



Titre : Biais dans les enquêtes épidémiologiques descriptives

Intervenant : Marianne Savès

Bonjour à Tous,

Aujourd'hui, nous allons nous intéresser aux biais dans les enquêtes épidémiologiques descriptives.

Commençons par définir ce qu'est un biais.

En épidémiologie, un **biais** est une différence entre ce qui est mesuré à partir des données de l'échantillon et la valeur exacte dans la population, cette différence étant liée à une erreur systématique. Le calcul réalisé à partir des données d'un échantillon pour estimer la valeur dans la population s'appelle « l'estimation ».

Dans le cas spécifique de l'**épidémiologie descriptive**, l'objectif est de décrire un phénomène de santé en termes de fréquence, dans un groupe de sujets en fonction des caractéristiques de ces sujets, du temps et du lieu. La mesure étudiée est donc ici la **fréquence du phénomène de santé dans la population**, et une enquête comportant des **biais** conduira donc à une **estimation erronée** de cette fréquence. On parlera de « résultats biaisés ».

Des **biais** peuvent survenir à **différentes étapes** d'une enquête descriptive :

- lors de la sélection des sujets : on parle de « biais de sélection »
- lors du recueil des informations : on parle dans ce cas de « biais d'information ».

Nous n'aborderons pas ici le cas des biais de confusion qui posent problème lors de l'analyse.

Nous allons approfondir tout d'abord la notion de **biais de sélection**.

L'échantillon sur lequel on va pouvoir estimer la fréquence du phénomène de santé résulte de plusieurs étapes de sélection.

Pour illustrer cela, nous allons prendre l'exemple de l'Etude nationale nutrition santé, l'ENNS. Dans cette étude transversale, l'objectif principal était de décrire les apports alimentaires et l'état nutritionnel des adultes et des enfants vivant en France métropolitaine en 2006-2007. La population-cible était donc représentée par les adultes et les enfants vivant en France métropolitaine en 2006-2007.

Pour que les données le concernant soient prises en compte dans l'analyse et contribuent à décrire les apports alimentaires selon l'âge et le sexe, un sujet devait être avant tout sélectionnable, c'est-à-dire habiter dans un domicile fixe et être accessible par téléphone. Puis, il devait être le sujet sélectionné dans l'un des foyers sélectionnés au sein d'une des communes sélectionnées. Ensuite, il devait avoir été contacté, pouvoir communiquer en français, avoir accepté de participer à l'enquête, et enfin avoir répondu aux questions concernant son âge, son sexe et ses apports alimentaires.

<http://invs.santepubliquefrance.fr/Dossiers-thematiques/Maladies-chroniques-et-traumatismes/Nutrition-et-sante/Enquetes-et-etudes/ENNS-etude-nationale-nutrition-sante>.

Dans une enquête épidémiologique descriptive, il y a biais de sélection si les sujets pris en compte dans l'analyse sont **non représentatifs** de la population **vis-à-vis du phénomène de santé** ; c'est-à-dire s'ils présentent plus fréquemment le phénomène de santé que la population (on a alors une sur-estimation de la fréquence du phénomène de santé), ou moins fréquemment le phénomène de santé (ce qui conduit à une sous-estimation de la fréquence).

Dans l'exemple de l'ENNS, la question à laquelle il est essentiel de pouvoir répondre est : est-ce que les apports alimentaires observés chez les sujets de l'échantillon finalement pris en compte dans l'analyse sont similaires à ceux de la population ? Si c'est le cas, l'estimation de la description des apports alimentaires faite à partir de l'échantillon est valide. Sinon, l'estimation sera erronée du fait d'un biais de sélection.

Mais, quelles sont les **sources potentielles** de biais de sélection dans une étude épidémiologique descriptive ? On voit bien qu'à chaque étape de sélection, un certain nombre de sujets appartenant à la population peuvent être exclus, et que cette exclusion peut éventuellement engendrer un biais de sélection. Nous allons présenter deux sources de biais de sélection rencontrées très fréquemment.

La première source correspond aux **sujets non-répondants**. Il s'agit de sujets sélectionnés pour participer à l'enquête mais n'y participant finalement pas, ou alors des sujets y participant mais ne répondant pas à toutes les questions dont celles permettant de répondre à l'objectif de l'enquête, et devant donc être retirés de l'analyse. Le problème des non réponses concerne tous les schémas d'étude, transversal ou longitudinal.

Dans l'exemple de l'ENNS, parmi les 5 217 sujets adultes composant l'échantillon prévu, 3 141 sujets ont finalement participé à l'enquête, et, parmi ces participants, 3 115 sujets ont répondu aux questions concernant les caractéristiques socio-démographiques et les apports alimentaires (soit 60% des sujets). Il y a donc eu 40% de non-répondants ((5217-3115)/5217=40%) en ce qui concerne l'objectif principal de l'enquête.

