

# La méthodologie de validation sociale: guide pratique d'utilisation dans une étude clinique.

Louise F. Tremblay et Pascale Brillon

*Centre de recherche Philippe Pinel  
Institut Philippe Pinel de Montréal*

## Résumé

La validation sociale est une stratégie de recherche qui permet de répondre à plusieurs questions cliniques. Elle comporte deux stratégies. Dans la comparaison sociale, on compare le comportement d'un groupe clinique et d'un groupe non clinique, avant ou après le traitement. L'évaluation subjective consiste à demander à des membres de la communauté d'évaluer le niveau de compétence de sujets d'un groupe clinique. L'objectif de l'article est de présenter, par ordre de priorité, les éléments requis pour utiliser cette méthodologie d'une façon optimale et réaliste. La sélection du groupe de référence, l'homogénéité du groupe clinique ainsi que le choix des instruments de mesure sont les principaux aspects à considérer.

La méthode de validation sociale a été proposée par Kazdin (1977) et Wolf (1978) comme un moyen d'évaluer si les procédures de traitement sont socialement acceptables et si les objectifs thérapeutiques sont cliniquement significatifs. C'est toutefois en tant que stratégie pour déterminer si l'impact thérapeutique est cliniquement significatif que cette procédure a été utilisée le plus fréquemment. La méthodologie de validation sociale s'inscrivait dans la lignée des propositions de Strupp et Hadley (1977) qui ont mis en évidence l'importance de considérer des perspectives multiples dans la mesure de l'impact thérapeutique. Ces auteurs ont proposé trois critères pour évaluer l'efficacité d'une intervention: le critère *théorique* (celui du thérapeute), le critère *subjectif-expérientiel* (celui du client) et le critère *social* (celui de la communauté).

---

Les auteurs tiennent à remercier M. Stéphane Sabourin pour ces commentaires pertinents d'une version préliminaire de l'article. Les demandes de tirés-à-part doivent être adressées à Louise F. Tremblay Ph.D., Centre de recherche Philippe Pinel, Institut Philippe Pinel de Montréal, 10,905 est Henri-Bourassa, Montréal, Québec, Canada, H3C 1H1.

La validation sociale est une méthode qui permet de mesurer le critère social. Ce critère est particulièrement pertinent à prendre en compte dans trois contextes. Tout d'abord, lorsque l'opinion de la communauté peut influencer de façon significative la réintégration de l'individu<sup>1</sup> dans l'environnement naturel, par exemple, dans le cas de déficience intellectuelle (ex. Castles et Glass, 1986; Storev et Horner, 1991) ou de troubles mentaux graves (ex. Bellack, Morrison, Mueser, Wade et Sayers, 1990; Donahoe, Carter, Bloem, Hirsch, Laasi et Wallace, 1990). Deuxièmement, si les personnes à traiter ne peuvent pas, ou difficilement, définir eux-mêmes les objectifs de traitement ou en évaluer l'impact, comme les enfants (ex. Flanagan, 1986; Horstman, 1988) ou les adolescents (ex. Hansen, Watson-Petzel et Smith-Christopher, 1989; Kazdin, Esveldt-Dawson, French et Unis, 1987; Minkin, Braukmann, Minkin, Timbers, Fixsen, Phillips et Wolf, 1976). Enfin, lorsque l'amélioration du comportement-cible implique directement le jugement des personnes de l'environnement comme, par exemple, pour les habiletés sociales (ex. Millbrook, Farrell et Curran, 1986; Monti et Fingeret, 1987).

La validation sociale d'un traitement repose sur deux stratégies. La *comparaison sociale* implique de contraster la performance d'un groupe clinique, avant ou après le traitement, avec le comportement d'un groupe de référence ne présentant pas les difficultés du groupe clinique. Et, l'*évaluation subjective* qui consiste à demander à des individus en contact quotidien avec les sujets du groupe clinique (Kazdin, 1977) ou des volontaires (Vézina et Ladouceur, 1984) d'évaluer la performance des sujets cliniques pré et post-traitement.

Utilisée de façon adéquate, la méthodologie de validation sociale permet de répondre à plusieurs questions cliniques. L'avantage principal de cette méthode est de fournir aux cliniciens des données comparatives concernant le comportement ou l'opinion de la population générale face aux difficultés que vivent leurs clients. En possédant des objectifs de traitement et des critères d'impact plus valides, les cliniciens seront mieux outillés pour traiter des personnes qui vivent des difficultés où le critère social doit être privilégié.

