

CHAPITRE 14

LES OUTILS DE MESURE Instruments et Questionnaires

Anne-Marie Schott, François Chapuis, Gilles Landrивon, Eric-Nicolas Bory, François Delahaye

La formulation de la question de recherche nécessite une bonne définition des facteurs étudiés et des critères de jugement. Ces derniers vont définir, dans le cadre du protocole, les données qui seront collectées et leur mode de recueil. La collecte des données fait intervenir des mesures et des questionnaires.

L'évaluation des instruments de mesure est essentielle à un stade précoce de la recherche. En effet, les défauts d'un instrument non seulement ont des conséquences sur la puissance de l'étude et la taille de l'échantillon nécessaire, mais ils peuvent compromettre aussi la validité même des conclusions de l'étude en introduisant des erreurs irréversibles dans l'estimation des variables (facteurs étudiés, critères de jugement, facteurs de confusion).

Le questionnaire constitue un mode particulier de collecte des données. Initialement utilisées en sociologie, psychologie et psychiatrie, les technologies de questionnaires intéressent aujourd'hui l'ensemble des spécialités médicales.

De la qualité d'un questionnaire dépend la qualité de la transmission, du stockage et de l'interprétation des renseignements recueillis. A travers le questionnaire, l'investigateur effectue une mesure du phénomène étudié. Un mauvais questionnaire, ou des questions mal posées, peuvent induire des biais de mesure : il peut y avoir alors une sur- ou une sous-estimation de l'importance du phénomène observé, invalidant totalement les conclusions de l'étude.

Plan du chapitre

I - CHOIX D'UN INSTRUMENT DE MESURE

A - Définition de l'objectif

- 1 - Utilisation transversale dans un but diagnostique ou pronostique
- 2 - Utilisation longitudinale pour évaluer l'efficacité d'une intervention

B - Recherche des instruments existants

C - Définir les phénomènes à mesurer

II - EVALUATION DE LA REPRODUCTIBILITE D'UN INSTRUMENT NOUVEAU

A - Evaluation de la reproductibilité

B - Mesure de la reproductibilité

C - Importance de la reproductibilité

D - Importance de l'échantillonnage

E - Méthodes statistiques pour l'évaluation des instruments de mesure

1 - Principes généraux

2 - Types de données

a - Données qualitatives

b - Données ordinales

c - Données continues

III - SAVOIR EXPRIMER CE QUE L'ON SOUHAITE

A - Définition et validation du champ d'exploration

B - Les divers types de questions

1 - La question ouverte

2 - La question semi-ouverte

3 - La question fermée

C - Un vocabulaire approprié

D - Formulation de la question

E - Organisation des questions

F - Codage

G - Utilisation d'un questionnaire étranger

IV - EXACTITUDE ET REPRODUCTIBILITE

V - PRESENTATION DU QUESTIONNAIRE

VI - LE PRE-TEST

Conclusion

PREMIERE PARTIE : INSTRUMENTS DE MESURE

Toute mesure est entachée d'un certain degré d'erreur. Il est fondamental de connaître les différents types d'erreurs, leurs origines, la façon de les quantifier et de les réduire.

Il n'y a pas d'instrument de mesure universellement bon. Le choix d'un instrument, parmi ceux qui existent et qui ont déjà été utilisés, dépend de plusieurs éléments dont les trois suivants qu'il faut prendre en compte : **l'objectif de la recherche, la maladie étudiée et la population cible.**

I - CHOIX D'UN INSTRUMENT DE MESURE

A - Définition de l'objectif

1 - Utilisation transversale dans un but diagnostique ou pronostique

On peut vouloir mesurer un symptôme ou une caractéristique permettant de classer une population en plusieurs groupes de sujets (dépistage). Il s'agit de déterminer la différence entre les sujets, pour une variable donnée, de façon transversale.

Par exemple:

- pour connaître la prévalence d'une affection dans une population déterminée ou pour dépister une affection dans une population déterminée, en épidémiologie;
- pour décider au niveau individuel d'une intervention thérapeutique ou pour déterminer si un individu est éligible dans un essai, en clinique.

Ainsi, on utilise le frottis cervical pour le dépistage du cancer du col utérin et pour en évaluer la prévalence dans une population déterminée.

Certains examens sont destinés à évaluer le risque du développement d'une affection chez des sujets sains ou de l'aggravation d'une maladie chez des patients. Il s'agit dans ces cas de mesurer des facteurs de risque et de pronostic. Par exemple, on recherche plusieurs types d'anticorps chez des patients atteints d'hépatite B. La présence et le taux de ces anticorps permettent d'apprécier le potentiel évolutif de la maladie et ainsi d'en établir le pronostic.

2 - Utilisation longitudinale pour évaluer l'efficacité d'une intervention

Il s'agit de déterminer l'évolution d'un sujet de façon longitudinale en comparant deux états d'un même sujet. Le plus souvent l'évaluation se fait avant et après une intervention thérapeutique dont on désire évaluer l'efficacité.

B - Recherche des instruments existants

Cette étape est obligatoire pour tous les types de mesure.

Paradoxalement, lorsqu'on s'intéresse à un objectif précis pour une population déterminée, il devient très difficile de trouver l'instrument adéquat. On est alors tenté d'être très critique vis-à-vis des instruments existants et de sous-estimer la difficulté du développement d'un nouvel instrument. Une erreur fréquente des cliniciens est d'éliminer trop facilement les échelles existantes et de se lancer dans le développement d'un nouvel instrument avec l'idée optimiste qu'ils peuvent mieux faire. En fait, le développement d'un instrument de mesure nécessite un investissement considérable de temps et d'argent. L'étape de revue exhaustive de tous les instruments développés dans le domaine est donc obligatoire.

A l'inverse, l'autre erreur fréquente est de considérer qu'une échelle est bonne parce qu'elle a été "validée", sans tenir compte des objectifs et des conditions dans lesquelles la validation a été faite. Une échelle validée pour déterminer le niveau des fonctions cognitives chez des personnes âgées n'est certainement pas valide pour mesurer l'état cognitif chez des étudiants de 20 ans.

