99 47 99
49 08 96 21 44	25 27 99 41 28	07 41 08 34 66	19 42 74 39 91	41 96 53 78 72
78 37 06 08 43	63 61 62 42 29	39 68 95 10 96	09 24 23 00 62	56 12 80 73 16
37 21 34 17 68	68 96 83 23 56	32 84 80 15 31	44 73 67 34 77	91 15 79 74 58
14 29 09 34 04	87 83 07 55 07	76 58 30 83 64	87 29 25 58 84	86 50 60 00 25
58 43 28 06 36	49 52 83 51 14	47 56 91 29 34	05 87 31 06 95	12 45 57 09 09
10 43 67 29 70	80 62 80 03 42	10 80 21 38 84	90 56 35 03 09	43 12 74 49 14
44 38 88 39 54	86 97 37 44 22	00 95 01 31 76	17 16 29 56 63	38 78 94 49 81
90 69 59 19 51	85 39 52 85 13	07 28 37 07 61	11 16 36 27 03	78 86 72 04 95
41 47 10 25 62	97 05 31 03 61	20 26 36 31 62	68 69 86 95 44	84 95 48 46 45
91 94 14 63 19	75 89 11 47 11	31 56 34 19 09	79 57 92 36 59	14 93 87 81 40
80 06 54 18 66	09 18 94 06 19	98 40 07 17 81	22 45 44 84 11	24 62 20 42 31
67 72 77 63 48	84 08 31 55 58	24 33 45 77 58	80 45 67 93 82	75 70 16 08 24
59 40 24 13 27	79 26 88 86 30	01 31 60 10 39	53 58 47 70 93	85 81 56 39 38
05 90 35 89 95	01 61 16 96 94	50 78 13 69 36	37 68 53 37 31	71 26 35 03 71
44 43 80 69 98	46 68 05 14 82	90 78 50 05 62	77 79 13 57 44	59 60 10 39 66
61 81 31 96 82	00 57 25 60 59	46 72 60 18 77	55 66 12 62 11	08 99 55 64 57
42 88 07 10 05	24 98 65 63 21	47 21 61 88 32	27 80 30 21 60	10 92 35 36 12
77 94 30 05 39	28 10 99 00 27	12 73 73 99 12	49 99 57 94 82	96 88 57 17 91
78 83 19 76 16	94 11 68 84 26	23 54 20 86 85	23 86 66 99 07	36 37 34 92 09
87 76 59 61 81	43 63 64 61 61	65 76 36 95 90	18 48 27 45 68	27 23 65 30 72
91 43 05 96 47	55 78 99 95 24	37 55 85 78 78	01 48 41 19 10	35 19 54 07 73
84 97 77 72 73	09 62 06 65 72	87 12 49 03 60	41 15 20 76 27	50 47 02 29 16
87 41 60 76 83	44 88 96 07 80	83 05 83 38 96	73 70 66 81 90	30 56 10 48 59

Ainsi il nous faudra examiner 4706 sujets dans un sondage aléatoire élémentaire. Puisqu'en pratique nous utiliserons un sondage en grappes, ce chiffre devra être multiplié par un facteur approprié reflétant le degré d'homogénéité à attendre à l'intérieur de chaque grappe en ce qui concerne le risque de la maladie.

Si dans l'exemple ci-dessus nous sommes prêts à accepter un ordre de précision inférieur, c'est-à-dire que nous pourrions accepter une valeur de $d = 30\%$ de p ($= 2\%$ dans ce cas), n sera $= 1,96^2 \cdot 2 \cdot (100 - 2) / 0,6^2 = 2091,54$. Ainsi, un sondage aléatoire élémentaire portant sur environ 2100 sujets aurait été suffisant. Dans le premier cas, nous pourrions dire que l'estimation est de $2\% \pm 0,4\%$; dans le second, $2\% \pm 0,6\%$. Il est intéressant de noter que le faible gain de précision (0,6 à 0,4) oblige à plus que doubler la taille de l'échantillon. En général, pour augmenter la précision de y x fois, il nous faut augmenter la taille de l'échantillon par y^2 fois. Ceci montre bien l'importance de décider a priori du minimum de précision acceptable.

6. EMPLOI D'UNE TABLE DE NOMBRES AU HASARD ET EXEMPLES PRATIQUES DE SELECTION D'ECHANTILLONS

Le processus du choix au hasard revient à choisir des nombres comme dans une loterie. Organiser une "loterie" chaque fois qu'il faut sélectionner un échantillon aléatoire pose des problèmes pratiques. Cependant, des nombres au hasard ont été sélectionnés plusieurs fois et diverses tables ont été publiées, qui peuvent très bien être utilisées. Un exemplaire d'une telle publication est reproduit ici.⁴

Pour utiliser la table, il faut réaliser que chaque nombre (à un chiffre) est indépendant. L'impression en blocs (grands blocs de 5 et petits blocs de 2) est uniquement faite pour éviter la fatigue des yeux lors de la lecture. On peut ainsi observer qu'il y a 50 rangées et 50 colonnes dans la table ci-jointe, comprenant au total 2500 nombres à un chiffre, c'est-à-dire nombres de 0 à 9, sélectionnés indépendamment. Le processus d'une sélection au hasard peut être illustré par l'exemple hypothétique suivant.

Supposons qu'il y ait 25 villages (de taille approximativement égale) et que l'on souhaite en sélectionner 6 au hasard. Imaginons que les villages soient numérotés 01, 02, 03, ... 25. Nous pouvons maintenant commencer à partir de n'importe quelle rangée et colonne dans la table ci-jointe et chercher des nombres à deux chiffres. Supposons que nous commençons par la rangée 14 et la colonne 16. Puisque nous sommes intéressés par des nombres à deux chiffres, nous devons lire les colonnes 16 et 17 ensemble. Dans ce cas, le premier nombre est 31. Puisque le nombre total de villages n'est que de 25, le nombre 31 doit être rejeté. Les deux nombres suivants, en lisant vers le bas, sont 46 et 73 qui doivent être rejetés pour la même raison. Le suivant est 19, et de ce fait le village 19 est sélectionné. En continuant de la même façon, on notera que les cinq nombres suivants sont supérieurs à 25 et doivent être rejetés, le prochain nombre acceptable étant 10. De ce fait, le village 10 est sélectionné. Le nombre suivant est "07", le village 7 est donc également sélectionné. Les quatre nombres suivants sont supérieurs à 25. Le cinquième est 24, et de ce fait le village 24 est également sélectionné. Les six nombres suivants sont supérieurs à 25. Le septième est 10 qui est acceptable mais le village a déjà été sélectionné; de ce fait nous devons chercher d'autres nombres. Toujours en descendant, les deux prochains nombres acceptables seront 14 et 15, indiquant que les villages portant ces nombres doivent être sélectionnés. Ainsi nous avons sélectionné au hasard les six villages correspondant aux nombres : 19, 10, 07, 24, 14 et 15. En plaçant les nombres à la suite, nous dirons que les villages 7, 10, 14, 15, 19 et 24 ont été sélectionnés.

La technique indiquée ci-dessus ignore la taille des villages; c'est-à-dire que chacun des 25 villages, quelle que soit sa taille, a une chance égale ($6/25$) d'être sélectionné. La technique est acceptable si les villages ont à peu près la même taille. Mais s'ils sont inégaux, il est préférable de faire la sélection de telle façon que la probabilité de sélection d'un village soit proportionnelle à sa taille. La méthode de sélection est bien illustrée par l'exemple suivant.

Supposons que la population des 25 villages soit celle indiquée dans le tableau suivant :

N° du village	Population	Total cum.	Intervalles
1	150	150	0001-0150
2	270	420	0151-0420
3	220	640	0421-0640
4	380	1 020	0641-1020
5	125	1 145	1021-1145
6	420	1 565	1146-1565
7	318	1 883	1566-1883
8	118	2 001	1884-2001
9	520	2 521	2002-2521
10	126	2 547	2522-2547
11	315	2 962	2548-2962
12	278	3 240	2963-3240
13	514	3 754	3241-3754
14	245	3 999	3755-3999
15	337	4 336	4000-4336
16	267	4 603	4337-4603
17	188	4 791	4604-4791
18	265	5 056	4792-5056
19	484	5 540	5057-5540
20	196	5 736	5541-5736
21	284	6 020	5737-6020
22	503	6 523	6021-6523
23	96	6 619	6524-6619
24	121	6 740	6620-6740
25	276	7 016	6741-7016
	<hr/> 7 016		

La population totale des 25 villages est de 7016 habitants. Si nous imaginons que les individus, commençant à partir du village 1, portent les nombres de 0001 à 7016, le procédé consiste à sélectionner 6 individus (c'est-à-dire 6 chiffres au hasard entre 0001 et 7016) identifiant les villages où ces individus se trouvent et de considérer ces villages comme ceux qui sont sélectionnés. L'opération est facilitée si l'on construit une troisième colonne dans la table ci-dessus appelée "Totaux cumulés". Les chiffres successifs situés dans cette colonne sont les populations du village 1, la population des villages 1 et 2, la population des villages 1, 2 et 3, etc., le dernier chiffre situé en face du village 25 est le total de tous les villages 1, 2, etc., jusqu'à et y compris la population du village 25. Une quatrième colonne est établie pour indiquer immédiatement les intervalles permettant d'identifier le village correspondant.