On constate toutefois que, malgré ses nombreux bénéfices, très peu d'auteurs ont utilisé cette stratégie de recherche (Tisdelle et St-Lawrence, 1986). Les premiers écrits concernant la validation sociale se sont surtout consacrés à en promouvoir les avantages comme mesure d'évaluation de l'impact thérapeutique. On constate, par ailleurs, que les critères requis pour utiliser cette méthode n'ont jamais été énoncés de façon suffisamment concrète et réaliste pour en rendre l'utilisation fréquente et valide dans les études cliniques où le critère social doit être privilégié (Fawcett, 1991). De plus, cette méthodologie a souvent été appliquée sans toute la rigueur requise (Storev et Horner, 1991).

Les exigences méthodologiques nécessaires à la réalisation de recherches cliniques sont déjà bien connues (Bergin et Lambert, 1978; Campbell et Stanley, 1963; Chassan, 1979; Cook et Campbell, 1979; Garfield, 1978). Un certain nom-

1. L'utilisation du masculin inclut également le féminin.

bre des critères à remplir dans la réalisation d'une étude de validation sociale se superposent à ces exigences. Toutefois, le recouvrement n'est pas complet.

Dans la réalité quotidienne, le chercheur fait souvent face à des contraintes qui l'obligent à faire des choix lorsqu'il s'agit d'appliquer une méthodologie de recherche. Il n'est, en effet, pas toujours possible de remplir de façon optimale toutes les conditions requises pour réaliser une étude clinique (Bellack, 1989 a). Il importe donc d'établir des priorités valides.

L'objectif de cet article est d'identifier, par ordre de priorité, les éléments qui doivent être satisfaits pour réaliser une étude de validation sociale, d'une façon optimale, mais réaliste. Cette analyse permettra au chercheur de décider où il doit d'abord diriger ses énergies. Les éléments-clés seront identifiés afin de présenter un mode d'emploi pratique visant l'utilisation adéquate et efficace de la validation sociale. Il se destine donc aux chercheurs intéressés à mettre en application cette méthodologie mais aussi aux cliniciens désireux d'évaluer de façon critique les études cliniques où on a eu recours à cette méthode.

Des études de validation sociale réalisées dans le domaine des habiletés sociales ont été sélectionnées afin de démontrer concrètement les recommandations proposées. La qualité méthodologique de ces études en fonction des critères suggérés est présentée au tableau 1.

## Recommandations

Deux éléments sont essentiels à considérer pour utiliser la méthodologie de validation sociale. Ce sont, par ordre de priorité, 1) la sélection des échantillons: a) la représentativité du groupe de référence; b) l'homogénéité du groupe clinique et 2) le choix des instruments.

### 1) La sélection des échantillons

#### a) La représentativité du groupe de référence.

Le groupe de référence désigne 1) le groupe de sujets non-cliniques (individus ne présentant pas la difficulté du groupe clinique) avec lequel le comportement des sujets du groupe clinique est contrasté dans une étude de comparaison sociale ou 2) le groupe d'individus de la communauté à qui on demande d'évaluer des sujets cliniques dans une étude d'évaluation subjective. La sélection du groupe de référence est l'élément crucial à considérer dans une étude utilisant la validation sociale. C'est en effet l'emploi d'un groupe de référence qui définit l'essence même de cette méthodologie.

**Tableau 1**

Présence de critères de qualité méthodologique de quelques études de validation sociale

Étude	Hasard	Jumelage	n > 20 par sexe	Diagnostic n > 20 par sexe	n > 20 par sous-groupe	Contrôle variables confon-dantes	Accords inter-observateurs	Mesures d'observation
1						H		✓
2								✓
3								✓
4				✓				✓
5		✓						✓
6								✓
7		✓		✓	S			✓
8		✓		✓	S			✓
9		✓		✓				✓
10				✓				✓
11		✓		✓	✓			✓
12				✓	pour 1 des s. groupes	H		
13	✓				s. groupes	H		✓

Not. V: le critère est présent; H: hospitalisations antérieures; S: symptomatologie (1990); 9: Mueser, Morrison et Wade (1990); 7: Donahoe, Carter, Bloem, Hirsch, Laasi et Wallace (1990); 11: Mueser, Morrison, Wixted et Mueser (1991); 12: Kazdin, Esveldt-Dawson, French, Unis (1987); 13: Hansen, St-Lawrence et Christoff (1985).