Une fois cette étape de recherche dans la littérature et auprès des experts terminée, il faut alors choisir parmi les instruments existants celui qui semble le plus adapté à l'objectif qu'on s'est fixé. Il faut pour cela déterminer leur pertinence par rapport à ce qu'on veut mesurer. Il faut également rechercher si l'exactitude et la reproductibilité de ces instruments ont été évaluées de façon correcte, dans un contexte identique et pour des objectifs similaires.

C – Définir les phénomènes à mesurer

Le phénomène que l'on veut mesurer peut être exprimé selon divers types d'échelle:

- Données qualitatives: elles sont exprimées en catégories qui ne peuvent être ordonnées les unes par rapport aux autres (couleur des yeux, catégories professionnelles...). Lorsqu'il n'existe que deux catégories possibles, les données sont dites dichotomiques ou binaires (statut vital, sexe).
- Données ordinales: à l'inverse des précédentes, ces données peuvent être classées par ordre croissant ou décroissant et combinées comme des nombres (par exemple, multipliées ou additionnées pour former des index).
 - Elles peuvent être discontinues (ou discrètes), si elles prennent seulement certaines valeurs entières, qui sont soit des catégories (différents stades de la dyspnée, grades successifs d'évolution d'une tumeur cancéreuse), soit des valeurs disposées régulièrement le long d'une échelle comportant un intervalle constant entre chaque valeur (nombre de crises d'épilepsie par mois, nombre d'articulations inflammatoires).
 - Elles peuvent être continues, si elles peuvent prendre virtuellement toutes les valeurs possibles entre les deux valeurs extrêmes de l'échelle de réponse (pression artérielle, poids). Les intervalles sont constants et connus.

Le choix du type d'échelle de mesure dépend des variables concernées et des objectifs. Dans un questionnaire destiné à recueillir des informations sur la consommation de calcium, on peut quantifier le lait consommé et construire un indice de consommation de calcium en mg par jour. Si l'objectif est simplement de partager les individus en groupes, on peut exprimer les résultats en catégories ordonnées (moins de 500 mg/j, de 500 à 1.000 mg/j, plus de 1.000 mg/j). Les catégories ont l'avantage d'avoir une signification pratique en clinique. Elles ont

cependant un caractère plus arbitraire et peuvent aboutir à méconnaître des différences significatives entre les groupes si elles n'ont pas été judicieusement choisies.

Il faut s'assurer que les variables qu'on mesure sont appropriées à l'objectif de la recherche. Pour cela on consulte un groupe d'experts, puis un échantillon de la population sur laquelle on désire appliquer l'instrument.

Une fois l'instrument choisi, il faut évaluer ses qualités dans le cadre des objectifs particuliers de la recherche envisagée.

II - EVALUATION DE LA REPRODUCTIBILITE D'UN INSTRUMENT

On distingue classiquement deux grands types de caractéristiques liées à un instrument de mesure: la reproductibilité et l'exactitude (tableau 1).

La reproductibilité d'un instrument est sa capacité à fournir une mesure identique de façon répétée (capacité du thermomètre à indiquer la même température de façon répétée, ...). La reproductibilité est essentiellement liée à l'erreur aléatoire. Plus l'erreur aléatoire est petite, plus la reproductibilité est bonne.

L'exactitude d'un instrument est sa capacité à fournir une mesure exacte du phénomène à mesurer (capacité du thermomètre à indiquer la température exacte, ...). L'exactitude est liée à la fois à l'erreur aléatoire et aux biais. L'exactitude (appelée encore « validité ») est définie par la sensibilité, la spécificité et est développée dans le chapitre IX.

A - Evaluation de la reproductibilité

Elle comporte les étapes suivantes:

- Enumérer les sources d'erreurs potentielles (inclinaison du rayonnement par rapport au corps du patient pour les radiographies, heure de la journée pour le poids ou la taille, saison pour l'évaluation de l'activité physique, conditions dans lesquelles se déroule la mesure de la pression artérielle, ...).
- Classer ces sources d'erreurs potentielles par ordre d'importance décroissante, en se fondant sur l'avis d'experts, les données de la littérature, et les conditions dans lesquelles on veut utiliser l'instrument. Dans un essai multicentrique, il est indispensable d'étudier la variation entre les centres au cours d'une étude pilote. En revanche, si l'étude envisagée a lieu dans un seul centre, cette source d'erreur n'existe pas.
- Déterminer si ces variations existent effectivement en pratique, et le cas échéant mesurer leur importance et tenter de les réduire au maximum.

Il est impossible d'établir ici une liste exhaustive de toutes les sources de variation possibles. Certaines sont spécifiques du type de mesure, d'autres existent, à des degrés variables, avec presque tous les types de mesures:

- la variation inter-investigateurs (variation entre des mesures du même phénomène par des investigateurs différents);

- la variation intra-investigateur (variation entre des mesures du même phénomène par le même investigateur à différents moments);
- la variation intra-patient (variation de la mesure ou du phénomène lui-même chez un même patient à deux moments différents).

Pour évaluer ces sources d'erreurs potentielles, on réalise des mesures répétées en faisant varier une seule source d'erreur à la fois. Il faut s'assurer qu'entre ces mesures répétées, la valeur réelle n'a pas changé. Ceci a une implication dans le choix du temps écoulé entre deux mesures répétées:

- si on mesure des variables labiles, il faut choisir un intervalle de temps très court entre les mesures successives afin d'éviter un réel changement de valeur;
- en revanche, pour des variables très stables telles que la taille à l'âge adulte ou la densité osseuse, le temps écoulé entre deux mesures peut être plus long, en conservant l'assurance qu'il n'y a pas eu de changement de la valeur réelle.