Nous pouvons maintenant choisir 6 nombres au hasard entre 0001 et 7016. Puisque nous sommes intéressés au choix de 6 villages séparés, si deux nombres indiquent le même village, le second nombre sera rejeté.

Nous pourrions, par exemple, commencer à la rangée 36 et à la colonne 31 de la table ci-jointe des nombres au hasard et, à partir de là, examiner tous les nombres à 4 chiffres. Le premier nombre que nous voyons est 6869. Si nous regardons les totaux cumulés, nous verrons que l'intervalle entre 6741 et 7016 (où se situe le nombre 6869 sélectionné) correspond au village 25, qui de ce fait est choisi. Le nombre suivant, 7957, est manifestement supérieur à 7016 et doit être rejeté. Le nombre suivant est 2245. La colonne des intervalles nous montre que les nombres 2002 à 2521 (qui comprennent 2245) correspondent au village 9, qui doit être de ce fait le second village sélectionné.

Le troisième nombre, 8045, est supérieur à 7016 et doit être rejeté. Le suivant est 5348, que l'on peut voir comme appartenant au village 19. Ainsi le troisième village sélectionné sera le village 19. Le nombre suivant est 3768, correspondant au village 14. Le suivant est 7779 et doit être rejeté et le suivant est 5566, correspondant au village 20. Le nombre suivant, 2780, correspond au village 11. Ainsi, les six villages sélectionnés dans l'ordre sont 19, 25, 9, 14, 20 et 11. En plaçant les nombres à la suite, les villages sélectionnés sont 9, 11, 14, 19, 20 et 25. (Si après avoir choisi un village, par exemple 14, nous nous heurtons encore à un nombre comme 3801 indiquant le même village, ce nombre sera rejeté.)

En pratique, pour des raisons expliquées dans la section 4.3, il sera beaucoup plus utile d'employer un sondage en grappes pour estimer la prévalence de la lèpre. L'exemple hypothétique suivant permet de mieux comprendre la méthode pratique :

Supposons que l'on souhaite estimer la prévalence de toutes les formes de lèpre dans une région qui consiste en 10 sous-districts avec la population suivante :

Sous-district N°	Population estimée
1	15 678
2	10 879
3	7 670
4	3 654
5	8 987
6	14 356
7	7 987
8	11 889
9	12 678
10	5 889
Total	99 667

La détermination de la dimension souhaitable d'un échantillon sera traitée dans une autre section. Pour le moment, supposons que la meilleure méthode soit de choisir comme échantillon 20 grappes de 300 individus chaque, c'est-à-dire une population totale de 6000 individus. Supposons également qu'une estimation pour la région tout entière soit suffisante et que les estimations de prévalence par sous-district ne soient pas nécessaires.

Dans ce cas, le problème revient à sélectionner 20 endroits au hasard parmi les 10 sous-districts. Il serait possible que certains de ces sous-districts n'apparaissent pas dans l'échantillon et que plus d'une grappe se trouve dans certains des sous-districts.

Le premier degré dans le processus de sélection est de 1) établir la liste des sous-districts; 2) indiquer pour chacun d'eux la population correspondante; 3) établir des totaux cumulés comme dans l'exemple précédent et 4) indiquer les intervalles correspondants à chaque sous-district. Ceci est illustré dans le tableau suivant :

Numéro du sous-district	Population estimée	Total cumulé	Intervalles correspondant aux sous-districts
1	15 678	15 678	00 001-15 678
2	10 879	26 557	15 679-26 557
3	7 670	34 227	26 558-34 227
4	3 654	37 881	34 228-37 881
5	8 987	46 868	37 882-46 868
6	14 356	61 224	46 869-61 224
7	7 987	69 211	61 225-69 211
8	11 889	81 100	69 212-81 100
9	12 678	93 778	81 101-93 778
10	5 889	99 667	93 779-99 667
Total	99 667		

La mesure suivante consiste à choisir 20 nombres indépendants au hasard, à 5 chiffres, entre 00001 et 99667. Si l'on regarde la table de nombres au hasard ci-jointe et que l'on commence au hasard, disons à la rangée 9, colonne 19, nous trouvons que le premier nombre est 949362 et ce nombre correspond au sous-district 10. Les nombres suivants et leur situation sont indiqués dans le tableau suivant :

<u>Nombre au hasard</u>	<u>Numéro du sous-district</u>
17479	2
09727	1
71092	8
25544	2
52662	6
53957	6
48442	6
60681	6
99672	
	Nombre à être rejeté, étant supérieur à 99667
38676	5
72865	8
51741	6
23765	2
86795	9
01773	1
14483	1
49891	6
50018	6
28074	3
29396	3
56328	6

Ainsi, 20 nombres ont été choisis, et à partir de ce choix, les sous-districts 1, 2, 3, 5, 6, 8 et 9 qui ont été sélectionnés et le nombre de grappes à être examinées dans chacun d'eux est le suivant :

<u>Sous-district</u>	<u>Nombre de grappes à être examinées</u>
1	3
2	3
3	2
5	1
6	8
8	2
9	1
	TOTAL
	20

Ce qui a été fait a donc été de sélectionner, au premier degré, des sous-districts et de répartir parmi eux les 20 grappes à être examinées. La sélection et la répartition ont été faites de telle façon que la probabilité de sélectionner un sous-district et que le nombre de grappes à être réparties entre ceux qui ont été sélectionnés ont été rendus proportionnels à la population des sous-districts.

Si un sous-district sélectionné est étendu et divisé en sous-unités plus petites, le processus indiqué peut être répété pour choisir les sous-unités (et le nombre de grappes dans chacune d'elles). Ce sera le deuxième degré et le processus peut être répété pour couvrir une sélection du troisième, du quatrième, etc., degré, jusqu'à ce qu'on atteigne une sous-unité ayant une population raisonnablement petite. Jusqu'à présent, pour la sélection à chaque degré, des estimations de population ont été utilisées. Elles peuvent être obtenues à partir d'anciens recensements ou d'autres chiffres. Cependant, une fois atteint un degré où les populations sont suffisamment petites (par exemple, groupes de villages), un recensement doit être effectué à ce niveau pour obtenir une estimation plus à jour et plus précise de la population.

A titre d'exemple, prenons le sous-district 3. Deux grappes ont été sélectionnées à partir de cette sous-unité. Disons que ce sous-district comprend 200 villages, certains petits et d'autres grands.

A ce stade, il est préférable de grouper de petits villages avoisinants en un seul pour aboutir finalement à 25 groupes de villages de taille à peu près égale (c'est-à-dire ayant une population de 300 individus).

La prochaine étape est de choisir deux nombres au hasard entre 01 et 25 et d'examiner complètement les groupes de villages ainsi indiqués.

6.1 Villages à large population

Dans l'exemple pratique ci-dessus, nous avons supposé que les villages étaient petits, avec une population de 500 individus ou moins. Il est possible dans de nombreux cas que certains des villages, ou tous, aient des populations plus larges. Si la dimension de la grappe a été fixée à 500, la question se pose de savoir comment sélectionner les 500 individus.

Si le village peut être assez bien divisé en petites sections par des limites très nettes comme un canal, un champ, etc., le plus simple sera de choisir un nombre voulu de ces sections au hasard, de la façon décrite plus haut pour le choix des villages.

