

LES SONDAGES

Spécialisé dans le traitement statistique des études cliniques, le Département Biométrie de FOVEA vous propose au travers de ce fascicule quelques notions simples concernant les sondages.

en pratique

Traitement statistique

DES ETUDES CLINIQUES

Traitement statistique

DES ETUDES CLINIQUES

INTRODUCTION

L'industrie Pharmaceutique a de plus en plus recours à des études d'observation appelées **observatoires**, soit dans un but marketing afin d'analyser le circuit de prise en charge par les différents acteurs de la santé d'un type particulier de patients, soit en recherche épidémiologique pour étudier l'apparition éventuelle ou l'évolution d'une maladie à travers le suivi d'un patient dans le temps.

Il s'agit alors d'une enquête épidémiologique d'observation au cours de laquelle il n'y a pas d'intervention humaine délibérée. On observe simplement la relation entre le facteur de risque et la maladie, ou bien entre le protocole de soins et la maladie.

Deux méthodes d'enquête sont alors disponibles : l'enquête exhaustive appelée **recensement** portant sur l'ensemble de la population ou bien l'enquête portant seulement sur un échantillon de la population appelée **sondage**.

Le recensement est plutôt réservé aux organismes officiels en raison de son coût et de la lourdeur de la logistique à mettre en place. Le sondage permet d'obtenir des informations sur une population importante pour un coût plus modéré, un délai raisonnable, un risque d'erreur minimale et une précision maîtrisée.

1 Définition du sondage

On appelle sondage une **méthode de sélection d'unités statistiques à étudier** telles que ces unités soient **représentatives de la population considérée**, méthode permettant de pratiquer des observations et des mesures pour en tirer **des conclusions sur la population entière**.

Cette méthode utilise le **calcul des probabilités** permettant ainsi de s'assurer qu'il existe un niveau mesurable de confiance dans la validité des résultats. La population d'où a été extrait l'échantillon est la **population source** ou **base de sondage**. La population sur laquelle on veut obtenir des informations et généraliser les observations rapportées au niveau de l'échantillon constitue la **population cible**.

Exemple : Si l'on souhaite connaître la prévalence du diabète chez les hommes de 35 à 55 ans en effectuant une étude chez les employés d'une grande entreprise, la population cible comprend tous les hommes de 35 à 55 ans et la population source est celle des hommes de 35 à 55 ans dans l'entreprise.

2 Types de sondage

Il existe **3 grands types de sondage**, en fonction des **modes de constitution de l'échantillon**.

■ **LES METHODES PAR "CHOIX RAISONNÉ" OU ÉCHANTILLONNAGE SUBJECTIF**, catégorie à laquelle appartient la méthode des quotas.

Il s'agit du choix d'un échantillon en s'aidant de la connaissance que l'on a au départ de la population. Ces méthodes reposent sur le principe que si certains individus sont, par certaines de leurs caractéristiques, "à l'image" de la population générale, les facteurs étudiés sur ces individus le seront aussi. L'échantillon est ainsi constitué de sujets censés refléter la population cible. Malgré leur caractère arbitraire, ces méthodes donnent en général **des résultats satisfaisants** et sont souvent utilisées par les **instituts de sondage politique**.

■ **L'ÉCHANTILLONNAGE "À PORTÉE DE MAIN"** C'est l'utilisation d'un échantillon que l'on a "sous la main" et qui de ce fait est **biaisé**.
Exemple : Sondage téléphonique

qui exclut les sujets qui n'ont pas le téléphone ou qui ne sont pas dans l'annuaire.

■ **LES SONDAGES ALÉATOIRES OU SONDAGES PROBABILISTES** dans lesquels sont tirés au sort un certain nombre d'individus qui, s'ils sont en nombre suffisant, peuvent être considérés comme représentatifs de la population dont ils sont issus (population source). **Plus l'effectif est important et plus la précision du sondage sera fine**.

3 Méthodes probabilistes de choix d'un échantillon

■ **SONDAGE ÉLÉMENTAIRE OU ÉCHANTILLONNAGE ALÉATOIRE SIMPLE** On tire au sort **directement** les individus qui constituent l'échantillon à partir de la population à étudier.

Cette méthode donne à chaque unité de la population une chance égale d'être inclu dans l'échantillon. Aucun contrôle n'est exercé sur la répartition géographique de l'échantillon, ce qui risque de donner **des échantillons peu représentatifs** car **mal répartis géographiquement**, mais **non biaisés**.