Prenons l'exemple du diagnostic radiologique de cancer pulmonaire. Pour évaluer la variabilité inter-investigateurs, chaque radiographie est lue par chacun des radiologues en insu et on mesure la concordance de leurs conclusions. La variabilité intra-investigateur est estimée en montrant plusieurs fois au même investigateur la même radiographie. La variabilité intra-patient est estimée en faisant plusieurs radiographies du même patient et en les faisant interpréter par un seul radiologue.

B - Mesure de la reproductibilité

Dans le domaine de la biologie, les laboratoires effectuent continuellement des mesures de reproductibilité en partageant les sérums en deux et en mesurant la variabilité de la mesure pour un même sérum. Cette variabilité est exprimée en écart-type autour de la moyenne. Par exemple, la reproductibilité pour le dosage du sodium est de $\pm 2,3$ mmol/l. Comme les valeurs normales sont comprises entre 130 et 150 mmol/l, il est facile de juger de l'acceptabilité de cette erreur.

En clinique, on est souvent amené à travailler avec des données discontinues. Supposons qu'on réalise des radiographies pulmonaires chez 110 sujets exposés pour déterminer la présence ou l'absence de signes de pneumoconiose. Les radiographies sont montrées successivement à deux radiologues et on examine dans quelle mesure leurs avis sont concordants (tableau 2). Dans cette table 2x2, le pourcentage de concordance entre les radiologues est représenté par la case en haut à gauche (présent-présent) et celle en bas à droite (absent-absent). Il est de $\frac{14 + 81}{110}$, soit 86 %.

Il peut y avoir plus de deux catégories. Prenons l'exemple de 100 femmes atteintes de cancer du sein examinées indépendamment par deux cliniciens qui doivent attribuer un degré de sévérité sur une échelle de I à IV selon des critères prédéfinis (tableau 3). La concordance parfaite entre les cliniciens est: $\frac{25 + 14 + 17 + 10}{100}$, soit 66 %, ce qui ne paraît pas très bon.

Cependant, lorsqu'on étudie les 34 cas dans lesquels les cliniciens ne sont pas d'accord, la différence n'est le plus souvent (27 fois) que d'un stade.

Pour ce type de données à plus de deux catégories de réponse, la concordance parfaite n'est pas un très bon reflet de la situation, car elle ne donne aucune indication sur l'importance de la discordance. Elle est également trop dépendante du nombre de catégories: plus celui-ci est grand, plus les chances de parfaite concordance diminuent.

Enfin, le simple calcul du pourcentage de concordance ne tient pas compte des cas où la concordance est due au hasard. Pour pallier ces inconvénients, plusieurs index statistiques de concordance ont été proposés (*cf infra*).

Ces index de concordance ne peuvent être appliqués que dans des études simples et lorsqu'on utilise des variables discontinues. Lorsque le plan d'étude est plus complexe, qu'il y a plus de deux observateurs concernés ou que les sources de variation sont plus nombreuses, il est préférable d'utiliser des méthodes mathématiques adaptées, permettant d'étudier plusieurs sources de variation à la fois.

Ces méthodes, fondées sur l'analyse de variance, évaluent la part de la variation liée aux différences réelles - soit entre les sujets, soit chez un même sujet après un traitement ou un intervalle de temps suffisamment long - et la part de la variation due aux erreurs de mesure.

On utilise également les principes de l'analyse de variance lorsque les variables étudiées sont continues. De ces méthodes sont dérivées des statistiques utilisées spécifiquement pour l'évaluation des instruments de mesure.

Une fois les sources d'erreurs identifiées et leur importance évaluée, il s'agit de les réduire. Dans l'exemple précédent, on peut réduire l'erreur due aux variations inter-investigateurs en n'utilisant qu'un seul radiologue pour lire toutes les radiographies ou encore en réalisant une formation identique de tous les radiologues. On peut diminuer l'erreur due à la variation intra-patient en appliquant des règles et des standards stricts pour la réalisation des radiographies (positionnement des patients, constantes radiologiques...).

C - Importance de la reproductibilité

L'exactitude d'un instrument est limitée si sa reproductibilité n'est pas suffisante (fig. 1). On doit donc évaluer d'abord la reproductibilité avant de se lancer dans une étude de validation complète car une mesure non reproductible est inutile.

Le manque de reproductibilité peut avoir de sérieuses conséquences sur tous les types d'études:

- dans les essais thérapeutiques, cela diminue la puissance de l'étude en augmentant la variance de la variable d'intérêt;
- dans les études épidémiologiques d'étiologie ou de causalité, on recherche le rôle d'un facteur ou d'une exposition dans la survenue d'une maladie. Si la mesure du facteur de risque ou celle de la maladie étudiée n'est pas reproductible, leur association est sous-estimée. Cela peut masquer une relation réelle et significative entre facteur de risque et maladie (augmentant ainsi le risque de ne pas détecter une relation réelle). Lorsqu'on veut mesurer les facteurs de confusion, cela réduit la capacité de contrôler ces facteurs, biaisant ainsi les conclusions de façon non prédictible.

D - Importance de l'échantillonnage

Certaines règles sont à respecter concernant le choix des sujets et la technique d'échantillonnage.

La reproductibilité et l'exactitude d'un instrument peuvent varier selon le type de sujets sur lesquels elles sont déterminées. Il faut donc les évaluer sur des sujets similaires à ceux à qui on veut appliquer l'instrument. Les mesures de densité osseuse sont moins reproductibles chez les sujets ostéoporotiques que chez les sujets sains, car les vertèbres sont moins nettement visibles, et la mesure de leur contour est entachée d'une plus grande erreur. Si on veut utiliser la densitométrie osseuse pour distinguer les sujets sains des sujets ostéoporotiques, il faut évaluer la reproductibilité de cet examen dans la population générale. En revanche, si on veut distinguer, parmi des sujets ostéoporotiques, différents stades de gravité, il faut évaluer la reproductibilité et l'exactitude de l'appareil dans un échantillon de sujets ostéoporotiques et non dans la population générale.