Plusieurs auteurs, dont Kazdin (1977), Van Houten (1979), Vézina et Ladouceur (1984) et Yeaton et Sechrest (1981) ont mis l'emphasis sur l'importance du choix du groupe de référence. Toutefois, aucun de ceux-ci n'a spécifié les conditions idéales de sélection. Il existe plusieurs façons de déterminer un échantillon (Bertrand, 1986). Trois méthodes sont utilisées plus fréquemment dans les études de validation sociale: 1) l'utilisation de volontaires (sélection non-systématique) 2) le jumelage des sujets cliniques avec ceux du groupe de référence et la vérification explicite du succès de ce pairage; et 3) la sélection au hasard.

La sélection d'un groupe de référence d'une façon non-systématique consiste à choisir des volontaires ou des personnes connues du chercheur, sans critères de sélection explicites. Si elle constitue une méthode de sélection peu coûteuse, aisément applicable et donc attirante, le choix de cette méthode résulte cependant en la constitution d'un groupe de référence très peu représentatif de la communauté de vie des sujets. Cette limite est très importante. En effet, la représentativité des membres de l'environnement de vie des sujets du groupe cliniques est une des conditions essentielles qui est recherchée chez le groupe de la communauté. Le but visé lorsqu'on utilise un groupe de référence est l'utilisation de données comparatives afin de définir des objectifs et des mesures de traitement qui sont plus proches de la réalité donc, plus valides écologiquement. Malgré les limites importantes de la méthode de sélection non-systématique, c'est celle qui est utilisée dans la majorité des études présentées au tableau 1<sup>2</sup>.

Le jumelage consiste à apparier, selon différentes variables, un groupe non-clinique à l'échantillon clinique et à évaluer, à posteriori, si les groupes diffèrent ou non sur ces dimensions. Le sexe, l'âge et le statut socio-économique sont les variables les plus souvent choisies pour constituer des groupes équivalents dans les études portant sur les habiletés sociales. Cette stratégie de sélection présente des avantages indéniables lorsqu'on réalise une étude comparative de l'efficacité de traitement, toutefois, elle comporte plusieurs lacunes lorsqu'il s'agit de réaliser une étude de validation sociale.

Tout d'abord, la validité de cette méthode repose sur la sélection de variables d'appariement pertinentes. Or, le choix de ces variables s'appuie rarement sur une analyse empirique préalable des corrélations existant entre les variables choisies et le construit à l'étude. Un choix inadéquat des variables d'appariement peut produire un groupe de référence non-représentatif de la communauté de vie des sujets cliniques. Cette situation risque de générer des résultats non valides puisque c'est la description du comportement du groupe de référence ou son opinion, selon le cas, qui serviront de normes pour identifier les objectifs de traitement et en évaluer l'impact. Par exemple, la relation entre le niveau socio-économique et le degré de compétence sociale est assumée par plusieurs auteurs ayant effectué des études dans le domaine des habiletés sociales. Cependant, le lien entre ces deux variables n'ayant jamais été démontré; l'appariement des groupes sur cette variable pourrait être non pertinent.

2. Il n'y a pas de colonne dans le tableau pour cette méthode de sélection, puisque son utilisation n'est pas recommandée.

Par ailleurs, il peut être irréaliste de tenter l'appariement d'un groupe clinique et non-clinique si les deux groupes présentent des caractéristiques très différentes. Par exemple, plusieurs tentatives de jumelage quant au niveau socio-économique de personnes souffrant de troubles mentaux graves et de la population générale, se sont soldées par un échec (voir par exemple, les études de Bellack *et al.* 1990; Bellack, Morrison, Wixted et Mueser, 1990; Mueser, Bellack, Douglas et Morison, 1991, au tableau 1). Il ne faut pas en conclure pour autant que ces études ont été mal menées mais plutôt qu'il est irréaliste d'apparier des personnes qui souffrent de troubles mentaux graves et des personnes de la population générale quant au niveau-socio-économique, les premières étant en général très défavorisées à cet égard. Ainsi, en l'absence de connaissances empiriques suffisantes pour jumeler les sujets du groupe clinique et non-cliniques, il est préférable de choisir les sujets du groupe de référence au hasard.

Dans une étude de validation sociale, on cherche à établir une norme. Il est donc plus important de constituer un groupe de sujets représentatifs de la population générale que de constituer un groupe en tout point identique à celui du groupe clinique. Il s'agit de déterminer des objectifs de traitement réalistes ou d'en mesurer l'atteinte après l'intervention et ce, en considérant comme critère, le comportement de membres de la communauté qui n'éprouvent pas la difficulté démontrée par le groupe clinique. La sélection au hasard du groupe de référence constitue la meilleure façon d'obtenir un échantillon représentatif du comportement de la population générale.