Dans une enquête médicale, les acteurs de santé risquent d'être regroupés (grandes villes). Le principe de l'échantillonnage aléatoire simple reste la base de toute technique d'échantillonnage.

■ **SONDAGE SYSTÉMATIQUE** Cette technique est souvent utilisée quand les éléments peuvent être rangés en ordre ou mis en liste d'une façon ou d'une autre. La liste est parcourue avec un pas égal à k (une unité prélevée toutes les k unités). Une bonne répartition géographique (en fonction de la densité) est assurée. C'est une méthode couramment employée en **Santé Publique**.

■ **SONDAGE STRATIFIÉ** Pour améliorer la précision d'un sondage, la population est divisée en sous-ensembles particuliers ou strates. **La population à l'intérieur d'une strate est considérée comme homogène et différente d'une strate à l'autre**. Un tirage au sort (élémentaire, à plusieurs degrés, en grappes) est réalisé à l'intérieur de chacune des strates constituées. Si les paramètres sont différemment représentés d'une strate à l'autre, dans un échantillon constitué par strates, chaque sous-ensemble est mieux représenté et l'erreur d'échantillonnage est ainsi réduite. Le sondage stratifié est fréquemment utilisé dans les **enquêtes vaccinales, les études animales et les problèmes d'environnement**.

■ **SONDAGE À PLUSIEURS DEGRÉS** Ce type de sondage est réalisé dans les enquêtes à grande échelle où une liste de toutes les unités serait trop importante. On procède à **plusieurs tirages au sort successifs**.

■ **SONDAGE EN GRAPPES** Le sondage en grappes est un cas particulier de sondage à deux degrés, la totalité des unités secondaires étant tirée dans les unités primaires sélectionnées.

■ **SONDAGE PAR UNITÉ DE SURFACE** Un système de quadrillage permet ainsi le dénombrement d'éléments comme des mouches, des moustiques ou des plantes.

■ **SONDAGE EN PLUSIEURS TEMPS** Utilisé quand on veut obtenir des informations supplémentaires, certains renseignements sont tirés d'un sous-échantillon. Par exemple, pour une enquête vaccinale, lorsque l'on veut avoir des renseignements à la fois par répartition géographique, socio-économique et par répartition sanitaire, avec possibilité de chevauchement entre les deux.

■ **SONDAGE EN GRAPPES** Le sondage en grappes est un cas particulier de sondage à deux degrés, la totalité des unités secondaires étant tirée dans les unités primaires sélectionnées.

■ **SONDAGE EN GRAPPES** Le sondage en grappes est un cas particulier de sondage à deux degrés, la totalité des unités secondaires étant tirée dans les unités primaires sélectionnées.

■ **SONDAGE EN GRAPPES** Le sondage en grappes est un cas particulier de sondage à deux degrés, la totalité des unités secondaires étant tirée dans les unités primaires sélectionnées.

■ **SONDAGE EN GRAPPES** Le sondage en grappes est un cas particulier de sondage à deux degrés, la totalité des unités secondaires étant tirée dans les unités primaires sélectionnées.

■ **SONDAGE EN GRAPPES** Le sondage en grappes est un cas particulier de sondage à deux degrés, la totalité des unités secondaires étant tirée dans les unités primaires sélectionnées.

■ **SONDAGE EN GRAPPES** Le sondage en grappes est un cas particulier de sondage à deux degrés, la totalité des unités secondaires étant tirée dans les unités primaires sélectionnées.

En effet, il est possible de considérer que les sujets constituant une grappe, par exemple les patients d'un même médecin, ont plus de critères en commun que des sujets pris au hasard. Ce phénomène, appelé **effet de grappe**, diminue la précision du sondage et doit être compensé par une augmentation de l'effectif étudié.

4 Calcul des effectifs

Détermination du nombre de sujets nécessaires (n) dans un sondage élémentaire

■ **POUR L'ESTIMATION D'UN POURCENTAGE** Le nombre de sujets à inclure dépend de **3 facteurs** :

- le pourcentage attendu (p) de sujets présentant un caractère donné
- la précision souhaitée (i) pour le pourcentage
- le risque de 1ère espèce (risque α), c'est-à-dire le risque que le pourcentage observé sur l'échantillon s'écarte du pourcentage de plus de 2 écart-types. En effet, n est donné par la formule :

$$n = \frac{e_a^2 \times pq}{i^2} \text{ ou } pq \times \left(\frac{e_a}{i}\right)^2$$

où n est le nombre à calculer
 E_a est l'écart réduit correspondant au risque consenti : par exemple, pour $\alpha = 0,05$, $E = 1,96$
p est le pourcentage approximatif du facteur étudié (la détermination exacte de p est précisément l'objet du sondage)
q est le complémentaire de p ($q = 1 - p$)
i est la précision souhaitée