Une autre possibilité est de procéder de la façon suivante. Supposons que la population du village soit de 2000 individus vivant dans 300 maisons. Si les maisons ne sont pas déjà numérotées, elles devraient l'être de 001 à 300. A partir d'une table de nombres au hasard, choisissons un nombre situé entre 001 et 300. Supposons que ce soit le nombre 056. Ceci sera notre point de départ. Tous les individus des maisons 056, 057, 058, etc., seront examinés jusqu'à ce que

le nombre requis de 500 soit atteint. Tous les membres des dernières maisons faisant partie de l'échantillon doivent être examinés. Par exemple, les maisons dont les nombres vont de 056 à 155 peuvent contenir 497 individus et la maison dont le nombre est 156, 8. Le principe est d'examiner tous ceux qui se trouvent dans la 156^e maison. La grappe sera alors constituée de 505 individus. Si au lieu de la maison ayant le nombre 056, une maison dont le nombre est plus proche de 300 était sélectionnée, disons le nombre 290, l'échantillon sera constitué par les individus vivant dans les maisons 290, 291 ... 300 et 001, 002 ... etc., jusqu'à ce qu'une grappe d'une taille d'environ 500 personnes soit obtenue.

6.2 Sélection de grappes dans des régions urbaines

La façon de procéder est bien illustrée par l'exemple hypothétique suivant : supposons que la commune ait une population estimée être de 90 000 individus, ceci d'après le recensement le plus récent (qui peut dater de quelques années) et que cette commune soit divisée en cinq localités, A, B, C, D et E ayant les estimations de population suivantes :

<u>Localité</u>	<u>Population estimée</u>
A	15 500
B	23 400
C	12 000
D	26 000
E	13 100
	<hr/>
TOTAL	90 000

Il arrive assez souvent qu'une majorité de gens aisés de la population vivent dans des zones très faciles à distinguer (disons la localité E). Il peut être considéré comme souhaitable de stratifier la population en au moins deux strates pour distinguer les gens aisés des autres et de choisir des grappes séparées à partir de chacune des strates. Supposons que pour des raisons théoriques et pratiques, il soit préférable de prendre une grappe à partir de la zone E des gens aisés et deux grappes du reste de la ville.

Une grappe sera donc choisie à partir de la localité E et deux autres à partir des localités A, B, C et D de la façon suivante :

Localité	Population estimée	Total cumulé	Intervalles indiquant la localité
A	15 500	15 500	00 001-15 500
B	23 400	38 900	15 501-38 900
C	12 000	50 900	38 901-50 900
D	26 000	76 900	50 901-76 900

Comme cela a été expliqué dans les sections précédentes, une table est construite comme indiqué ci-dessus. Supposons maintenant que l'on désire avoir deux grappes n'appartenant pas à la même localité.

Nous pouvons maintenant chercher un nombre à cinq chiffres (car le total est de 76 900) dans la table des nombres au hasard. En utilisant la table donnée dans ce manuel et en partant d'un point choisi au hasard, rangée 15 et colonne 17, nous rencontrerons le nombre à cinq chiffres 67539. On peut voir que ce chiffre correspond à la localité D qui est de ce fait sélectionnée. Le nombre suivant est 34484 qui correspond à la localité B. Ainsi, les localités B et D sont sélectionnées.

La sélection de la grappe sera illustrée pour la localité B.

Le chef du projet (ou le statisticien s'il y en a un), en collaboration avec les autorités locales, dessine une carte de la localité en la divisant en blocs, estime le nombre d'habitants

dans chaque bloc et attribue des numéros aux blocs. Supposons que la localité soit divisée en 100 blocs. Les blocs seront disposés en ordre de succession, de même que la population estimée. Un nombre au hasard est choisi comme précédemment pour déterminer le point de départ (c'est-à-dire le bloc). A ce stade, il est pratique d'examiner toutes les personnes vivant dans le bloc sélectionné et dans les blocs suivants jusqu'à ce que le nombre requis (dimension de la grappe) soit atteint.

6.3 Sélection de grappes dans des populations largement disséminées et dans des bidonvilles

En pratique, il y a deux situations dans lesquelles la sélection de grappes comme celles indiquées dans les sections précédentes devient impossible :

- 1) lorsque la communauté est petite et vit dans des zones extrêmement disséminées;
- 2) lorsque la communauté consiste en une "population marginale" qui se rassemble dans les zones urbaines et n'est pas prise en compte dans le recensement. Ces populations vivent généralement dans des bidonvilles, à la périphérie des localités urbaines, dans des abris improvisés.

Dans ces cas, le mieux est d'avoir recours au "sondage par région".

La méthode consiste à dessiner tout d'abord une carte délimitant la région. Un point de repère facile à distinguer, comme la hutte du chef, une école locale, etc., est choisi et marqué sur la carte aussi exactement que possible. A partir de ce point, des coordonnées sont tracées et la carte est alors divisée en carrés égaux (voir illustration). Le numérotage des carrés peut être effectué dans un ordre quelconque.

Supposons qu'il en résulte 64 carrés. Un nombre au hasard entre 01 et 64 est maintenant choisi dans la table des nombres au hasard. Supposons que 09 soit sélectionné. Le carré 9 est recherché sur la carte du pays en utilisant un compas et en mesurant les distances en pas (ou en utilisant un ruban métrique ou le compteur kilométrique d'un véhicule motorisé). Les habitants du "carré 9" sont maintenant enregistrés en commençant par la partie du carré la plus proche du carré 8, l'enregistrement continue dans les carrés 10, 11, etc., jusqu'à ce qu'une grappe d'une taille de 500 individus soit atteinte.

Les carrés inclus dans le groupe échantillon doivent être marqués par des piquets, des bâtons, des pierres, etc., pour faciliter l'identification du groupe pendant les examens. Suivant cette méthode, la population vivant dans les régions limitrophes a plus de chances d'être sélectionnée que les autres, ce à quoi on peut remédier en pondérant de façon différente (d'après la population estimée, etc.) les différents carrés et en choisissant un point de départ comme précédemment pour la sélection des villages.

28							
	13	12	61	36	37	52	53
27	14	11	62	35	38	51	54
26	15	10	63	34	39	50	55
25	16	9	64	33	40	49	56
24	17	8	1	32	41	48	57
23	18	7	2	31	42	47	58
22	19	6	3	30	43	46	59
21	20	5	4	29	44	45	60

○ Point de repère
(point de départ)

limite de la
région

7. EXPERIENCE PROVENANT DES ENQUETES LAT

Au début des années 60, l'équipe consultative de la lèpre de l'OMS (LAT) était responsable d'un certain nombre de sondages dont la plupart étaient limités à certaines parties des pays; les estimations obtenues ne pouvaient donc pas être considérées comme des estimations nationales de la lèpre. Néanmoins, les observations faites peuvent être utiles pour planifier de futurs sondages.

Le tableau suivant* indique le nombre d'individus examinés, le nombre de cas de lèpre relevé par le sondage, le taux de prévalence et les limites de confiance à 95 %. On observera que la prévalence varie de 0,14 % (à Parana et San José en Argentine) à 3,86 % (à Shwebo et Myingyan en Birmanie). En Birmanie, où plus de 19 000 individus ont été examinés, la précision est plus élevée - les limites de confiance se situant à ± 20 % de l'estimation. Ailleurs, sauf en Argentine, les limites sont le plus souvent à ± 30 % des estimations du sondage. L'enquête effectuée dans les provinces Khon Kaen et Lampang en Thaïlande donne peut-être une indication de la taille d'un échantillon. L'estimation était de 1,22 % avec des limites de confiance de 1,55 et de 0,88 - estimation que l'on pouvait considérer comme raisonnablement précise pour la plupart des objectifs. La méthode de sondage consistait à examiner 33 grappes d'environ 500 individus chacune. Le nombre total d'individus examinés était de 16 862.

PREVALENCE DE LA LEPRE DANS CERTAINS PAYS SELECTIONNES

	Exam.	Grappes	Positives	Taux de prévalence	Lim. de conf.	
					supérieures	inférieures
Argentine Parana et San José	5 169	50	7	0,14	0,23	1,04
Argentine Chaco I et II	1 635	15	42	2,57	4,10	1,04
Libéria	5 327	35	87	1,63	2,09	1,18
Cameroun	14 473	30	374	2,58	3,26	1,91
Nigéria occidental	6 496	16	101	1,55	2,04	1,07
Thaïlande	16 862	33	205	1,22	1,55	0,88
Birmanie	19 493	38	753	3,86	4,58	3,15
Philippines	10 361	20	69	0,67	0,87	0,46

* En fait, les méthodes de sondage utilisées dans ces enquêtes variaient et étaient quelque peu plus compliquées. Dans cette section, avant tout pour illustrer les effets d'un sondage en grappes, les données ayant trait à différentes grappes dans différents sondages sont reproduites et les calculs sont faits en supposant qu'un sondage en grappe unique a été effectué, comme décrit dans ce manuel.

