



# L'épidémiologie pour tous

## Nombre de sujets nécessaire pour une enquête descriptive à visée qualitative

Rédigée par B. Toma, B. Dufour, E. Garin, J. Rivière et J.J. Bénet ; validée par le bureau de l'AEEMA

Dans certains cas, l'objectif d'une enquête d'épidémiologie descriptive est, non pas d'estimer le pourcentage de prévalence d'une maladie dans une population, mais de vérifier sa **présence** ou, au contraire, son **absence** par rapport à un **seuil de prévalence donné**. Autrement dit, l'objectif est non pas *quantitatif* (quantifier la prévalence), mais *qualitatif* (situer le niveau de prévalence par rapport à ce seuil).

La présente fiche traite le cas d'une enquête à visée qualitative.

### Problématique

#### Prouver la présence d'une maladie

Prouver la **présence d'une maladie** dans une population est relativement **facile**, surtout si sa prévalence est élevée. Il suffit de disposer d'un test de dépistage suffisamment spécifique (pour éviter les réponses faussement positives) et suffisamment sensible (pour éviter les réponses faussement négatives). Dès lors que la maladie n'est pas rare, un échantillon représentatif, de taille même modérée, conduit à une forte probabilité d'avoir au moins un sujet fournissant une réponse positive au test de dépistage (un « vrai positif », si le test est spécifique).

Intuitivement, on comprend que plus la prévalence de la maladie dans la population est élevée, plus facile sera sa détection avec un échantillon représentatif :



Le **nombre de sujets nécessaire** pour obtenir au moins une réponse positive au sein de l'échantillon est **inversement proportionnel au pourcentage de prévalence**.

Il est possible de calculer le nombre nécessaire de sujets dans un échantillon représentatif pour avoir de grandes chances que dans l'échantillon figure au moins un sujet atteint. Ce nombre est ainsi directement fonction du niveau de prévalence que l'on se fixe comme devant être détecté (au moins un sujet atteint dans l'échantillon) avec une probabilité élevée.

#### Démontrer l'absence d'une maladie

Essayer de démontrer **l'absence d'une maladie** dans une population est **plus difficile**, surtout si son expression clinique est rare et protéiforme. A la limite, il serait nécessaire de soumettre à un test de dépistage sensible et spécifique la totalité des sujets de la population pour pouvoir affirmer, en l'absence de réponse positive, que la maladie n'a pas été détectée parmi elle.

Lorsque la population est composée de quelques unités ou dizaines d'unités, ce dépistage exhaustif est envisageable ; mais lorsqu'elle est de très grande taille, il entraînerait un coût prohibitif, d'autant qu'à tout moment un risque d'introduction de l'agent pathogène peut survenir et qu'il faudrait donc,

périodiquement, procéder à un nouveau dépistage. L'enquête portera donc sur un échantillon de la population d'intérêt :



*Le **nombre de sujets** nécessaire est **d'autant plus grand** que le **niveau de prévalence** que l'on veut être capable de détecter est **faible**.*

## Détermination du nombre de sujets nécessaire

La démarche ressemble à celle de l'approche quantitative (cf. Fiche [Nombre de sujets nécessaire pour une enquête descriptive à visée quantitative](#)) : il existe une **interdépendance** entre le **seuil de prévalence fixé**, que l'on souhaite détecter, le niveau du **risque acceptable** d'erreur par défaut et le **nombre de sujets** nécessaire.

Le choix du niveau de prévalence limite à détecter (« le seuil ») et du risque acceptable d'erreur par défaut conditionne donc le nombre de sujets nécessaire dans l'échantillon.

Pour déterminer le nombre de sujets nécessaire dans l'échantillon à étudier, il faut donc fixer :

- le pourcentage de prévalence limite (le « seuil ») que l'on veut pouvoir détecter,
- et le risque d'erreur par défaut que l'on accepte (risque de ne pas avoir au moins un sujet atteint au sein d'un tel échantillon).

### Fixation d'un seuil de prévalence

Il convient de définir un *pourcentage de prévalence limite* qui correspond au seuil de sensibilité devant être atteint par le protocole d'échantillonnage.



*Autrement dit, le protocole d'échantillonnage doit permettre d'avoir **au moins un sujet atteint** de la maladie recherchée, au sein de l'échantillon.*

Pour une maladie à faible contagiosité (ou transmissibilité), il est possible d'admettre un très faible pourcentage de prévalence (par exemple, 0,2 %) comme un indicateur d'approche de la disparition

espérée de cette maladie et, par conséquent, d'utiliser des termes évoquant le caractère « indemne » de cette maladie pour la population. Bien sûr, pour une maladie à contagiosité (ou transmissibilité) élevée, cette démarche n'est pas possible car la présence de foyers, même en petit nombre, entraînerait un risque élevé de diffusion.

### Définition du niveau du risque accepté d'erreur par défaut

Intuitivement, on comprend que plus le niveau du risque d'erreur par défaut accepté est fixé bas, plus le nombre de sujets à étudier sera élevé. Habituellement, le choix de ce risque se situe à 5 % ou à 1 %.



*Si par exemple, un **risque d'erreur par défaut de 5 %** a été retenu, ceci signifie que pour une **population** dans laquelle le **pourcentage de prévalence est du niveau fixé** comme seuil à détecter, **5 fois sur 100 on peut ne pas avoir d'unité malade dans l'échantillon** dont la taille permet normalement, **95 fois sur 100, d'en comprendre au moins un**.*

### Détermination du nombre de sujets nécessaire

Lorsque le pourcentage de prévalence seuil a été déterminé et le niveau du risque d'erreur par défaut, fixé, il est possible, selon une démarche semblable à celle de l'enquête quantitative (cf. Fiche [Nombre de sujets nécessaire pour une enquête descriptive à visée quantitative](#)), de consulter le tableau de la page 208 du livre *Epidémiologie appliquée* pour connaître le nombre de sujets nécessaire correspondant (ou d'utiliser le tableur de l'annexe 15 « [Tableur pour calculer le nombre de sujets nécessaire pour une approche qualitative](#) » du site de l'AEEMA).