E - Méthodes statistiques pour l'évaluation de la reproductibilité des instruments de mesure

Pour étudier la reproductibilité d'un instrument, on utilise la concordance entre des valeurs obtenues par le même instrument dans différentes conditions. Pour étudier la variation entre différents radiologues étudiant les mêmes radiographies, on compare les résultats obtenus par chaque radiologue et on regarde dans quelle mesure ces résultats sont concordants en utilisant des statistiques de concordance.

Il faut distinguer la concordance entre deux variables d'autres relations plus vagues et moins strictes telles qu'une simple association ou corrélation. Cette distinction importante entre concordance et association est la base des principes généraux utilisés pour l'analyse statistique des mesures.

1 - Principes généraux

Pour décrire la relation qui existe entre deux variables dans une population donnée, les index utilisés habituellement représentent la mesure dans laquelle les variations d'une des variables sont similaires à celles de l'autre variable. Ces deux variables peuvent ou non être exprimées dans des échelles de mesure similaires.

Exemples:

- Si on s'intéresse à la relation entre taux de cholestérol et âge (le cholestérol étant une variable continue et l'âge étant exprimé en catégories de 10 ans), un indice d'association décrit dans quelle mesure le taux de cholestérol monte (ou descend) d'une décennie à l'autre. Ces indices d'association sont le coefficient de corrélation linéaire (r), le coefficient de régression, le coefficient rho de Spearman (ρ), le coefficient tau de Kendall (τ).
- On veut étudier la relation entre le taux d'hémoglobine et le taux de créatinine chez les patients atteints d'insuffisance rénale. Le coefficient de régression indique avec quelle force la valeur de la créatinine sérique influence celle du taux d'hémoglobine (le coefficient de régression correspondant à la pente de la droite de régression), et le coefficient de corrélation r détermine la relation linéaire qui existe entre ces deux variables, c'est-à-dire à quel point ces deux variables varient ensemble et de la même façon.

Concernant les instruments de mesure, lorsque deux instruments sont sensés mesurer le même phénomène, il ne suffit pas d'apporter la preuve que leurs mesures sont simplement corrélées

au-delà du simple hasard. Ces mesures devraient en théorie être identiques s'il n'y avait pas d'erreur de mesure. Ainsi la relation qu'on veut mettre en évidence doit être très forte, il ne s'agit pas d'une corrélation mais d'une concordance. Les index d'association classiques sont inadaptés pour décrire la concordance parce que deux variables peuvent avoir une relation très proche sans jamais être en parfait accord. Deux instruments A et B mesurant la densité osseuse peuvent varier dans le même sens et dans la même proportion même si A donne toujours des mesures plus élevées que B. Par ailleurs, deux variables peuvent avoir une très forte corrélation négative, ce qui correspond à une très mauvaise concordance malgré la très forte corrélation.

Plusieurs index de concordance ont été proposés pour évaluer la reproductibilité et l'exactitude des instruments. Seul le coefficient kappa est étudié ici.

Il est nécessaire, pour utiliser des indices de concordance, que les deux variables soient exprimées dans la même unité. Si on doit comparer deux variables qui s'expriment sur des échelles différentes, on ne peut utiliser des index de concordance, mais seulement des mesures d'association.

2 - Type des données

Les index de concordance utilisés dépendent du type des données.

a - Données qualitatives

On peut calculer simplement le pourcentage de concordance. Cependant, cette mesure englobe le pourcentage de concordance qui existerait par hasard, même si les deux variables étaient totalement différentes et indépendantes. Le coefficient kappa est plus approprié car il mesure la concordance entre les variables en tenant compte de l'effet du hasard.

Ne pas prendre en compte l'effet du hasard peut conduire à de fausses conclusions en surestimant la concordance réelle. Pour une variable binaire, la concordance prévisible par pur hasard est donnée par la formule:

$$[p_1 \times p_2] + [(1-p_1) (1-p_2)]$$

où p_1 est la proportion de personnes ayant une caractéristique selon une mesure X, et p_2 la proportion de personnes ayant la même caractéristique selon une mesure Y.

Le coefficient kappa est défini ainsi:

$$\text{Kappa} = \frac{\text{concordance observée} - \text{concordance prévisible}}{1 - \text{concordance prévisible}}$$

Lorsque deux mesures sont concordantes à un degré qui n'est pas supérieur au pur hasard, la valeur de kappa est zéro. Lorsque les deux mesures sont parfaitement concordantes, kappa est égal à 1.

Exemple: une étude compare l'information sur l'observance d'un traitement à la réserpine obtenue par l'interrogatoire des patients d'une part et par les dossiers médicaux d'autre part (tableau 4).

$$\text{Pourcentage de concordance due au hasard} = \frac{(21 \times 39) + (196 \times 178)}{(217)^2} = 0,7583$$

$$\text{Pourcentage de concordance observée} = \frac{14 + 171}{217} = 0,8525$$

$$\text{Kappa} = \frac{0,8525 - 0,7583}{1 - 0,7583} = 0,39$$

Dans cet exemple, ne pas prendre en compte le hasard aurait conduit à surestimer nettement la concordance (85 % au lieu de 39 %).

b - Données ordinales

Dans ce cas également, kappa est la statistique de choix mais on peut le modifier en kappa pondéré. Le principe est de prendre en compte l'importance de la discordance en considérant tous les différents degrés de concordance partielle.

Par exemple, s'il y a quatre catégories de réponse (tableau 3): une discordance entre un stade I selon l'observateur A et un stade II selon l'observateur B est moins grave que si l'observateur A estime le stade à I et l'observateur B le stade à IV.

c - Données continues

Les index à utiliser sont les coefficients de corrélation intra-classes qui associent une mesure de corrélation à une statistique testant la différence entre les moyennes. En termes de régression linéaire, cela signifie que ces coefficients testent non seulement la similarité des pentes des droites de régression, mais également que la droite passe par 0.

L'utilisation d'une analyse de variance à plus de deux entrées permet d'étudier les variations provenant de plusieurs sources simultanément et correspond à des méthodes qui sont au delà des objectifs de ce chapitre.