Les caractéristiques supplémentaires suivantes qui se dégagent des sondages LAT peuvent donner des idées pour de futurs sondages prévus pour l'évaluation ou pour obtenir des informations de base.

	% de la population de plus de 14 ans	% de tous les cas de lèpre dans ce groupe d'âge	Prévalence (%) (dans ce groupe d'âge)
Cameroun	62,4	93	3,9
Nigéria occidental	51,6	89	2,7
Thaïlande	49,6	93	2,3
Birmanie			
Shwebo	55,3	71	5,1
Myingyan	59,8	74	5,5
Philippines	57,5	87	1,01
Argentine			
Parana et			
San José	65,5	91	0,14
Chaco	57,0	93	0,9

* Généralement, 40 à 60 % des cas présentaient des invalidités.

En général, dans de nombreuses régions d'hyperendémicité (dans les pays en développement), environ 70 à 90 % des cas ont donc des chances d'être trouvés dans les groupes d'âge supérieurs à 14 ans, qui comprennent un peu plus de 50 % de la population. De ce fait, une possibilité est de limiter le type d'enquête de population (visites de foyers, etc.) à la population adulte (âgée de plus de 14 ans). La généralisation à partir du sondage ne s'appliquera alors qu'à la population adulte de la communauté ou du pays. Dans de nombreuses situations, la prévalence s'est montrée déjà passablement élevée dans le groupe d'âge 10-14 (par exemple 2,9 % dans la région de Singy de la Birmanie du Nord). Dans ces conditions, il faudrait que le sondage porte sur le groupe d'âge de 10 ans et au-dessus. Les enfants examinés dans ces foyers pourraient fournir des informations sur les groupes d'âge inférieurs, mais avec une précision plus faible. Cependant, la prévalence d'ensemble pour tous les groupes d'âge obtenue de cette façon ne serait pas exagérément affectée. Ceci est davantage discuté dans le paragraphe suivant.

8. QUELQUES CONSIDERATIONS PRATIQUES IMPORTANTES

Les généralités suivantes, concernant les sondages, doivent retenir l'attention.

1. Une fois l'échantillon sélectionné, il est impératif que chaque individu (couverture de 100 %) de l'échantillon soit examiné. En pratique, ceci est souvent impossible. Il faut s'efforcer d'obtenir une couverture aussi étendue que possible et faire l'impossible pour examiner un sous-échantillon aléatoire de la population sondée, non examinée au début. L'information obtenue du sous-échantillon indiquera les moyens d'ajuster l'estimation obtenue de la population sondée réellement couverte. La précision d'ensemble peut être réduite dans une certaine mesure parce que nous ajoutons la variation du sondage du sous-échantillon à la variation du sondage de l'échantillon initial. Cependant, l'estimation ainsi obtenue ne sera pas biaisée. Le biais est une des erreurs les plus importantes non dues au sondage et qu'il faut éviter.

2. Comme cela a été mentionné dans la section précédente sur l'expérience provenant des sondages LAT, la majorité des cas se trouvent dans le groupe d'âge adulte, mais la pratique montre que les enfants sont plus faciles à couvrir. La plupart des sondages ne font état que

d'une couverture d'ensemble. Une bonne couverture d'ensemble peut consister en une couverture des enfants extrêmement satisfaisante couplée à une mauvaise couverture des adultes. Pour cette raison, la méthodologie suivante est suggérée :

- a) traiter le sondage comme s'il était stratifié, en enfants et en adultes;
- b) choisir une taille d'échantillon qui convienne à l'estimation chez les enfants et une autre taille d'échantillon qui convienne à l'estimation chez les adultes;
- c) sur le plan opérationnel, les mêmes groupes de villages choisis pour les adultes seront utilisés pour les enfants. Il faut prévoir qu'un échantillon plus large sera nécessaire pour les enfants car la prévalence de la maladie est plutôt faible chez eux. Des groupes de villages supplémentaires peuvent être sélectionnés au hasard pour procurer le nombre requis d'enfants. Dans ces villages supplémentaires, les examens peuvent être limités aux enfants seuls. S'il est nécessaire d'examiner certains adultes pour des raisons opérationnelles, ceci est possible mais il n'en sera pas tenu compte dans l'analyse;
- d) si une couverture insuffisante, en particulier des adultes, peut introduire de sérieux biais dans les estimations, il peut en être de même si l'on examine plus d'adultes dans des villages supplémentaires pour compléter le total voulu.

Il faut se souvenir qu'un sondage n'est pas une détection de masse de cas. Le seul but du sondage est de procurer des estimations de prévalence non biaisées et raisonnablement précises pour la communauté entière d'où les échantillons sont tirés. Le temps, les efforts et les ressources nécessaires au suivi d'un sous-échantillon de l'échantillon de population non couvert, a beaucoup plus de sens que l'examen d'un grand nombre d'individus n'appartenant pas à l'échantillon sélectionné initial.

Un exemple pratique illustrera les points mentionnés ci-dessus. Supposons que la prévalence attendue dans tous les groupes d'âge soit d'environ 1,2 % et que la prévalence dans le groupe d'âge 15 et au-dessus soit de 2,3 %. De plus, la population dans ce groupe d'âge adulte représenterait 50 % du total. Indiquons que la précision des estimations devrait être telle que le coefficient de variation pour les adultes devrait être inférieur à 15 % et celui des enfants inférieur à 20 %.

La prévalence attendue spécifique d'âge dans une situation semblable est généralement :

Groupe d'âge au-dessous de 15 - 0,18 %

Groupe d'âge au-dessus de 15 - 2,3 %

Dans des conditions de sondage aléatoire élémentaire, les tailles appropriées des échantillons seraient de 1835 adultes et de 13 863 enfants. Dans un sondage en grappes, la taille des échantillons devrait être multipliée par quatre. On aboutirait ainsi à 7340 adultes et à environ 55 000 enfants. En utilisant une méthode en grappes typique (c'est-à-dire villages ou grappes de villages d'une taille de 500) et en imaginant que 50 % de la population aient 15 ans et plus, ceci signifierait qu'il faut examiner une population totale de 14 680 individus ou environ 29 grappes pour les adultes et une population de 55 000 enfants qui sera obtenue à partir d'une population totale de 110 000 individus ou 220 grappes.

La technique consisterait à choisir 220 grappes au hasard et d'en choisir 29 à nouveau au hasard. La population totale des 29 grappes sera examinée et dans les 191 restantes, ne le seront que les enfants au-dessous de 15 ans.

Si, pour une raison quelconque, environ 60 % seulement des adultes sélectionnés peuvent être examinés, à peu près 4400, il est suggéré qu'un échantillon de 10 % des restants, c'est-à-dire 10 % de 2940 soit 294 adultes, soit sélectionné au hasard et que tous les efforts soient faits pour les retrouver.

La population totale de 55 000 enfants dans l'exemple ci-dessus semble anormalement large. Ceci est dû au fait que la prévalence dans ce groupe d'âge est très faible. Le degré désiré

de précision signifie que si la prévalence dans l'échantillon est en fait de 0,18 %, nous serions en mesure de dire que les limites de confiance à 95 % sont $0,18 \pm 0,07$ (c'est-à-dire 40 % de 0,18), soit entre 0,25 % et 0,11 %. Il pourrait être suffisant de démontrer que la prévalence des enfants est inférieure à 0,3 %. Dans ce cas, nous demanderions une précision exprimée par un coefficient de variation qui serait de $0,06/0,18$, soit 33,33 %. On parviendrait ainsi à une taille d'échantillon de 4941 individus par sondage aléatoire élémentaire. Si l'on tient compte de l'effet de grappes, ceci signifierait qu'il faut en examiner 4 fois^{*} plus, soit environ 20 000 enfants. Ceci serait obtenu à partir d'une population totale de 40 000 individus ou 80 grappes.

Les 29 grappes sélectionnées à l'origine pour les adultes procureront en fait 7340 enfants. Avec la prévalence observée de 0,18 %, ceci amènerait à une limite supérieure de confiance à 95 % de 0,38 %.

Les tailles d'échantillon requises pour les enfants pourraient peut-être être réduites si l'on pouvait établir une relation entre la prévalence chez les adultes et celle chez les enfants. Si une telle relation est observée, même pendant le sondage (par exemple dans les grappes sélectionnées), des méthodes plus raffinées d'estimation peuvent être utilisées et la taille de l'échantillon pour les enfants considérablement réduite. A ce stade, une consultation statistique spéciale est souhaitable.