■ **POUR L'ESTIMATION D'UNE MOYENNE** Dans la formule ci-avant, la variance pq est remplacée par S^2

TRAITEMENT STATISTIQUE des études cliniques

DES ETUDES CLINIQUES

Traitement statistique

DES ETUDES CLINIQUES

$$n = \frac{e_a^2 \times S^2}{i^2} \text{ ou } S^2 \times \left(\frac{e_a}{i}\right)^2$$

■ **CORRECTION DE LA TAILLE DE L'ÉCHANTILLON** Si la **taille de l'échantillon (n) est supérieure à 10% de la population totale (N)**, il faut employer la **formule de correction** suivante :

$$n_c = \frac{n}{1 + \frac{n}{N}}$$

où n_c est la taille corrigée de l'échantillon
n est la taille de l'échantillon estimée en premier calcul
N est la taille de la population totale

5 Expression de l'estimation du paramètre étudié

L'estimation d'une variable qualitative ou d'une variable quantitative doit toujours être assortie du **calcul de l'intervalle de confiance (IC) à 95%**.

■ **POUR L'ESTIMATION D'UN POURCENTAGE**

$$p \pm e_a \sqrt{\frac{pq}{n}} \text{ ou } p \pm e$$

■ **POUR L'ESTIMATION D'UNE MOYENNE**

$$x \pm e_a \sqrt{\frac{S^2}{n}} \text{ ou } x \pm e$$

L'IC est déterminé par la valeur de e qui en détermine les bornes ($\pm e$). **La précision est d'autant plus grande que l'IC est petit**.

LES SONDAGES EN PRATIQUE

6 En conclusion

La technique des sondages ou d'échantillonnage basée sur la théorie probabiliste permet d'obtenir, dans un délai relativement court, sur un nombre restreint de sujets tirés au sort dans un certain type de population, des informations sur cette population. La technique des sondages se décline en un certain nombre de procédures toutes basées sur le **tirage aléatoire**, du **tirage aléatoire simple** où les sujets sont tirés au sort directement dans l'ensemble de la population au **tirage stratifié** où les sujets sont tirés de façon aléatoire dans chacune des strates constituant l'échantillon ou **tirage à plusieurs degrés** où l'on tire successivement des unités primaires (UP) puis, dans chacune de ces unités primaires, des unités secondaires (US) qui feront l'objet de l'étude.

Le **sondage en grappes**, très utilisé dans ses observatoires par l'industrie Pharmaceutique, est un **cas particulier des sondages à plusieurs degrés**. C'est un sondage à 2 degrés où les unités primaires (UP) sont les médecins et les unités secondaires (US) les patients qui sont l'objet de l'enquête. **La qualité de l'échantillon** constitué dépend de sa **bonne représentativité**, dépendante elle-même d'une **bonne répartition géographique**, d'une **précision optimale** et d'un **risque d'erreur minimal**. **La précision d'un sondage** dépend de la **taille de l'échantillon**. Pour avoir une

précision 10 fois meilleure, il faut multiplier la taille de l'échantillon par 100. La précision recherchée doit également être en rapport avec le paramètre à évaluer. **Plus la valeur de ce paramètre est faible, plus la précision doit être grande**. Dans un observatoire, la taille de l'échantillon dépend de la prévalence dans la population du critère retenu et par conséquent du choix des critères pris en compte et du niveau souhaité de la représentativité de l'échantillon par rapport à la population étudiée.

7 Bibliographie

1. **Bouyer J., Hemon D., Cordier S., Derriennic F., Stücker I., Stengel B., Clavel J.**

Epidémiologie – Principes et méthodes quantitatives
Les Editions INSERM 1993

2. **Ferley JP.**

Présentation d'une méthode d'échantillonnage pour une enquête en population générale
CAREPS – Grenoble
Septembre 1986 - IDEA

3. **Peavy V.**

Enquêtes et sondages
CDC Juillet 1986 - IDEA

4. **Rumeau-Rouquette C., Bréart G., Padiou R.**

Méthodes en épidémiologie
Flammarion – Médecine-Sciences