III – QUESTIONNAIRE : SAVOIR EXPRIMER CE QUE L'ON SOUHAITE

Questionnaire: de quoi s'agit-il ? Du document imprimé ? De l'ensemble constitué des questions posées à un individu ? Ou du concept global "entretien-questions-réponses-enregistrement des réponses" qui conduit *in fine* à l'élaboration du document imprimé ?

Dans un essai clinique, la forme papier ou électronique du questionnaire, souvent nommée "bordereau" ou "formulaire" est destinée à stocker les informations récoltées lors de l'étude. Cette version est d'autant plus facilement élaborée qu'elle suit la séquence chronologique de l'essai (par exemple, un questionnaire par visite médicale, dit bordereau de visite) et qu'elle est destinée à consigner des données ne prêtant pas sujet à interprétation (par exemple, les valeurs diastolique et systolique de la pression artérielle).

Il en va tout autrement d'un questionnaire destiné à l'évaluation de données subjectives concernant le vécu des patients, leur douleur, leur qualité de vie. Comment mesurer objectivement le subjectif ? Mesure-t-on bien ce qu'on souhaite mesurer ? Le vocabulaire

utilisé est-il compréhensible par tous et de la même manière (en Afrique centrale, fièvre signifie à la fois paludisme et fièvre) ?

Le questionnaire, qui constitue un exercice de communication pour les deux parties, ne peut être élaboré que lorsqu'une des parties sait ce qu'elle veut communiquer et obtenir de la part de l'autre. Il n'est pas envisageable de le construire avant que la question de recherche à laquelle on souhaite répondre, ainsi que le plan de l'étude ne soient connus avec précision

A - Définition et validation du champ d'exploration

Bien avant la rédaction des questions, il est primordial :

- de définir avec précision ce qui sera mesuré. Un questionnaire est d'abord destiné à la mesure de l'élément d'intérêt (par exemple : douleur, satisfaction...);
- et de s'assurer qu'un terme employé recouvre le même concept et la même définition pour tous.

La définition correcte des termes représente un tel obstacle que l'on fait souvent appel à un groupe multidisciplinaire d'experts lors de la construction d'un nouveau questionnaire.

L'étude de la littérature est d'un précieux secours car de nombreux questionnaires ont déjà été validés. Ils peuvent, sous certaines conditions, être réutilisés en l'état.

B - Les divers types de questions

On distingue trois grandes catégories de questions, non exclusives. Le choix dépend de la nature de l'information recherchée et de la façon dont l'analyse ultérieure est envisagée.

1 - La question ouverte

Exemple : "Comment décririez-vous votre douleur ?"

C'est la moins contraignante pour le sujet questionné. Il lui est possible d'employer ses propres mots pour exprimer toute la panoplie de ses sentiments. Ceci est particulièrement important lorsque l'expérimentateur cherche avant tout à définir l'étendue de ce qui est ressenti par le patient, la variété des opinions de l'équipe soignante ...

La synthèse, à des fins de codage et d'analyse statistique, d'un ensemble aussi divers de réponses à une même question est compliquée. Elle peut conduire à des erreurs d'interprétation. La grande latitude laissée au sujet dans sa réponse expose l'investigateur à sa bonne volonté et/ou sa capacité à exprimer et transmettre ce qu'il ressent. Le codage, qui implique une réduction (par un processus de standardisation) de l'information, conduit à un appauvrissement de l'information initialement recueillie.

2 - La question semi-ouverte

Exemple : "Lors de votre dernière consultation médicale, votre médecin a-t-il également examiné un (ou plusieurs) membre(s) de votre famille: père, mère, frère(s), sœur(s), enfant(s), autre(s) ?"

Avec ce type de question, l'effort de mémoire est minimum. La réponse est suggérée et orientée.

3 - La question fermée

Exemple : "Etes-vous né(e) un vendredi 13 ?"

Elle impose une réponse univoque parmi celles proposées par l'équipe qui a rédigé le questionnaire (oui, non, ne sait pas). Aucun effort (ou presque) de mémorisation n'est demandé. Le codage de la réponse est facile et (au risque d'erreur près) fiable. En revanche, aucun choix n'est possible en dehors des réponses proposées. Ce n'est pas dommageable pour l'exemple donné ci-dessus. Cela l'est certainement pour l'évaluation des émotions, du comportement, de l'apport calorique, de la douleur...

C - Un vocabulaire approprié

Y a-t-il des mots à éviter ou d'autres à employer plus particulièrement ? Chaque mot simple et familier, faisant référence à un concept clair peut être utilisé. Dans le cas contraire, il doit être évité.

Comment le savoir ? Une équipe expérimentée choisit le vocabulaire en fonction du public visé et évite, lors de la construction d'un questionnaire destiné au "patient moyen", l'emploi de termes culturellement marqués, nouveaux, de termes scientifiques ou médicaux précis mais trop spécialisés, de termes insuffisamment répandus dans la population, ou à l'inverse trop flous ou recouvrant un sens large et imprécis, ou diversement interprétables, ... Un individu peut s'avérer incapable de décrire son "statut marital" alors qu'il est assurément capable de dire s'il est célibataire, concubin, marié, divorcé, ou veuf.

Un vocabulaire simple, clair et précis dont le sens est identique pour tout sujet potentiel quel que soit son statut socio-économique, sa catégorie socioprofessionnelle ou son niveau d'éducation est un gage de qualité du questionnaire.

D - Formulation de la question

Il faut tendre vers la clarté et l'absence d'ambiguïté. Il s'agit avant tout d'éviter les formulations trop générales, complexes, ambiguës, litigieuses. Quelques exemples sont donnés dans le tableau 5.

E - Organisation des questions

Le but d'un questionnaire est de recueillir des données de qualité, communiquées librement par le sujet qui participe à l'étude. Le souci est d'éviter que le questionnaire soit mal ou incomplètement rempli.