La sélection au hasard est particulièrement désignée pour obtenir la norme la plus juste possible. Cependant, lorsque l'on parle de sélection au hasard, on ne veut toutfois pas dire le hasard complet. En effet, il est nécessaire de limiter le recrutement des sujets à la région géographique de vie des sujets cliniques afin de connaître le comportement ou le jugement des personnes avec qui ces sujets pourraient interagir. Une liste des membres de la communauté, par exemple, par le bottin téléphonique<sup>3</sup>, permet de sélectionner au hasard des candidats potentiels à qui on demandera de participer à l'expérimentation.

Un autre avantage de la sélection au hasard consiste en la possibilité de documenter la représentativité de l'échantillon, ce qui n'est pas possible avec les deux autres procédés. Celle-ci nous permet, en effet, de savoir, entre autre, combien de personnes ont été rejointes, combien ont accepté ou non de participer et pourquoi. Elle permet donc de documenter les taux et les raisons du refus de participer ce qui nous permet d'identifier, s'il y a lieu, des biais dans l'échantillonnage et d'en tenir compte dans l'interprétation des résultats.

Une sélection au hasard comporte un avantage supplémentaire, celui de permettre de constituer un groupe de référence qui comporte des sujets présentant différents degrés d'habiletés par rapport à la variable mesurée. Ceci permet de comparer le comportement des sujets cliniques à ceux de sujets qui présentent des niveaux d'habiletés supérieurs, inférieurs ou égaux à la moyenne de la population générale. L'utilisation de cette méthode est préconisée, entre autres, par Ja-

3 Numéro spécial sur la mesure du changement thérapeutique dans la revue *Behavioral Assessment* Vol. 10 (1988)

cobson, Folette et Revenstorf (1984) et Kendall et Grove (1988). Il est important de noter que certains auteurs ne partagent pas cette position. Van Houten (1979, par exemple, favorise une sélection du groupe de référence qui serait composée de sujets non-clinique très compétents plutôt que de sujets représentatifs de la population générale choisis au hasard. Ce débat concernant le choix du critère auquel est comparé le groupe clinique dépasse largement le contexte de cet article. Le lecteur peut se référer au numéro spécial traitant ce sujet en détail, publié en 1988 dans la revue *Behavioral Assessment*<sup>4</sup>.

Malgré ses avantages inégalables afin de constituer un groupe de référence représentatif de la communauté de vie des sujets cliniques, la sélection au hasard du groupe de référence est très rarement utilisée. L'étude de Hansen, St-Lawrence et Christoff (1989) (voir tableau 1) est la seule où des sujets non cliniques, représentatifs de la population clinique ont été sélectionnés au hasard dans le milieu de vie des sujets (étudiants dans trois écoles secondaires publiques).

Enfinement, afin de maximiser l'impact d'un choix approprié du groupe de référence, celui-ci doit comporter un nombre suffisant de sujets. Un échantillon doit comporter vingt sujets et plus par sous-groupes afin de permettre l'utilisation adéquate de statistiques (Kraemer, 1981). L'examen du tableau 1 nous montre à nouveau que ce critère est rarement satisfait dans les études présentées au tableau 1 puisque seuls Millbrook, Farrell et Curran (1986) ont rempli ce critère dans leur étude.

Le nombre de sous-groupes requis dépend du nombre de variables à contrôler. Par exemple, dans une étude portant sur les habiletés sociales de personnes souffrant de troubles mentaux graves, il faut constituer des sous-groupes en fonction du sexe. La compétence sociale des hommes et des femmes souffrant de troubles mentaux graves s'est en effet avérée différente dans l'étude de Mueser, Bellack, Morrison et Wade, 1990.

Il est donc essentiel d'identifier les variables qui peuvent être des hypothèses rivales ou confondantes et de prévoir un nombre suffisant de sujets par sous-groupes. S'il n'est pas possible de recruter beaucoup de sujets (comme s'est souvent le cas dans les études cliniques), il est préférable de restreindre le nombre de sous-groupes et donc, la validité externe. Par exemple, le chercheur pourrait ne recruter que des hommes dans son échantillon. Cette décision permettra d'établir des conclusions solides parce qu'une ou plusieurs hypothèses rivales sont plausibles pour interpréter les résultats seront contrôlées.