Stratification

Le sondage par stratification peut être nécessaire dans les enquêtes sur la lèpre pour les raisons suivantes :

1. Les pays sont généralement divisés en provinces ou États et la lutte contre la lèpre dépend souvent des autorités sanitaires de l'État. Des informations sur la situation de la lèpre devront donc être obtenues séparément par État, ou peut-être même par plus petites sous-unités administratives.
2. Des informations sont généralement requises séparément pour les régions urbaines et rurales, l'organisation de la lutte contre la lèpre pouvant être différente dans les deux cas.
3. Les données statistiques montrent que la stratification d'un pays ou d'un État par niveaux d'endémicité augmentera l'efficacité de la méthode de sondage. Dans ce but, la région peut être divisée, en se basant sur les expériences précédentes, en régions d'hyperendémicité, d'endémicité modérée et autres.

Il existe un certain nombre d'autres variables épidémiologiques connues pour être associées à la maladie telles que l'âge, le sexe, la situation des contacts familiaux, etc. La stratification basée sur ces variables peut contribuer théoriquement à augmenter l'efficacité, mais peut avoir des inconvénients sur le plan opérationnel.

Lorsque les strates sont décidées, des échantillons aléatoires sont tirés de chacune d'elles, de façon indépendante. La taille de l'échantillon (ou la fraction de la population qu'elle représente) peut être, et est en général, très différente suivant les différentes strates. L'estimation globale (c'est-à-dire l'estimation pour le pays, etc.) est l'association des estimations pour chacune des strates, pondérées de façon voulue (c'est-à-dire en tenant compte des différences de taille de l'échantillon).

9. ORGANISATION D'UN SONDAGE

Lorsqu'il est prévu d'effectuer un sondage dans une région, le chef du projet devrait visiter cette dernière environ un mois avant la date choisie pour débiter le sondage. Il devrait alors trouver la ou les personnes "clés" dont la coopération sera nécessaire pour la réussite des opérations. Les objectifs et la méthode du sondage devraient lui être expliqués en détail.

* D'après l'expérience obtenue en Thaïlande.

Il faut l'assurer que l'information obtenue des sondages ne sera utilisée qu'à des fins de planification et profitera finalement à la communauté. Le secret sera gardé sur les noms des individus et les publications ne comporteront que des données globales. Les personnes "clés" expliqueront elles-mêmes pourquoi certains individus pourraient ne pas coopérer. Il faudra décider en commun des moyens à utiliser pour surmonter ces obstacles. En général, ils sont dus à l'opprobre social et à la crainte de perdre son emploi. S'il y a une réelle absence de coopération, peut-être serait-il préférable d'éviter le mot "lèpre" en parlant du sondage. Une possibilité est de parler simplement "d'enquête" dermatologique. Le diagnostic de lèpre doit rester secret entre le malade et l'agent de lutte contre la lèpre. Si nécessaire, la section diagnostique du dossier peut être codée.

Un problème important est souvent celui de l'examen des femmes. Les moyens d'y faire face doivent être décidés très tôt. Par exemple, une sage-femme locale ou une femme de confiance appartenant à la communauté peut être invitée à se joindre à l'équipe de sondage et à faire part à un membre de l'équipe de toutes les anomalies cutanées et déformations observées. Ceci doit être discuté avec la personne "clé" qui sera alors en mesure d'obtenir le consentement de la population.

Le moment est alors propice pour obtenir une carte détaillée de la région, des informations sur le plus récent recensement, avec le plus de détails possibles (âge, sexe, etc., par village ou pâté de maisons, liste des foyers, etc.). S'il n'existe pas de données provenant d'un recensement, il faut recourir à d'autres moyens pour obtenir des estimations de la population. En général, il existe quelques indications sur la taille de la population, provenant soit de recensements (même anciens) ou de sondages ou d'enquêtes effectuées dans d'autres buts, au moins dans certaines parties de la région. Dans un cas, il existait une liste des contribuables. Le nombre obtenu a été multiplié par un facteur reflétant la proportion estimée de contribuables dans la communauté, pour obtenir une première indication approximative de la taille de la population. En principe, à moins que ceci soit absolument nécessaire, les listes de contribuables ne devraient pas être utilisées pour éviter toute association possible entre le sondage et les impôts. Si après une recherche intense on n'obtient aucune information, on peut avoir recours à un sondage de région.

Au moment de la première visite, des informations devraient être obtenues sur les groupes ethniques, les régions géographiques, etc., à exclusion du sondage. Une telle exclusion peut concerner par exemple des nomades qui ne peuvent être atteints, des régions de troubles, et d'autres, où le sondage ne sera pas autorisé. Pour tous ces groupes, leur taille et leur emplacement devraient être précisés et il faut savoir, en admettant qu'ils soient inclus dans le recensement, s'ils sont assez bien définis pour être exclus des chiffres de la population par district et communauté. D'autres informations intéressantes sont :

- quelle est la distribution de la population par région rurale et urbaine;
- à quel point est-elle disséminée sur de larges espaces;
- l'accessibilité aux communautés, en général, et pendant chaque saison;
- les périodes pendant lesquelles les gens se déplacent pour leur travail, etc.
- en général, le moment de l'année, les semaines et les jours où l'on peut espérer rencontrer facilement la population.

Puisque le sondage en grappes sera la méthode recommandée, il faudra à ce moment décider de la taille des grappes. Il est pratique de considérer comme une grappe un village ou un groupe de villages dans des régions rurales. Le nombre moyen de personnes qui peuvent être examinées pendant une période de temps donnée (disons trois semaines) doit être déterminé en tenant compte du temps nécessaire aux voyages et à la préparation du travail.

Enregistrement de la population du sondage

Avant le début réel des examens, le chef d'équipe et/ou un membre de l'équipe responsable des statistiques, doit commencer à enregistrer la population du sondage. Accompagné d'une personne considérée comme une autorité par la communauté, il doit visiter tous les quartiers

d'habitation (y compris les hôpitaux, les maisons de repos, les pensionnats, etc., de même que les maisons) dans les grappes sélectionnées et enregistrer les résidents. L'intervalle entre ces visites et les examens doit être aussi court que possible, la population pouvant être mobile dans certains cas.

Chaque ménage doit recevoir un numéro qui est peint à l'entrée de la maison ou à celle de l'ensemble de maisons. Une carte individuelle doit être remplie pour chaque personne appartenant au ménage, c'est-à-dire celui qui habituellement dort ou a dormi dans la maison et se propose d'y retourner. Les informations suivantes doivent figurer sur la carte :

Numéro de la grappe :

Nom du village :

Numéro du foyer :

Numéro de série :

Nom : (prénom et nom du père/du mari)

Age : (si l'âge n'est pas stipulé, il doit être estimé. Pour les enfants de moins d'un an, l'âge doit être donné en mois)

Sexe :

Relation avec le chef du foyer :

Des cartes doivent être également préparées pour les membres absents du foyer et pour ceux qui résident temporairement dans la maison. Peuvent être définis comme résidents temporaires : "ceux qui prévoient de passer la nuit dans la maison la nuit précédant le premier jour de l'examen mais qui n'appartiennent pas au ménage". Si un ménage entier est absent au moment de l'enregistrement, le nombre de personnes qu'il comprend doit être estimé (ou obtenu de voisins, etc.) et un nombre correspondant de cartes doivent être remplies avec la seule indication des numéros de la grappe, du ménage et des numéros de série.

La couverture, et par suite la fiabilité des estimations, dépendent beaucoup de la précision de l'enregistrement et il faut donc, à ce stade, être extrêmement rigoureux. Le principe le plus important est d'inclure tout membre du ménage lui-même inclus dans le sondage et d'exclure les autres. Si la situation exacte d'un individu est difficile à établir, il faut l'interroger soigneusement, de même que le chef du foyer, et si un doute persiste, il faut indiquer sur la carte, situation "douteuse".

Une personne est classée comme étant "temporairement absente" (TA) si elle appartient au ménage mais est absente de la communauté et pense en être absente la nuit précédant les examens. Une personne qui est simplement absente de la maison parce qu'elle travaille aux champs ou dans une usine, ou se livre à un commerce dans le village ou une cité voisine, mais qui pense retourner à la maison la nuit, doit être incluse dans les examens et ne pas être classée comme TA. Si l'on s'efforce en vain de la localiser et de l'examiner, elle ne doit pas être classée comme TA, mais simplement enregistrée comme étant absente pendant les examens.

Une personne est classée comme étant temporairement présente (TP) si elle n'appartient pas au ménage, mais est présente au moment de l'enregistrement et pense passer la nuit dans la maison la nuit avant le début des examens. Aucune carte n'est nécessaire pour les visiteurs occasionnels qui n'ont pas l'intention de passer la nuit.