Il faut intéresser le patient, lui donner envie de répondre et donc se concentrer sur la conception du fond autant que de la forme du questionnaire:

- au début figurent les questions auxquelles on tient absolument à avoir une réponse et auxquelles le patient a envie de répondre. C'est l'accroche initiale, attractive, garantie de réponse;

- puis viennent les questions par thème, groupées, car cela facilite grandement le déroulement de l'entretien (avez-vous déjà été opéré de la prostate ? Si oui, répondez aux 6 questions qui suivent, sinon allez directement à la question 22);
- les questions les plus personnelles prennent en général place à la fin, lorsque le sujet est confiant et cherche moins qu'au début à organiser ou à contrôler l'information qu'il délivre.

F - Codage

Convertir les réponses en chiffres, plus facilement utilisables à des fins d'analyse, implique l'emploi d'un système de conversion.

Deux possibilités s'offrent au chercheur :

- l'élaboration d'un système personnel de codage, indispensable si le champ de l'étude est nouveau ou très spécialisé,
- ou l'application de systèmes pré-établis et déjà testés : connaître la prévalence des maladies et leurs complications dans un hôpital peut se faire en appliquant une classification nationale ou internationale, par exemple la Classification Internationale des Maladies de l'Organisation Mondiale de la Santé, dont les codes sont accessibles à tous.

Le codage est fonction du type de question.

Une question fermée, imposant une réponse parmi celles proposées, permet un codage facile d'emblée.

En revanche une question ouverte ne permet pas de prévoir *a priori* l'ensemble des réponses. Les réponses sont alors regroupées par catégories logiques, de façon subjective, à l'issue de l'étude, par l'investigateur.

Où faire figurer les instructions de codage ? Deux options s'offrent à l'investigateur:

- sur le questionnaire lui-même

Au niveau de chaque question, figurent les valeurs de codage pour chaque variable. Il est utile pour l'épidémiologiste, dans la gestion des variables, ainsi que plus tard pour le statisticien au cours de l'analyse, de situer la place relative occupée par la valeur codée de la variable dans l'ensemble du questionnaire (tableau 6). Le codage figure en permanence sous les yeux, les risques d'erreur de transcription sont moindres. Mais l'espace nécessaire impose une augmentation de la longueur du questionnaire, particulièrement lorsque les instructions de codage sont longues et complexes. Le volume, le poids ainsi que le coût généré sont généralement plus élevés.

- dans un document annexe (tableau 7)

Le questionnaire est allégé dans sa présentation: ne figure que la question. Mais l'information de codage est moins facilement et moins rapidement disponible, un document annexe s'égaré, ...

Le codage est aussi fonction de la question de recherche et de l'analyse statistique prévue. Il est de tradition de réserver certains codes à des valeurs particulières: "non" prend souvent la valeur 0, "oui" la valeur 1, les valeurs manquantes les valeurs 9, 99, ...

La question de recherche peut cependant conditionner les valeurs des variables d'une façon particulière. Le statisticien peut, lors de l'analyse, recoder des variables (par transformation) si le codage initial s'avère inadéquat.

Consulter un épidémiologiste ou un biostatisticien à ce stade est certainement sage.

G - Utilisation d'un questionnaire étranger

Le chercheur peut être tenté d'utiliser un questionnaire développé à l'étranger pour un problème de même ordre. Outre la difficulté d'une traduction parfaite se pose la question du transfert et de l'applicabilité des concepts culturels qui ont présidé à l'élaboration du questionnaire.

La demande exprimée par un patient devenu lombalgique à la suite d'un accident du travail n'est pas identique selon qu'il est heureux ou non dans l'accomplissement de son travail quotidien. Le chercheur doit en tenir compte dans l'analyse des données qu'il a recueillies sur ce thème. Mais a-t-il initialement pris soin de vérifier qu'une même demande était exprimée de façon identique par les sujets de la population de l'étude ? Des patients de cultures différentes n'expriment pas des difficultés similaires par les mêmes symptômes. Conduire une recherche dans ce domaine impose de tenir compte de ces particularités culturelles dès la phase d'élaboration du questionnaire.

De même, un questionnaire destiné à l'évaluation de divers aspects de l'état de santé de patients britanniques, bien qu'ayant été validé dans d'autres pays d'Europe du Nord et ayant conduit à des résultats similaires, ne traduirait probablement pas le même concept s'il était appliqué tel quel dans les pays latins de l'Europe du Sud: on ne se fait pas culturellement la même idée de la santé dans chacun de nos pays européens. Il en est de même entre Américains et Français.

Pourquoi utilise-t-on alors parfois des questionnaires étrangers ?

Parce que cela est inévitable dans certaines circonstances, par exemple lorsqu'une étude multicentrique se déroule simultanément dans plusieurs pays dont la langue est différente, ou lorsqu'il s'avère nécessaire de fonder une nouvelle étude sur la méthodologie et le questionnaire d'une étude précédemment publiée, considérée comme la référence en la matière, afin de la répliquer et d'en comparer les résultats. Quelques étapes sont nécessaires à la bonne transposition d'un questionnaire d'un pays (d'une culture) à un autre (tableau 8).

IV - EXACTITUDE ET REPRODUCTIBILITE

Deux qualités sont indispensables à un questionnaire : son exactitude et sa reproductibilité.

L'exactitude est la capacité du questionnaire à fournir une mesure exacte de ce qui est à mesurer, comme la capacité pour le champion de tir à l'arc à atteindre le centre de la cible, ou celle d'un thermomètre à indiquer l'exakte température.

Par analogie avec un test diagnostique, le questionnaire parfaitement exact serait celui qui aurait une sensibilité et une spécificité parfaites (100 %) et pour lequel les valeurs prédictives

positive et négative seraient également de 100 %. Un tel test a-t-il jamais existé ? Le questionnaire administré aux candidats de l'examen du code de la route est non seulement supposé dépister les candidats qui connaissent bien le code de la route (sensibilité du test) et éliminer ceux qui le connaissent insuffisamment (spécificité du test), mais également prédire qu'un candidat qui a réussi le test connaît bien son code de la route (valeur prédictive positive) ou qu'un autre candidat a échoué parce qu'il le connaissait mal (valeur prédictive négative).