Un dernier aspect mérite d'être signalé. Les notions de groupe de référence et de groupe de contrôle peuvent paraître interchangeables mais elles ne le sont pas. En effet, le groupe de contrôle est utilisé lorsqu'on veut comparer l'efficacité d'un traitement expérimental à un traitement factice ou au meilleur traitement connu. Dans les deux cas, il s'agit de contrôler les effets liés aux attentes des sujets. On veut s'assurer que l'amélioration des sujets est dû au traitement reçu et non pas

4 Le bottin téléphonique ne permet pas d'établir une liste qui remplit les critères du hasard au sens statistique (Bertrand, 1986) car uniquement 70% des adultes y sont inscrits. Toutefois, malgré ces limites, cette stratégie d'échantillonnage est souvent celle qui est la seule disponible, les autres types de listes plus complètes étant, en général, confidentielles.

uniquement à l'attention qu'on leur a accordée. Il s'agit donc de déterminer s'il y a eu ou non changement. Par ailleurs, on utilise un groupe de référence afin de déterminer si l'impact est cliniquement significatif en comparant le comportement des sujets cliniques à ceux du groupe non-clinique ou en se référant à l'évaluation des membres de la communauté.

b) L'homogénéité du groupe clinique

Les critères de sélection du groupe clinique sont identiques à ceux qui prévalent pour réaliser toute étude clinique. La standardisation de la procédure de recrutement des sujets est un des éléments des plus importants à satisfaire afin de garantir l'homogénéité de l'échantillon (Bergin et Lambert, 1978; Chassan, 1979; Gottman et Markman, 1978).

Un échantillon homogène de sujets ne peut être recruté sans le recours à une méthode diagnostique standardisée (Bellack, 1989a). Une procédure standard d'évaluation et l'utilisation d'un minimum de deux observateurs constitue un investissement essentiel afin de sélectionner un échantillon homogène. Il faut de plus évaluer l'accord inter-observateurs obtenu pour l'évaluation du diagnostic et rapporter le degré d'entente entre les juges selon un indice approprié (Garfield, 1978; Chassan, 1979). Le diagnostic n'est pas établi de façon standardisée dans un peu plus de la moitié des études répertoriées au tableau 1. Cet aspect essentiel pour réaliser une étude clinique de haute qualité est donc souvent négligé.

Les études doivent inclure, comme pour le groupe de référence, un nombre suffisant de sujets par sous-groupes de diagnostic ( $N > 20$ ) (Kraemer, 1981). Un nombre trop restreint de sujets par sous-groupes ne permet pas de réaliser des analyses statistiques. Le nombre de sujets par sous-groupes est insuffisant dans la majorité des études. Comme le montre le tableau 1, le nombre de sujets est adéquat uniquement dans les quelques études (no. 5, 8, 9, 10). L'emploi d'un nombre insuffisant de sujets par groupe a souvent comme conséquence l'obligation de regrouper les sous-groupes afin de réaliser les analyses statistiques. Cette décision équivaut pourtant à ne pas avoir évalué le diagnostic. Regrouper, par exemple, des personnes présentant des diagnostics aussi différents que celui de schizophrénie ou de maladie affective dans une même analyse (par exemple, Hansen, St-Lawrence et Christoff (1985); Fingeret, Monti et Paxson (1985); Monti et Fingeret (1987)) permet difficilement de tirer des conclusions valides parce que le groupe de sujets est trop hétérogène.

Ainsi, s'il est impossible de recruter un nombre suffisant de sujets par sous-groupes, il est vraiment préférable de restreindre le nombre de diagnostics étudiés ou de procéder à une étude impliquant plusieurs centres, ceci permettant d'obtenir un nombre approprié de sujets. D'autre part, l'inclusion de différents sous-types d'un même diagnostic dans un même groupe des schizophrènes différenciés et paranoïdes, par exemple, est un compromis réaliste et valide si cette distinction ne constitue pas en elle-même la problématique étudiée.

Le contrôle des variables confondantes est le dernier élément essentiel à prendre en considération dans la sélection du groupe clinique. Ceci est nécessaire afin de minimiser la possibilité que des variables puissent être invoquées comme étant des hypothèses rivales aux conclusions. Par exemple, dans une étude où les

sujets souffrent de troubles mentaux graves, le sexe, l'âge, la durée et le nombre d'hospitalisations précédentes, l'état de la symptomatologie, la présence de déficience intellectuelle, de syndrome organique et le statut des individus au moment de l'étude (hospitalisé ou non) sont toutes des variables qui doivent être contrôlées parce qu'elles pourraient en elles-même être considérées comme les causes des différences observées.

La description de l'échantillon quant aux variables qui ont été contrôlées doit de plus être précise. Ceci est d'autant plus nécessaire puisque, dans la pratique quotidienne, il est rarement possible de contrôler adéquatement toutes les variables potentiellement confondantes; un portrait précis de l'échantillon peut s'avérer une solution de compromis.