Si, pour certaines raisons, il est nécessaire d'inclure dans les examens des individus qui n'appartiennent pas à l'échantillon sélectionné (par exemple à la demande du chef du village, etc.), une carte doit être remplie, mais la mention "n'appartient pas à l'échantillon" doit être portée en diagonale sur la carte, en caractères gras.

Les informations figurant sur la carte peuvent être vérifiées le jour de l'examen. Ceci est particulièrement important pour les rubriques comme TA et TP puisque ceci se rapporte, comme

mentionné ci-dessus, à l'absence ou à la présence de la personne dans la maison la nuit précédant les examens.

Des personnes apparaissant lors des examens et qui n'ont pas été enregistrées reçoivent des cartes et sont incluses dans l'examen seulement après que l'équipe ait soigneusement vérifié qu'elles appartiennent bien au groupe échantillon.

Le travail peut être organisé par des visites de porte en porte ou en demandant à tous les membres du groupe échantillon de se rendre pour les examens à un lieu central.

Les visites de porte en porte sont en général préférables dans les sondages sur la lèpre. Une des raisons en est que l'équipe se familiarise ainsi avec le groupe échantillon et peut vérifier si l'enregistrement et les examens sont bien complets. Cependant dans certaines régions, il est difficile d'effectuer un examen satisfaisant de la peau à l'intérieur des maisons quand celles-ci ont de petites fenêtres et que la lumière est insuffisante. Dans ces cas, il faut se demander s'il n'est pas possible d'examiner les gens chez eux mais en dehors de leur maison ou dans un lieu central avec une lumière suffisante. L'examen en un lieu central ne permet pas de bien contrôler la présence et l'identité des participants et l'on ne doit y recourir que lorsque la population est tellement disséminée que l'équipe ne peut raisonnablement espérer couvrir la région à pied. Il peut être également extrêmement difficile lors d'un examen en un lieu central d'identifier un individu et de trouver sa carte, particulièrement lorsqu'il n'est pas accompagné par ses proches.

Quelle que soit la façon dont le travail est organisé, l'identité de l'individu est vérifiée à chaque examen en lui demandant son nom et le nom de son père (et le nom de la mère dans certaines sociétés) en vérifiant que ses réponses correspondent bien à ce qui est inscrit sur sa carte personnelle. Il faut demander au sujet examiné son nom; le membre de l'équipe ne doit pas le lui suggérer ou lui demander s'il est correct.

Il faut s'efforcer d'examiner tout individu enregistré dans le groupe échantillon; ceci nécessite un contrôle constant de la couverture pour être en mesure de trouver les personnes qui ne se présentent pas à l'examen; des rendez-vous doivent leur être donnés dans la soirée ou tôt le matin à un moment où elles restent chez elles. Dans de nombreux cas, des visites répétées à domicile peuvent être nécessaires avant qu'une personne soit en mesure, ou puisse être persuadée, de se présenter aux examens.

10. PROCEDURES ET CRITERES DE DIAGNOSTIC

10.1 Introduction

Les procédures et critères de diagnostic de la lèpre, lors de sondages, doivent en général être limités à ce qu'il est possible de faire sur le terrain et à des méthodes qui peuvent être appliquées d'une façon raisonnablement uniforme et logique.

Les procédures diagnostiques sont essentiellement cliniques, et à un certain point bactériologiques; de ce fait les examinateurs doivent avoir reçu une formation suffisante sur les méthodes spécifiques à employer et posséder également une certaine expérience, ce qu'il peut être utile, dans certaines situations, de vérifier en demandant à un examinateur indépendant et sûr, si possible un médecin, de revoir les sujets classés comme cas, suspects et normaux.

Des problèmes diagnostiques surgissent surtout dans deux circonstances. a) En cas de lésions précoces, en particulier quand certaines d'entre elles ne présentent que des signes minimes ou limite de lèpre. Ceci suppose que les examinateurs devraient essayer d'être aussi objectifs que possible. b) En cas de lésions inactives qui le sont devenues soit en raison du traitement soit spontanément. Dans toute la mesure du possible, ces lésions doivent être reconnues comme telles et exclues de la liste des cas actifs.

10.2 Procédures diagnostiques

Les procédures diagnostiques doivent comprendre les mesures suivantes :

Inspection : inspection de toute la surface du corps sous une bonne lumière et de façon systématique. Dans certains cas, il peut être nécessaire d'avoir recours à des examinatrices. Le temps minimum d'inspection devrait être de deux minutes. Il peut être utile de demander au sujet examiné s'il n'a pas remarqué une tache ou une perte de sensibilité. L'inspection comprend la recherche de macules, de papules, d'infiltration diffuse, de nodules, de régions sèches de la peau, d'ulcères plantaires, et de difformités.

Palpation : la palpation des troncs nerveux doit toujours comprendre celle du nerf cubital et du nerf poplité externe. Les autres nerfs doivent être examinés suivant les besoins.

Recherche d'une perte de sensibilité : la recherche d'une perte de sensibilité doit être effectuée a) sur toute lésion cutanée dont l'aspect est compatible avec la lèpre, b) quand un malade se plaint d'engourdissement et c) quand il y a épaissement d'un tronc nerveux même en l'absence de lésions cutanées. L'examen de la sensibilité doit être limité à un toucher léger (toucher) et à une piqûre d'épingle (douleur). Dans certains cas, seule la piqûre d'épingle est possible.

Examen de frottis cutané : l'examen d'un frottis cutané doit être effectué chez tous les sujets pour lesquels le diagnostic de lèpre est certain ou douteux. L'examen, y compris la lecture de la lame doit être effectué d'une façon standard. Référ : Smear Technique; measurement of the Bacteriological Index (BI) : Doc. TDR/THELEP/Protocol/82.1 Appendice 4.

Biopsies cutanées : au moment de la planification, une décision doit être prise en ce qui concerne les biopsies cutanées. En général, il semble préférable de limiter les biopsies aux cas douteux ou lorsqu'il est difficile de faire la distinction entre état actif et inactif.

10.3 Critères et catégories diagnostiques

Pour un sondage, la population examinée peut être classée de la façon suivante :

I. Lèpre multibacillaire : lésions cutanées compatibles avec la lèpre, avec frottis positif en bacilles acido-résistants (BAR).

II. Lèpre active paucibacillaire : a) lésions cutanées avec anesthésie, compatibles avec la lèpre et avec frottis BAR négatif, ou b) zones de perte de sensibilité sur une peau apparemment normale avec nerf épaissi et sensible, et avec frottis BAR négatif.