Si on ne sait pas quelle mesure choisir, parce que plusieurs peuvent s'appliquer à ce concept, il faut en utiliser plusieurs. Si on ne sait pas quelle dimension du concept mesurer, il faut les mesurer toutes.

La reproductibilité est la capacité du questionnaire à fournir une mesure identique de façon répétée, comme pour le tireur à l'arc à mettre toutes ses flèches au même endroit, ou pour le thermomètre à indiquer la même température de façon répétée. La condition de reproductibilité est vérifiée dès que le même test, appliqué plusieurs fois de suite dans les mêmes conditions, conduit à chaque fois au même résultat. Il s'agit donc d'une qualité bien différente de la précédente mais tout aussi importante.

La figure 1 illustre ces deux concepts.

Comment s'assurer de la qualité d'un questionnaire avant son utilisation dans l'étude ?

- En le testant pendant une période de rodage (pré-test) et en l'appliquant à quelque temps de distance (re-test). La concordance entre les réponses doit être élevée, si l'information recherchée n'a eu aucune raison de changer. Ceci teste la reproductibilité.
- Puis en comparant les résultats obtenus par son propre questionnaire à ceux d'un questionnaire de référence précédemment publié par une autre équipe dans des circonstances similaires sur une population similaire. Ceci teste l'exactitude.
- Ou en utilisant une méthode de mesure déjà validée par une autre équipe: il en existe parfois des dizaines pour une même situation à évaluer. Il faut malgré tout re-tester la méthode de mesure dans ce nouveau contexte puisqu'elle a été validée dans des circonstances différentes.

V - PRESENTATION DU QUESTIONNAIRE

Tout questionnaire doit être très clairement identifié.

Le nom de l'étude (par exemple, ETCI2 - ETude de Comparaison de 2 Interventions) et le titre du questionnaire (par exemple, "Evaluation de la qualité de la vie à 6 mois") doivent figurer en gros sur la couverture ainsi que le numéro de code du centre (dans une étude multicentrique), la date à laquelle il a été complété, et l'identification du patient.

Si plusieurs questionnaires différents sont utilisés chez un même patient, chaque questionnaire doit pouvoir être distingué, par exemple grâce à une couleur spécifique, une identification spécifique facilement reconnaissable (par exemple, Q|U|A|V|0|6 et Q|U|A|V|2|4 représentent respectivement les questionnaires d'évaluation de la qualité de la vie à 6 et à 24 mois).

Les questionnaires du même type, utilisés chez des patients différents, sont identifiés par un code unique qui respecte l'anonymat tout en évitant les confusions entre questionnaires. On

utilise souvent le format suivant: |K|E|R|C|E|, où les trois premières lettres représentent les trois premières du patronyme et les deux suivantes les deux premières du prénom.

Les questions doivent être numérotées dans un ordre séquentiel, sans omission. Il en est de même pour les pages.

Tout symbole qui peut favoriser la rapidité de compréhension du questionnaire, sans en altérer le sens, est le bienvenu. Ceci est particulièrement vrai d'une flèche qui évite l'emploi d'une suite de mots, d'un point (•) ou d'un astérisque (*) placé devant chaque idée afin de la distinguer de la précédente. Il en est de même, pour éviter de se tromper de ligne, des pointillés qui relient la fin de la question à la case dans laquelle figurera la réponse.

Né(e) un Vendredi 13 ? (non = 0; oui = 1; ne sait pas = 9) |_|

La lisibilité et la compréhension sont toujours facilitées par un espacement suffisant entre les questions.

VI - LE PRE-TEST

Le pré-test représente la phase ultime de la préparation du questionnaire. Il peut être difficilement vécu car la remise en question est parfois douloureuse ! Il représente cependant une étape indispensable.

Il est conduit dans des conditions identiques à celles de l'étude réelle et doit porter sur au moins 20 à 30 individus. Combien de pré-tests ? Il est rare que le questionnaire soit d'emblée parfaitement rectifié : un second pré-test est probablement à prévoir dès la planification de l'étude.

A l'occasion de ce pré-test sont obtenues pour la première fois:

- des données quantitatives, comme le temps nécessaire pour compléter le questionnaire (durée moyenne, durées minimale et maximale), le nombre de questions qui n'obtiennent pas de réponse, le nombre de réponses incohérentes (en désaccord avec une réponse à un précédent item), l'existence d'éventuels questionnaires non complétés, ...
- des données qualitatives, peut-être les plus importantes, comme l'impression générale de facilité, ou non, à compléter le questionnaire, ou les difficultés liées au niveau culturel des questionnés (problème plus souvent lié au vocabulaire employé qu'au concept abordé par la question), ou relatives à la séquence logique des questions, ...

Les commentaires des questionnés sont très riches d'informations. Ne jamais les négliger et tenter d'en obtenir le plus possible et de la meilleure qualité sont deux règles primordiales. Ces commentaires peuvent aboutir à la résurgence de concepts initialement abandonnés ou ignorés, ou à l'émergence de nouveaux concepts. Ils peuvent aussi concerner la lisibilité du questionnaire, la répétition de questions, l'insuffisance de place laissée pour la réponse aux questions ouvertes...

Conclusion

Les instruments de mesure jouent un rôle important, parfois sous-estimé, dans la qualité d'une étude. Pour déterminer avec précision les relations entre facteurs étudiés et critères de jugement, ces facteurs doivent être mesurés de façon reproductible et exacte. La validation des instruments de mesure est donc capitale et correspond à une démarche délicate et souvent très longue.

Le clinicien ou l'équipe de recherche qui utilisent un questionnaire poursuivent deux buts:

- obtenir l'information nécessaire à leur étude ou à leur enquête,
- mesurer cette information avec le maximum de qualité (en termes d'exactitude et de reproductibilité),

tout en respectant la dignité des patients. La règle d'or est certainement d'établir une relation de confiance dès les premières questions.