La création d'un ou de plusieurs groupes cliniques homogènes relève souvent du défi. Cependant, il est beaucoup plus facile de dégager des conclusions valides lorsque l'échantillon est homogène. Ainsi, restreindre le type de sujets étudiés à une catégorie diagnostic ou à un type de problème s'avère souvent le compromis le plus judicieux pour remplir à la fois les critères associés à la validité interne et externe. Augmenter le nombre de sujets de l'étude ne constitue pas une solution miracle. En effet, une réduction de la variabilité de l'erreur de mesure exige de sélectionner de très grands groupes de sujets. Ceci est souvent tout à fait irréaliste étant donné les coûts, les contraintes techniques et éthiques imposées lors de la réalisation d'une étude clinique et ce, même dans les études multi-centres.

2) Le choix des instruments

L'utilisation de mesures d'observation dans les études de validation sociale est très fréquente. L'emploi de données comportementales est devenu la norme, en particulier, dans certains domaines comme les habiletés sociales. Les recherches répertoriées au tableau 1 comportent, pour la plupart, des données d'observation. On constate cependant que l'observation systématique n'y est pas toujours utilisée de façon adéquate. Recueillir des données d'observation exige des efforts considérables. Ces énergies ne sont pas investies de façon optimale si les conditions requises pour garantir la fiabilité et la validité des données ne sont pas satisfaites.

La condition fondamentale à remplir lorsqu'on recueille des données d'observation est la mesure de l'accord inter-observateurs et son évaluation selon un indice approprié (Bakeman et Gottman, 1986). Deux observateurs, travaillant de façon indépendante, évaluent les sujets selon une grille commune; ceci nous permettra d'évaluer l'accord inter-observateurs. Le degré d'entente doit toujours être évalué lorsqu'on demande à des personnes d'observer un comportement afin de garantir la qualité des données recueillies.

L'accord doit être mesuré par l'indice Kappa de Cohen (1960) lorsque des catégories de fréquence sont utilisées et par des coefficients intra-classe (Shrout et Fleiss, 1979), lorsqu'on se sert d'une échelle de type Likert (ex. 1 à 6). Ces deux indices d'accord permettent de contrôler le taux d'accord dû à la chance. Un exemple aidera à mieux comprendre pourquoi il est si important de contrôler ce phénomène. Si un observateur doit choisir parmi trois catégories de comportement à chaque fois qu'il observe, il a trente pour-cent (30%) des chances de don-

ner la même réponse que l'autre observateur et ce, sans avoir à poser un jugement sur la cote qui décrit le mieux le comportement. Il est donc essentiel de tenir compte de ce facteur lorsqu'on évalue l'accord inter-observateurs. Si l'accord inter-observateurs est toujours mesuré dans les études répertoriées au tableau 1, le choix des indices pour l'évaluer ne tient pas compte du contrôle de la chance dans neuf des treize études. Ce choix limite de beaucoup la démonstration de la fiabilité des données.

La nécessité d'évaluer l'accord inter-observateurs est souvent perçue comme une exigence peu utile; ce n'est pas le cas. En effet, en plus de favoriser la répétition des résultats, un accord inter-observateurs élevé permet de diminuer l'erreur de mesure. Étant donné que la variation individuelle est élevée lorsqu'on utilise de petits groupes de sujets, toute réduction de l'erreur de mesure (dans ce cas-ci l'erreur due aux observateurs) améliore grandement la qualité des données. Des données fiables permettent de tirer des conclusions claires et solides et favorisent la possibilité de confirmer les hypothèses. La mesure de l'accord inter-observateurs est cependant une question dont l'élaboration dépasse le cadre de cet article; le lecteur peut se référer à deux sources particulièrement utiles: Bakeman et Gottman (1986) et Tinsley et Weiss (1975) pour une discussion approfondie.

De façon générale, l'utilisation de mesures d'observation, même globales, est préférable à celle d'instruments papier-crayon. Ceci est d'autant plus vrai avec certains types de clientèles, pour lesquelles une étude de validation sociale est à privilégier, comme les enfants, les personnes qui souffrent de troubles mentaux graves ou d'un handicap intellectuel. Avec ces personnes, les mesures papier-crayon sont, la plupart du temps, très peu utiles. Des troubles de concentration et de lecture, entre autres, réduisent de beaucoup la validité et la fiabilité de ce genre d'instrument.