Comme dans la démarche quantitative, ce tableau donne les nombres nécessaires de sujets pour une population « infinie ».

Il faut alors vérifier si la situation correspond ou non à ce que l'on appelle une « **population infinie** ».



Lorsque la taille de l'échantillon est faible par rapport à la population d'intérêt, on considère qu'il s'agit d'une population infinie. **Le seuil a été fixé à 10 %.**

Si le rapport (appelé « *fraction de sondage* ») : *effectif nécessaire de sujets dans l'échantillon / effectif de la population* est **inférieur à 10 %**, le nombre lu dans le tableau de la page 208 est celui des unités à introduire dans l'échantillon.

Si ce rapport est **supérieur à 10 %**, le nombre de sujets à introduire est inférieur à celui de ce tableau et, ce, d'autant plus que le pourcentage du rapport est élevé (c'est-à-dire que le nombre de sujets dans l'échantillon se rapproche de celui de la population totale) : il est lu dans les tableaux des pages 209 et 210. Ceci permet de réduire un peu le nombre d'unités à prendre dans l'échantillon.

**Nota bene 1 :** les valeurs trouvées dans les tableaux cités ci-dessus ont été calculées pour un test de dépistage de sensibilité parfaite. Si la sensibilité est inférieure à 1, il convient de corriger ces valeurs en les divisant par la valeur de la sensibilité.



Exemple : nombre de sujets nécessaire (issu du tableau) : 100 ; sensibilité du test : 0,9 ; nombre ajusté de sujets nécessaire en tenant compte de la sensibilité du test :  $100 / 0,9 = 111$ .

**Nota bene 2 :** si l'on introduit dans l'échantillon un nombre de sujets supérieur à celui lu dans le tableau adéquat, on augmente la probabilité de détecter la présence de la maladie d'intérêt ; si on diminue ce nombre, on diminue alors cette probabilité.

## Interprétation des résultats obtenus avec le nombre de sujets nécessaire

Si, après l'étude, une ou plusieurs unités ont fourni une **réponse positive**, la **présence de la maladie a été démontrée**, tout en évitant d'avoir à étudier un nombre excessif de sujets.

Si au contraire, après l'étude, **aucune unité n'a fourni de réponse positive**, ceci signifie soit que la **maladie est réellement absente**, soit, si elle est présente, que son **pourcentage de prévalence a de grandes chances d'être inférieur au seuil ayant été défini comme objectif de capacité de détection** permis par la taille de l'échantillon. Toutefois, à cause des inévitables fluctuations d'échantillonnage dues au hasard, il existe un risque pour qu'au sein de l'échantillon de taille pourtant adaptée au pourcentage de prévalence à détecter, aucun sujet atteint ne figure alors que le pourcentage de prévalence dans la population est au niveau du seuil fixé : il s'agit du **risque d'erreur par défaut** (ne pas reconnaître la présence d'un élément qui existe) (cf. Fiche [Erreurs](#)) qu'il est nécessaire de prendre en considération.

Les éléments conditionnant le nombre de sujets nécessaire dans une enquête descriptive, quantitative ou qualitative, sont indiqués dans le tableau 1.

Tableau 1. **Éléments interdépendants en relation avec le nombre de sujets nécessaire pour une enquête descriptive, quantitative ou qualitative**

<b>Approche quantitative</b>	<b>Approche qualitative</b>
Prévalence supposée de la maladie (ou prévalence attendue)	Pourcentage de <i>prévalence</i> à détecter (prévalence limite = seuil)
Précision relative du résultat	Niveau d' <i>erreur par défaut</i> acceptable (risque d'erreur bêta)
<b>Nombre de sujets nécessaire</b>	<b>Nombre de sujets nécessaire</b>

Dans une **enquête descriptive à visée qualitative**, le **nombre de sujets nécessaire** dans l'échantillon est conditionné, d'une part, par le **pourcentage de prévalence à détecter** (seuil de prévalence), d'autre part, par le **risque d'erreur par défaut acceptable**.



Plus le **pourcentage de prévalence à détecter est faible**, plus élevé est le **nombre de sujets nécessaire**, pour un même niveau de risque d'erreur.

De même, pour un pourcentage donné de prévalence à détecter, le **nombre de sujets nécessaire est d'autant plus élevé que le risque d'erreur par défaut jugé acceptable est faible**.

Après l'étude :



- Si **une ou plusieurs** unités ont fourni une **réponse positive**, la **présence de la maladie a été démontrée**.
- Si **aucune** unité n'a fourni de **réponse positive**, ceci signifie :
  - **soit la maladie est réellement absente,**
  - **soit, elle est présente mais son pourcentage de prévalence a de grandes chances d'être inférieur au seuil de détection défini.**

Toma B. et al. : *Epidémiologie appliquée*, 2018, AEEMA éditeur, 614 p.

Annexe : « [Tableur pour calculer le nombre de sujets nécessaire pour une approche qualitative](#) »

Terminologie : des liens hypertextes (en bleu souligné) vous permettent d'accéder aux définitions sur le site de l'AEEMA

Fiches AEEMA : [Nombre de sujets nécessaire pour une enquête descriptive à visée quantitative ; Erreurs](#)