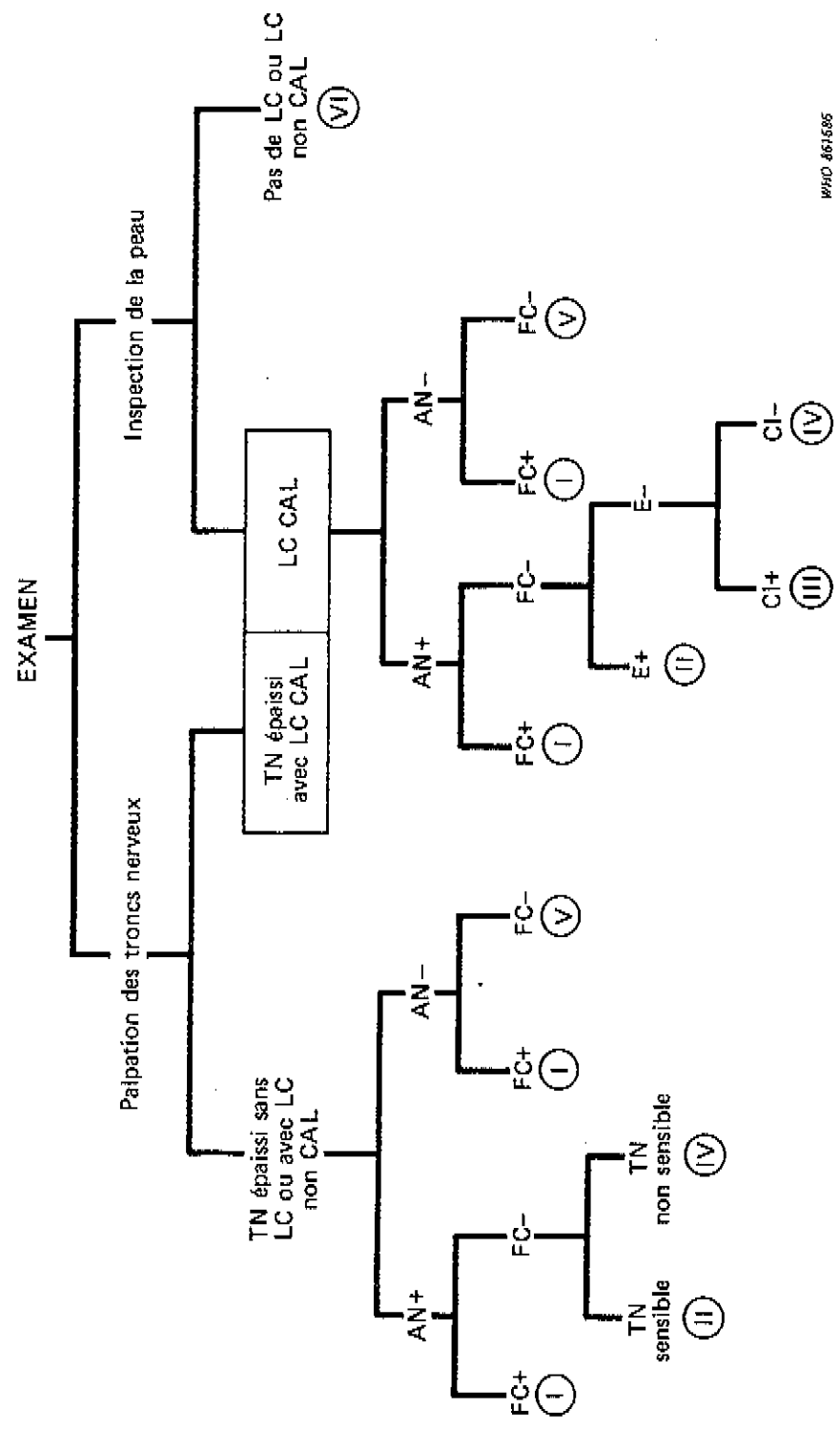
III. Lèpre inactive : lésions cutanées cicatricielles avec anesthésie qui pourraient être compatibles avec la lèpre, avec frottis BAR négatif.

IV. Lèpre d'activité douteuse : a) lésions cutanées avec anesthésie, compatibles avec la lèpre, sans érythème ni infiltration, ni cicatrice et avec frottis BAR négatif, ou b) zones de perte de sensibilité sur une peau apparemment normale avec nerf épaissi et non sensible, et avec frottis BAR négatif.

V. Lèpre douteuse : a) lésions cutanées à frottis négatif, sans anesthésie, pouvant être compatibles avec la lèpre ou b) épaissement d'un tronc nerveux sans zones de perte de sensibilité et avec frottis BAR négatif ou c) zones de perte de sensibilité sur une peau apparemment normale sans épaissement d'un tronc nerveux et avec frottis BAR négatif.

VI. Pas de lèpre : pas de lésion cutanée, pas d'épaississement nerveux et pas de perte de la sensibilité compatible avec la lèpre.

Les procédures d'examen sont schématiquement illustrées dans le tableau ci-joint.



WHO 461585

- | | | | | |
|-----|------------------------------|-----|---|--------------------------|
| I | Lèpre multibacillaire | LC | = | Lésion cutanée |
| II | Lèpre active paucibacillaire | CAL | = | Compatible avec la lèpre |
| III | Lèpre inactive | TN | = | Tronc nerveux |
| IV | Lèpre d'activité douteuse | AN | = | Anesthésie |
| V | Lèpre douteuse | E | = | Erythème |
| VI | Pas de lèpre | FC | = | Frottis cutané |
| | | CI | = | Cicatrice |

11. ERREURS NON DUES A L'ECHANTILLONNAGE

Les principes statistiques sont suffisamment avancés pour qu'il soit possible d'éviter ou de minimiser les variations d'échantillonnage. Par ailleurs, dans des enquêtes médicales, la suppression des erreurs non dues à l'échantillonnage est étroitement liée à la précision du diagnostic et à la couverture que l'on obtient. Des principes statistiques peuvent être incorporés dans les opérations de sondage pour supprimer les variations, ou tout au moins en donner des estimations auxquelles on peut s'attendre en raison des variations de diagnostic. La possibilité de réussir une couverture étendue dépend largement du milieu socioculturel et de l'autorité du personnel du sondage.

La couverture joue un rôle particulièrement important dans les sondages sur la lèpre. La maladie est très souvent étroitement liée aux caractéristiques des éléments de la population non couverts (appelés de ce fait "non répondants"); ceci perturbera sérieusement les estimations basées sur la population couverte et annulera tous les avantages attendus d'une modalité de sondage élaborée pour éliminer les erreurs d'échantillonnage. Pour cette raison, une part considérable des dépenses totales prévues pour le sondage doit être réservée aux solutions du problème des non-réponses.

Il est difficile de trouver des solutions à ce problème et de les généraliser; il n'est pas inutile de faire la liste des causes de non-réponse le plus souvent observées et d'envisager quelques mesures permettant de les minimiser.

11.1 Opprobre lié à la maladie

Dans la plupart des parties du monde, l'opprobre lié à la maladie reste considérable et les sujets atteints craignent d'être connus comme étant lépreux. Dans une certaine mesure, il est possible de surmonter cette difficulté par une intense campagne auprès du public en assurant la population que les informations seront strictement confidentielles et que la publication des résultats ne portera que sur des données globales. De plus, si une personne est soupçonnée d'être atteinte lors de l'examen général, avec recherche de taches, etc., sa convocation pour un examen plus détaillé, si nécessaire, peut être faite avec la plus grande discrétion possible de façon à ce que personne n'en soit au courant. De plus, il faudrait obtenir l'assurance de la part des organismes sanitaires, des employeurs, etc. que les informations obtenues des sondages seront traitées de façon confidentielle et ne seront utilisées que pour des analyses statistiques, et cette assurance devrait être largement divulguée.

11.2 Examen incomplet des femmes

Il s'agit encore d'un phénomène assez fréquent car il peut être difficile et embarrassant, ou totalement inutile, de demander à des femmes de se dévêtir complètement. Cette difficulté peut être surmontée en s'adressant à des femmes (de préférence des femmes âgées ou des sages-femmes) pour les examens et en s'assurant que les régions cutanées normalement recouvertes par les vêtements, etc. sont examinées.

11.3 Ignorance de la maladie et méconnaissance de lésions discrètes

Les sondages devraient être précédés par une campagne intensive d'éducation sanitaire pour faire mieux connaître la lèpre à la population échantillon et pour lui montrer l'importance d'un diagnostic précoce des lésions. Pendant cette campagne d'éducation sanitaire, tous les individus devraient être encouragés à se présenter aux séances d'examen. Il faut faire comprendre l'importance de ces examens non seulement pour ceux ayant remarqué l'existence de lésions et qui soupçonnent la maladie, mais également pour ceux qui n'ont aucun doute, peut-être parce que certains signes leur ont échappé.

11.4 Refus de certaines classes de la population, comme des mendiants, d'être enregistrés

C'est une raison de considérer un sondage comme étant très différent d'une détection générale de cas dans un programme de lutte. La propagande initiale doit souligner le fait que les malades découverts ne recevront de traitement ou ne seront adressés au programme général que s'ils sont volontaires.

11.5 Attitude indifférente des agents de la lutte contre la lèpre

Là encore, il y a la possibilité d'éduquer les agents de la lutte contre la lèpre et peut-être d'obtenir la coopération d'un assistant social ou d'un membre du personnel sanitaire de la communauté.

11.6 Indifférence à l'égard du sondage

L'importance des sondages dans des buts de planification ou de surveillance n'est pas très bien comprise par la plupart des gens. Pour beaucoup, il s'agit d'un exercice purement académique. Tout programme d'information du public pour lui expliquer l'objectif d'un sondage doit être complété par l'assurance que tout cas découvert sera traité si nécessaire et qu'il en sera de même pour toute autre maladie. Il est extrêmement souhaitable que l'équipe transporte avec elle une réserve de quelques médicaments courants et peu coûteux et du matériel de premiers soins. Rien n'attire plus la sympathie du public et sa coopération que l'ancienne coutume du "guérisseur", c'est-à-dire la capacité de recevoir avec sympathie les plaintes formulées et si possible d'aider à les soulager.