Par ailleurs, dans certaines problématiques, en particulier lorsqu'elles s'expriment par de l'évitement (comme l'agoraphobie par exemple), les mesures papier-crayon peuvent s'avérer plus valides que des mesures comportementales. Il est en effet très difficile de recréer une mise en situation qui nous permettra d'observer l'évitement. Afin de déterminer s'il est nécessaire d'utiliser des mesures d'observation ou si des mesures d'auto-évaluation sont plus adéquates, il faut d'abord se demander si le sujet peut ou non nous donner une évaluation juste de son comportement. Si ce n'est pas le cas, les mesures d'observation doivent être utilisées. Si l'auto-évaluation du sujet est jugée valide, le chercheur peut alors choisir d'utiliser ou non des mesures comportementales. Enfin, lorsqu'on choisit d'utiliser l'observation, les variables mesurées doivent être sélectionnées avec parcimonie (Mahoney, 1978). Il vaut donc mieux mesurer peu de variables, mais le faire de façon fiable et valide.

## Conclusion

La méthode de validation sociale comporte de nombreux avantages. Le principal réside dans l'utilisation de critères d'évaluation de l'impact et d'objectifs de traitement plus valides écologiquement. Par validité écologique, on entend des critères d'impact ou des objectifs plus représentatifs du comportement ou de la perception moyenne de la population générale.

La sélection du groupe de référence est la pierre angulaire de la méthodologie de validation sociale. Les sujets du groupe non-clinique doivent être représentatifs de la population générale et le nombre de sujets doit être suffisant. Comme nous on peut le constater au tableau 1, dans la majorité des études de validation sociale examinées, ces critères essentiels ne sont pas satisfaits. Il n'est pas suffisant de comparer un groupe non-clinique à un groupe clinique pour réaliser un étude de validation sociale de qualité. En effet, l'utilisation d'un groupe de référence non représentatif peut nous conduire au choix d'objectifs de traitement non appropriés ou à conclure, à tort à l'impact du traitement. De plus, il n'est pas recommandé d'utiliser uniquement des mesures de validation sociale (le critère social) et d'ignorer les critères théorique (celui du thérapeute) et subjectif-expérientiel (celui du client) lorsqu'on mesure l'impact thérapeutique. La méthode de validation sociale est pertinente pour nous permettre de tenir compte du critère social. Toutefois, elle ne nous permet pas de saisir toutes les facettes du changement thérapeutique.

Si le choix du groupe de référence est l'élément essentiel pour réaliser une étude de validation sociale de qualité, les critères de qualité méthodologique reconnus pour la réalisation d'une étude clinique doivent aussi être pris en compte. Le groupe clinique doit être homogène quant au diagnostic et les variables confondantes doivent être contrôlées. Trop souvent, il n'est pas possible de tirer des conclusions solides parce qu'on a voulu inclure trop de catégories de diagnostics dans une même recherche. La réalisation d'une étude clinique est très exigeante; les chercheurs sont souvent tentés de vouloir répondre à plusieurs questions simultanément. Cette décision résulte souvent en des énergies mal investies et peu rentables, en particulier dans la sélection des groupes cliniques trop hétérogènes ou trop petit et l'utilisation trop rigoureuse de l'observation systématique.

Le recours à des mesures d'observation est souvent perçu comme une stratégie d'évaluation idéale dans le domaine de la thérapie comportementale. Elle l'est effectivement, mais uniquement dans les cas où elle est utilisée de façon adéquate et lorsqu'il est approprié de le faire. Sans une évaluation de l'accord inter-observateurs, selon un indice adéquat et, l'utilisation de données non seulement globales mais aussi de fréquence (descriptives), les mesures d'observation sont peu fiables, peu valides ou les deux. Il est donc préférable de restreindre le nombre de variables mesurées, mais d'utiliser une stratégie d'évaluation valide et fiable, que d'augmenter le nombre de dimensions étudiées.

Utilisée de façon adéquate et dans des contextes pertinents, la méthodologie de validation sociale permet de répondre à plusieurs questions cliniques. L'avantage principal de cette méthode est de fournir aux cliniciens des données comparatives concernant le comportement de la population générale face à des difficul-

tés que vivent leurs clients. En utilisant des objectifs de traitement et des critères d'impact plus valides, les cliniciens seront mieux outillés pour les traiter. Par ailleurs, si les chercheurs fournissent aux cliniciens des données plus pertinentes écologiquement, ceux-ci seront plus intéressés à les utiliser dans leur pratique clinique quotidienne. L'utilisation de la méthodologie de validation sociale est une stratégie de recherche qui devrait favoriser ce rapprochement entre la clinique et la recherche.