La seconde source correspond aux **sujets perdus de vue**. On désigne ainsi les sujets ayant participé à l'enquête, au moins initialement, mais pour lesquels on n'a plus de nouvelles à un moment du suivi prévu. Ce cas de figure ne concerne donc que les enquêtes longitudinales.

Prenons l'exemple de l'Etude NutriNet-Santé, qui est une enquête longitudinale, dans laquelle il est prévu de suivre 500 000 sujets pendant 5 ans, à partir de questionnaires remplis en ligne chaque année, portant sur l'alimentation, l'activité physique, la santé, etc.

<https://www.etude-nutrinet-sante.fr/>

La difficulté dans ce type d'enquête est de motiver les sujets à participer aux différents suivis programmés dans le cadre de l'enquête. Et on imagine aisément qu'un certain nombre des sujets ayant initialement répondu aux questionnaires de l'Etude Nutrinet Santé pourront, pour raison de santé, lassitude, ou toute autre raison, ne pas continuer à participer à l'enquête à un moment du suivi.

Enfin, comment prévenir les biais de sélection ?

Il faut agir en amont de la réalisation sur le terrain de l'enquête, en prévoyant dans le protocole :

- quel que soit le type d'enquête :
 - o des mesures incitant les sujets sélectionnés à participer à l'enquête,
 - o des procédures de relance pour inciter les sujets n'ayant pas répondu initialement à participer finalement à l'enquête,
 - o des outils simples de recueil pour ne pas décourager les sujets ayant accepté de participer, et les inciter à répondre à toutes les questions posées,
- et spécifiquement dans les enquêtes longitudinales :
 - o des procédures de relance des participants à chaque suivi,
 - o des moyens pour fidéliser ces participants.

Toutes ces actions relèvent d'une stratégie de communication à bien déterminer et spécifier dans le protocole de l'enquête avant sa mise en œuvre.

Après avoir étudié les biais de sélection, nous allons aborder à présent les **biais d'information**.

Une fois que les sujets contactés dans le cadre de l'enquête ont accepté d'y participer, on passe à l'étape de recueil des informations. Cette étape est fondamentale puisque tous les résultats issus de l'enquête dépendent de sa qualité.

Un biais d'information survient dans une enquête épidémiologique descriptive lorsque le **classement des sujets de l'échantillon vis-à-vis du statut du phénomène de santé** étudié est **erroné**. On parle aussi de « **biais de classement** » ou de « **biais de mesure** ». C'est-à-dire que l'on va considérer que certains sujets présentent ce phénomène de santé alors qu'ils ne le présentent pas en réalité (on parle alors de « faux positifs »), ou qu'*a contrario*, certains sujets ne présentent pas le phénomène de santé alors qu'ils en sont en réalité atteints (on parle dans ce cas de « faux négatifs »). Dans le premier cas, cela conduit à une sur-estimation de la fréquence du phénomène de santé puisque certains sujets seront détectés comme présentant le phénomène de santé alors qu'ils en sont indemnes, et dans le second, à une sous-estimation, puisque certains sujets présentant le phénomène de santé ne seront pas détectés.

L'erreur de classement est liée à une erreur lors du recueil de l'information concernant le phénomène de santé, venant soit de l'outil de recueil, soit du mode de recueil retenu, et donc de l'enquêté ou de l'enquêteur.

Quelles sont alors les **sources potentielles** de biais d'information dans une étude épidémiologique descriptive ? Nous allons aborder ici les sources les plus fréquentes au travers d'exemples, et indiquer des solutions pour **prévenir les biais d'information**.

Intéressons-nous tout d'abord aux questionnaires. On peut obtenir des réponses erronées si l'enquêté rencontre des **difficultés de compréhension** vis-à-vis des questions, du fait du **contexte culturel** ou de la **langue** utilisée.

Reprenons l'exemple de l'ENNS. Il a été décidé de ne pas enquêter auprès des sujets qui ne comprenaient pas la langue française. On imagine bien que ce choix vise à limiter les biais d'information qui seraient liés à un défaut de compréhension dû à la langue.

Il faut insister sur le fait qu'un questionnaire doit être testé préalablement si possible en situation réelle, afin de juger notamment de son adéquation au contexte de l'enquête.

Très souvent, les questionnaires font appel à la **mémoire** des sujets, ce qui peut entraîner également un biais d'information si le sujet donne une réponse erronée du fait d'un oubli. Ce biais d'information particulier est appelé « **biais de mémorisation** ».