11.7 Inaccessibilité de villages sélectionnés, disparition de villages figurant sur la liste de recensement, absence de travailleurs saisonniers

Ces problèmes sont communs à la plupart des enquêtes sanitaires et n'ont pas de lien particulier avec la maladie en question. Ils soulignent l'importance d'une organisation initiale pour vérifier les recensements, connaître les particularités saisonnières telles que les moussons qui rendent certains lieux inaccessibles. Les saisons de semence et de moisson, les jours de marché et de fête doivent être évités. En bref, le calendrier des visites doit être organisé pour s'assurer que la population échantillon sera bien présente dans les villages sélectionnés lors de la visite de l'équipe.

Le sondage ne doit pas finir trop tard dans la journée (par exemple 12 h-1 h de l'après-midi) pour permettre aux gens d'accomplir leurs tâches quotidiennes. Si le sondage doit réellement continuer toute la journée, la zone de l'échantillon doit être divisée en deux ou trois parties bien délimitées et le moment de la visite de chacune d'elles, clairement annoncé; si cette précaution n'est pas prise, de nombreuses personnes se fatigueront d'attendre et partiront. Les personnes temporairement absentes doivent être invitées à se présenter pour examen dans l'après-midi.

Il est possible que malgré toutes les précautions prises, un village sélectionné à partir d'une liste ancienne n'existe plus ou que pour une raison imprévisible un village ne soit pas accessible. Habituellement, un autre village est choisi au hasard pour compléter le programme de travail. Dans de telles situations, il faut admettre qu'un certain biais peut avoir été introduit. Un jugement doit être basé sur d'autres indices épidémiologiques.

Dans certains cas, comme dans la savane africaine, les hommes adultes seront absents en raison de leur travail, pour des périodes assez longues.

11.8 Recherche des personnes "clés"

Trouver une personne clé ayant une influence dans le village, obtenir sa collaboration et lui donner une certaine responsabilité, en lui faisant sentir l'importance de son rôle, permettra de s'assurer la coopération de la population. Le temps et les efforts nécessaires à soigner les relations avec une telle personne sont largement payants.

11.9 Une fois toutes les phases de préparation du sondage terminées, il sera très souhaitable d'entreprendre un présondage avant le début du sondage lui-même. Ce présondage permettra : de savoir si les formulaires d'enregistrement correspondent bien aux objectifs du sondage et, si nécessaire, de les modifier; de s'assurer que les méthodes à employer sont claires pour le personnel chargé du sondage; de revoir le manuel de sondage préparé pour ce personnel et de l'aider à parfaire sa formation.

Un tel présondage devra être effectué sur une population limitée, typique de la population sélectionnée mais pas dans les villages (ou les localités) en cas de populations urbaines. Ces présondages aideront non seulement à perfectionner les procédures statistiques mentionnées précédemment mais peuvent fournir des indications relativement précises sur les besoins en personnel et en matériel, la rapidité du travail, et des estimations globales du coût du sondage.

12. SONDAGES PÉRIODIQUES POUR ÉVALUATION

Le sondage de base détermine l'ampleur du problème sous tous ses aspects et aide à organiser les activités de lutte appropriées. Des sondages effectués périodiquement dans la région (disons tous les 3 ou 4 ans) permettront de surveiller l'efficacité du programme de lutte et de mettre en oeuvre des stratégies de remplacement, si celle qui a été choisie se montre insuffisamment efficace.

Les indicateurs suivants peuvent être choisis pour mesurer l'efficacité du programme :

- 1) distribution des cas par âge et par sexe;
- 2) proportion des cas non détectés (découverts pendant le sondage) par rapport au total des cas enregistrés;
- 3) proportion des cas actifs par rapport au total;
- 4) proportion des cas multibacillaires par rapport au total;
- 5) proportion de malades découverts au cours du sondage et qui prennent régulièrement leurs médicaments;
- 6) proportion de malades ayant des infirmités.

Un biais peut s'introduire dans les estimations si le même échantillon est utilisé pour chaque sondage périodique. De ce fait, il est suggéré d'utiliser un système d'échantillonnage par liste avec remplacement partiel. Dans ce système, une partie de l'échantillon initial est retenue et une autre est choisie. Un exemple en rendra l'application plus claire.

Supposons qu'une région rurale d'une population d'un million d'habitants ait été choisie pour un programme de lutte intense contre la lèpre. Vingt-cinq grappes (villages ou groupes de villages) d'une taille d'environ 500 personnes peuvent avoir été choisies pour le sondage de base. Dans le sondage suivant, à effectuer après, disons trois ans, un nouveau groupe de 15 grappes peut être sélectionné dans la même région. Celles-ci, avec les 10 grappes sélectionnées au hasard parmi les 25 premières, constitueront le nouvel échantillon.

L'expérience acquise dans le passé montre que lorsqu'un bon programme de lutte a commencé, la réduction de l'incidence se manifeste après des périodes de temps relativement courtes. Cependant, pour des raisons théoriques, et étant donné la longue période d'incubation de la maladie, la réduction de l'incidence chez les adultes peut sembler ne pas être proportionnée à l'intensité du programme de lutte contre la lèpre. La réduction chez les enfants, particulièrement les nouveau-nés, devrait être d'emblée substantielle et refléter directement les efforts de lutte. Cependant, l'incidence chez les enfants n'est pas élevée et il faudrait avoir recours à des méthodes autres qu'à des sondages de population. Des écoles et centres de santé maternelle et infantile sont des possibilités quand ces établissements couvrent raisonnablement bien la population de ces groupes d'âge.

REMERCIEMENTS

La rédaction de ce manuel a largement bénéficié des commentaires très utiles adressés par :

le Dr Jair Ferreira, Porto Alegre, Brésil; le Dr M. D. Gupte, Indian Council of Medical Research, Inde; le Dr K. Kersauze, Ecole nationale des Ingénieurs des Travaux ruraux, Strasbourg, France; le Dr M. Lechat, Ecole de Santé publique, Université catholique de

Louvain, Bruxelles, Belgique; le Dr D. L. Leiker, Institut tropical royal, Amsterdam, Pays-Bas; M. K. Uemura, Directeur, Division des Statistiques sanitaires, OMS, Genève; le Dr B. Zuiderhoek, Léprologue OMS, Ujung Pandang, Indonésie; le Bureau régional OMS pour les Amériques et le Bureau régional de l'OMS pour le Pacifique occidental.

BIBLIOGRAPHIE

1. Organisation mondiale de la Santé, Directives concernant la lutte contre la lèpre, Genève, 1980
2. Organisation mondiale de la Santé, Série de Rapports techniques N° 607, 1977 (Cinquième rapport du Comité d'experts de la lèpre)
3. Nations Unies, Département des Affaires économiques et sociales internationales, Bureau de statistiques. Recommendations for the preparation of sample survey reports (Provisional issue). Statistical Papers, Series C, N° 1, Rev. 2 (New York, 1964)
4. Fisher, R. A. & Yates, F. Statistical tables for biological, agricultural and medical research, 6th ed., revised and enlarged. Edinburgh, Oliver and Boyd, 1963, p. 137
5. Martinez Dominguez, V. et al. Epidemiological information on leprosy in the Singu area of Upper Burma. Bulletin de l'Organisation mondiale de la Santé, 58 : 81-89 (1980)

FORMULES POUR LES ERREURS TYPES - SONDAGE EN GRAPPES

I. Toutes grappes de dimension égale

a = nombre de grappes

b = taille de la grappe

y_i = nombre de sujets avec la maladie dans la grappe i

f = fraction de l'échantillon*

$r_i = y_i/b$

L'estimation de la proportion de sujets avec la maladie $r = \frac{\sum y_i}{a \cdot b}$

$$v = \text{variance de } r = \frac{(1-f)^*}{a(a-1)b^2} \left\{ \sum y_i^2 - \frac{(\sum y_i)^2}{a} \right\}$$

e.t = erreur type de r = \sqrt{v}

Les limites de confiance à 95 % de r sont $r \pm 1,96$ (e.t)

II. Grappes inégales

a = nombre de grappes

x_i = taille de la grappe i

y_i = nombre de sujets avec la maladie dans la grappe i

r = estimation de la proportion positive (c'est-à-dire avec la maladie)

x = x_i = taille de l'échantillon total

$r = y_i/x_i$; f = fraction de l'échantillon*

$$v = \text{variance de } r = \frac{(1-f)^*}{x^2} \frac{a}{(a-1)} \left\{ r^2 \sum x_i^2 + \sum y_i^2 - 2r \sum x_i y_i \right\}$$

e.t = erreur type de r = \sqrt{v}

Les limites de confiance à 95 % de r sont $r \pm 1,96$ (e.t)

* Peut être ignoré si l'échantillon représente une petite proportion (par exemple moins de 5 %) de la population.