Tableau 1 - Les deux grands types de caractéristiques liées à un instrument de mesure

Biais	erreur systématique	exactitude (ou validité)
-------	---------------------	--------------------------

Chance	erreur due au hasard	reproductibilité (ou fiabilité)
--------	----------------------	---------------------------------

Tableau 2 - Concordance des avis de deux radiologues pour 110 sujets suspects de pneumoconiose

		Radiologue n° 2		
		Présent	Absent	
Radiologue n° 1	Présent	14	7	21
	Absent	8	81	89
		22	88	110

Mis en forme : Police :Gras

Mis en forme : Police :Gras

Tableau 3 - Concordance des avis de deux cliniciens sur la sévérité de l'atteinte de 100 femmes présentant un cancer du sein

		Clinicien n° B					
		I	II	III	IV		
Clinicien n° A	I	25	7	2	1	35	Mis en forme : Police :Gras
	II	4	14	5	0	23	Mis en forme : Police :Gras
	III	3	6	17	3	29	Mis en forme : Police :Gras
	IV	0	1	2	10	13	Mis en forme : Police :Gras
		32	28	26	14	100	Mis en forme : Police :Gras

Tableau 4 - Concordance de l'information sur l'observance d'un traitement à la réserpine obtenue par l'interrogatoire des patients d'une part et par la lecture des dossiers médicaux d'autre part, chez 217 patients

		Dossier médical		
		Oui	Non	
Interrogatoire	Oui	14	7	21
	Non	25	171	196
		39	178	217

Mis en forme : Police :Gras

Mis en forme : Police :Gras

Tableau 5 - Formulations de questions à éviter

-
- la double négation : pensez-vous que les gens qui n'ont pas d'emploi ne sont pas heureux ?
 - la double interrogation (deux questions à la fois) : êtes-vous favorable à l'adjonction systématique, moyennant 2 euros par mois, de fluor dans l'eau de boisson pour éviter les caries chez vos enfants ?
 - la suggestion appuyée malgré la forme interrogative : ne trouvez-vous pas que la fluoruration systématique de l'eau de boisson est une atteinte à la liberté individuelle ?
 - la combinaison des deux précédentes (suggestion appuyée et deux items explorés par une même question) : préférez-vous une source d'énergie propre comme le nucléaire à une centrale à charbon polluante ?
 - le(s) mot(s) au sens imprécis : êtes-vous souvent constipé(e) ?
 - la question imprécise : seriez-vous favorable à une réduction modérée de l'apport calorique chez les personnes inactives ?
-

Tableau 6 - Exemple de codage inclus dans le questionnaire

N° Question	Libellé de la Question	Codage	Repérage
06	En quelle année êtes-vous né(e) ? (Indiquer l'année complète)	_ _ _ _	(58-61)
07	Etes-vous né(e) un Vendredi 13 ? (non = 0, oui = 1, ne sait pas = 9)	_	(62-62)
08	Etes-vous né(e) en France métropolitaine ? (non = 0, oui = 1, ne sait pas = 9)	_	(63-63)
09	Si oui, dans quel département ? (Indiquer le numéro du département)	_ _	(64-65)

Tableau 7 - Informations minimales lorsque les règles de codage figurent dans un document annexe au questionnaire

-
- numéro de la variable
 - son nom complet (par exemple, jour de naissance)
 - son abréviation: le plus souvent huit symboles au maximum, pour des raisons de compatibilité des fichiers informatiques entre eux (par exemple, JOURNAIS)
 - un bref descriptif (par exemple, jour de la semaine lors de la naissance)
 - son type: caractère, numérique, logique, date, ... (par exemple, numérique)
 - sa largeur une fois codée, incluant éventuellement la virgule et les décimales (par exemple, 4 caractères - 2 chiffres, virgule, 1 décimale - pour la variable "température corporelle" - 37,5°C)
 - la page (et éventuellement la référence) du questionnaire à laquelle elle se réfère
 - la base de données (fichier informatique) à laquelle elle appartient, au cas où il y aurait plusieurs fichiers
 - enfin et surtout, les modalités de codage (par exemple, non = 0, oui = 1, ne sait pas = 9)
-

Tableau 8 - Etapes d'une validation "trans-culturelle"

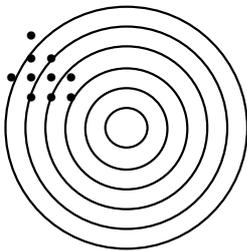
-
- sélection du questionnaire étranger le plus susceptible de répondre au besoin
 - double traduction parallèle et indépendante (de la langue d'origine à la langue actuelle), en évitant le piège de la traduction littérale, par deux personnes connaissant bien l'idée sous-jacente à chaque question (en général, par au moins un des chercheurs bilingues de l'équipe)
 - double traduction inverse (de la langue actuelle à la langue d'origine), par d'autres traducteurs que les premiers, afin de vérifier que les concepts véhiculés par chaque question sont encore présents
 - arbitrage des différences lors d'une réunion de consensus
 - pré-test du questionnaire auprès de futurs interrogés potentiels
 - modifications finales
-

Fig. 1 - Exactitude et reproductibilité: la classique illustration de la cible

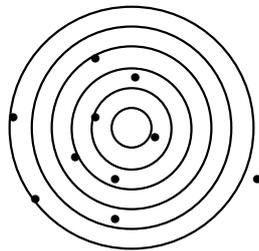
Sur la cible du centre, le tir est soumis à une erreur aléatoire et tous les impacts se distribuent au hasard autour du centre: le tir est très peu reproductible, et donc peu exact.

Sur la cible de gauche, les impacts sont regroupés mais systématiquement à côté du but recherché: le tir est très reproductible mais très inexact.

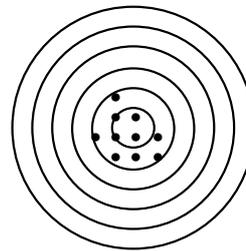
Sur la cible de droite, les impacts sont regroupés autour du centre: le tir est exact et reproductible.



Reproductible,
mais non-exact



Non reproductible,
et donc peu exact



Reproductible
et exact