

FOVIEA

STAT'

LES
ÉCHELLES
D'ÉVALUATION

TRAITEMENT STATISTIQUE
DES ÉTUDES CLINIQUES

Les échelles sont des instruments d'évaluation très fréquemment employés dans les essais thérapeutiques.

HAMILTON, l'auteur d'une des échelles les plus utilisées actuellement dans l'étude de la dépression, remarquait que le "travail que constitue la construction d'une échelle satisfaisante est considérable et ne peut être entrepris par des amateurs".

Il est notamment très important que des travaux de validation sérieux soient réalisés lorsqu'une nouvelle échelle est créée.

Spécialisée dans le traitement statistique des études cliniques, FOVEA STAT' vous propose au travers de ce fascicule quelques notions simples qui vous permettront de mieux appréhender cet outil.

Définition



Une échelle est **un ensemble d'items**, chaque item étant une partie du phénomène à évaluer et comportant **plusieurs modalités**.

Certaines échelles permettent de calculer un **score global** ou **partiel** : c'est le score de plusieurs items.

Les différents types



- 1 - L'ÉCHELLE NOMINALE est constituée d'items dont les modalités ne sont **pas hiérarchisées**.
- 2 - L'ÉCHELLE ORDINALE est constituée d'items dont les modalités sont **ordonnées, hiérarchisées**.
- 3 - L'ÉCHELLE D'INTERVALLES est une échelle ordinaire pour laquelle **les intervalles** entre les modalités sont **égaux**.
- 4 - L'ÉCHELLE CARDINALE permet l'obtention d'un **score synthétique** de plusieurs (*score partiel*) ou de la totalité (*score global*) des items.

Classification



- 1 - Une échelle est dite **unidimensionnelle** quand **1 seule dimension** est évaluée à **plus de 80 %**. Dans le cas d'une échelle comportant 20 items, il est ainsi nécessaire que 16 d'entre eux mesurent la dimension psychologique pour que cette échelle soit classée en "Psychologique".
- 2 - Une échelle est dite **multidimensionnelle** lorsqu'elle comporte **N dimensions** et que ces N dimensions sont évaluées dans **des proportions semblables**.



Validité



La validité d'une échelle est **double** :

Validité interne (*rigueur intrinsèque de l'échelle*),
Validité externe (*permanence des résultats pour des phénomènes comparables*).

LA VALIDITÉ INTERNE

La validité interne d'une échelle est **sa faculté à mesurer correctement le phénomène que l'on veut évaluer**. Une échelle est dite valide si elle ne déforme pas la réalité et si son utilisation n'induit pas d'erreur systématique dans les résultats.

Il existe plusieurs types de validité interne :

La validité par référence à un critère (Criterion validity) comprend :

La validité de simultanéité qui apprécie la **concordance** entre 2 épreuves exécutées pratiquement simultanément.

L'élément de comparaison (*critère reconnu et valide*) peut être **une autre épreuve de même nature** (*validité concurrente*) ou **un autre moyen d'évaluation de valeur reconnue**.

La validité prédictive qui exprime l'aptitude de l'échelle à **prévoir une évolution** ou **un état ultérieur à partir d'un résultat actuel**.

La validité de la teneur (Content validity) apprécie la **concordance** entre le **contenu de l'échelle** (*ses différents items*) et **l'ensemble des situations pouvant être rencontrées lors de l'évaluation du phénomène**.

La validité de la conception (Construct validity) évalue la **logique de construction de l'échelle**. Par exemple, le contrôle de qualité des échelles ordinales doit pouvoir déceler s'il existe une bonne hiérarchisation entre les modalités des items.

L'analyse factorielle des correspondances permet d'apprécier la **cohérence interne** de l'échelle, c'est-à-dire son **homogénéité**.

LA VALIDITÉ EXTERNE

La validité externe renvoie à **la reproductibilité des résultats**.

La reproductibilité ou "**fidélité**" d'une échelle est **sa capacité à reproduire des résultats uniformes** : les mêmes conditions doivent engendrer les mêmes résultats.

Il existe différentes expressions de la fidélité :

Par référence à l'individu : la fidélité inter-juges apprécie la **corrélation** entre les résultats obtenus par **2 ou plusieurs investigateurs** évaluant **au même moment le même individu** avec **la même échelle**.

Elle est exprimée par le **coefficient de concordance Kappa**.

2 - **Par référence au temps : la fidélité test-retest** apprécie la **corrélation** entre 2 ou **plusieurs évaluations du même individu** effectuées dans **les mêmes conditions expérimentales à des temps différents**.

Elle est mesurée par le **coefficient r de Pearson** pour les **valeurs numériques** ou de **Spearman** pour les **valeurs ordinales**.

3 - **Par référence à l'échelle : 3 moyens**

- **La cohérence interne** mesure l'**homogénéité de l'échelle** et vérifie l'**étroitesse des corrélations** entre les items. Elle est évaluée par le **test Alpha de Cronbach**.

- **La méthode des formes parallèles** met en évidence la **corrélation** entre **des versions différentes** de la même échelle, similaires par le nombre d'items, l'ordre de leurs difficultés, mais différentes dans leur contenu.

L'indice d'équivalence est le **coefficient de corrélation** entre les résultats obtenus à partir des 2 formes.

- **La méthode "split-half"** mesure la **corrélation** entre les **2 moitiés de l'échelle** ou entre les **items pairs et impairs**. Elle s'exprime par un **coefficient d'homogénéité**.

Nous ont déjà fait confiance

BEIERSDORF S.A.,
B.M.S., CHAUVIN, CIBA-
GEIGY, CLINIQUE, CRINEX, DU PONT
PHARMA, GALDERMA, INNOTHERA,
KNOLL FRANCE, LEURQUIN MEDIOLANUM,
LIPHA, LTM, M.S.D., NEGMA, NUTRICIA,
OBERLIN, PIERRE FABRE, PROCTER & GAMBLE
PHARMACEUTICALS, RHONE POULENC,
ROCHE, SANOFI, SCHERING-PLOUGH,
SERVIER, S.K.F., SMITH & NEPHEW,
RICHARDS FRANCE, SOFAMOR,
SYNTEX, SYNTHELABO, TAKEDA,
3M SANTÉ, WYETH-FRANCE,
YAMANOUCHI PHARMA

FOVIA
STAT