L'ENNS comportait une enquête alimentaire reposant sur des questionnaires de type rappel des 24 heures. Le principe est de demander au sujet de décrire la nature et la quantité de tous les aliments et boissons consommés la veille de l'entretien. Afin d'aider le sujet à s'en souvenir, le rappel est fait en laissant d'abord le sujet se rappeler spontanément de sa consommation, puis en revenant sur des consommations qui sont souvent oubliées (par exemple les sucreries), afin de réduire le plus possible le biais de mémorisation. En outre, le diététicien s'appuie sur un manuel validé de photographies des portions.

Une autre difficulté fréquemment rencontrée concerne les questions qui abordent des **thèmes sensibles**, tels que les pratiques sexuelles ou la consommation d'alcool par exemple. Cela peut conduire l'enquêté à répondre sciemment de façon erronée. Vous retrouverez souvent ce type de biais d'information sous le nom de **biais de prévarication**.

Par ailleurs, par rapport à des questions ramenant à des **normes sociales**, l'enquêté peut avoir tendance à donner la réponse qui lui paraît projeter l'image la plus valorisante pour lui, ce qui le conduira là aussi à donner une réponse erronée. On parle dans ce cas de **biais de désirabilité sociale**.

Si on s'intéresse à l'activité physique par exemple, certains sujets enquêtés pourront répondre qu'ils pratiquent une activité physique régulière alors que ce n'est pas le cas, si la pratique d'une activité physique régulière est valorisée socialement.

Les **biais** d'information illustrés ci-dessus sont **dus à l'enquêté** qui a fait le choix, volontaire ou non, de donner des réponses discordantes par rapport à l'information « réelle ». D'autres peuvent être **liés à l'enquêteur**. Ainsi, en cas de questionnaire très ouvert, ou pour lequel des consignes pour la passation (c'est-à-dire la façon de faire passer le questionnaire à l'enquêté) n'ont pas été clairement données, et surtout dans le cas d'enquêtes impliquant plusieurs enquêteurs, le risque est que l'enquêteur oriente ou interprète les réponses des enquêtés en fonction de ses propres opinions ou connaissances.

Par exemple, l'enquêteur aborde avec l'enquêté la question de l'activité physique, et celle-ci est graduée en activité 'nulle'/'faible'/'modérée'/'intense'. En cas d'hésitation de l'enquêté entre deux modalités et sans consigne précise donnée en amont par l'équipe responsable de l'enquête, l'enquêteur pourrait choisir de noter celle des deux indiquant l'activité la plus intense si l'enquêté lui paraît « sportif »...

Afin de limiter ces biais, il est indispensable de standardiser le plus possible les questionnaires, c'est-à-dire que tous les sujets doivent être enquêtés à partir d'un même questionnaire, mais également de standardiser la passation du questionnaire, grâce à la définition des concepts pouvant poser problème, à la rédaction d'un guide de l'enquêteur et à la formation du ou des enquêteurs, afin de garantir un maximum de neutralité et d'homogénéité.

Dans l'ENNS, les informations concernant l'activité physique étaient recueillies à partir d'un questionnaire utilisé internationalement, et dans lequel les différents degrés d'intensité de l'activité physique (modéré, intense) sont détaillés.

La nécessité de standardiser l'outil et la façon de l'utiliser est vraie quel que soit cet outil.

Ainsi, dans l'ENNS, des mesures anthropométriques, comme la mesure de la taille et du poids, de même que la mesure de la pression artérielle, étaient proposées. Chez les adultes, elles étaient réalisées dans un centre d'examen de santé de l'Assurance maladie, ou à domicile par un infirmier. Il était donc essentiel de prévoir dans le protocole des consignes afin que ces mesures soient réalisées de la même manière pour tous. Vous voyez ici l'exemple du protocole concernant la mesure du poids.

Enfin, un outil quel qu'il soit (questionnaires, mesures anthropométriques, dosages biologiques, etc.), même parfaitement standardisé, peut être imparfait, et ainsi induire des classements erronés des enquêtés. On parlera de problème de **sensibilité** si l'outil classe des sujets réellement atteints par le phénomène de santé comme en étant indemnes, et de problème de **spécificité** si l'outil classe des sujets réellement indemnes comme en étant atteints. Il faut donc toujours veiller à choisir les outils les plus performants compte tenu du contexte de l'enquête (donc en fonction du coût, de la disponibilité, du temps).

Cette présentation des biais dans les enquêtes épidémiologiques descriptives s'achève. Beaucoup de noms différents sont donnés à ces biais, mais on peut toujours les regrouper en biais de sélection ou biais d'information. Retenez qu'ils sont un sujet de préoccupation majeur pour les épidémiologistes, et que la meilleure façon de les prévenir est d'avoir prévu des mesures spécifiques dans le protocole de l'enquête.

A très bientôt !