## Références

- Bakeman, R. et Gottman, J. M. (1986). *Observing interaction: An introduction to sequential analysis*. New York: Cambridge.
- Bellack, A. S. (1989a). Treatment outcome evaluation methodology with schizophrenics. *Advances of Behavior Research and Therapy*, 11, 191-200.
- Bellack, A. S. (1989b). (Éd.). *A clinical guide for the treatment of schizophrenia*. New York: Plenum Press.
- Bellack, A. S., Morrison, R. L., Wixted, J. T. et Mueser, K. T. (1990). An analysis of social competence in schizophrenia. *British Journal of Psychiatry*, 156, 809-818.
- Bellack, A. S., Morrison, R. L., Mueser, K. T., Wade, J. H. et Sayers, S. L. (1990). Role play for assessing the social competence of psychiatric patients. *Psychological Assessment*, 2, 248-255.
- Bergin, A. E. et Lambert, M. J. (1978). The evaluation of therapeutic outcomes. Dans S. L. Garfield et A. E. Bergin (Éd.). *Handbook of psychotherapy and behavior change*. New York: Wiley (pp. 139-189).
- Bertrand, R. (1986). *Pratique de l'analyse statistique des données*. Montréal: Les Presses de l'Université du Québec.
- Campbell, D. T. et Stanley, J. C. (1963). *Experimental and quasi-experimental design for research*. Chicago: Rand McNally.
- Castles, E. E. et Glass, C. R. (1986). Empirical generation of measures of social competence for mentally retarded adults. *Behavioral Assessment*, 8, 319-325.
- Chassan, J. B. (1979). *Research design in clinical psychology and psychiatry*. (éd. rev.) New York: Irvington.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- Cook, T. D. et Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field setting*. Chicago: Rand McNally.
- Donahoe, C. P., Carter, M. J., Bloem, W. D., Hirsch, G. L., Laasi, N. et Wallace, C. J. (1990). Assessment of interpersonal problem-solving skills. *Psychiatry*, 53, 329-339.
- Fawcett, S. B. (1991). Social validity: A note on methodology. *Journal of Applied Behavior Analysis*, 24, 235-239.
- Fingeret, A. L., Monti, P. M. et Paxson, M. A. (1985). Social perception, social performance and self-perception. *Behavior Modification*, 9, 345-356.
- Flanagan, R. (1986). Teaching young children responses to inappropriate approaches by strangers in public places. *Child and Family Behavior Therapy*, 8(2), 27-43.
- Garfield, S. L. (1978). Research problems in clinical diagnosis. *Journal of Consulting and Clinical Psychology*, 46, 596-607.
- Gottman, J. M. et Markman, H. J. (1978). Experimental designs in psychotherapy research. Dans S. L. Garfield et A. E. Bergin (Éd.). *Handbook of psychotherapy and behavior change* (pp. 23-62). New York: Wiley.

- Haley, W. E. (1985). Social skills deficits and self-evaluation among depressed and non depressed psychiatric inpatients. *Journal of Clinical Psychology*, 41, 163-168.
- Hansen, D. J., St-Lawrence, J. S. et Christoff, K. A. (1985). Effects of inter-personal problem-solving training with aftercare patients on problem-solving component skills and effectiveness of solutions. *Journal of Consulting and Clinical Psychology*, 53, 167-174.
- Hansen, D. J., Watson-Perzel, M. et Smith-Christopher, J. (1989). Clinical issues in social skills training with adolescents. *Clinical Psychology Review*, 9, 365-391.
- Hansen, D. J., St-Lawrence, J. S. et Christoff, K. A. (1989). Conversational skills of inpatient conduct disordered youths: Social validation of component behaviors and implications for skills training. *Behavior Modification*, 12, 424-444.
- Holmes, M. R., Hansen, D. J. et St-Lawrence, J. S. (1984). Conversational skills training with aftercare patients in the community: Social validation and generalization. *Behavior Therapy*, 15, 84-100.
- Horstman, A. M. (1988). Peer ratings of children's assertive and unassertive behavior: A social validation study. *Child and Family Behavior Therapy*, 10, 143-159.
- Jacobson, N. S., Follette, W. C. et Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, 15, 336-352.
- Kazdin, A. E. (1977). Assessing the clinical or applied importance of behavior change through social validation. *Behavior Modification*, 1, 427-451.
- Kazdin, A. E., Esveltd-Dawson, K., French, N. H. et Unis, A. S. (1987). Problem-solving skills training and relationship therapy in the treatment of antisocial child behavior. *Journal of Consulting and Clinical Psychology*, 55, 76-85.
- Kendall, P. C. et Grove, W. M. (1988). Normative comparisons in therapy outcome. *Behavioral Assessment*, 10, 147-158.
- Kraemer, H. C. (1981). Coping strategies in psychiatric clinical research. *Journal of Consulting and Clinical Psychology*, 49, 309-319.
- Mahoney, M. J. (1978). Experimental methods and outcome evaluation. *Journal of Consulting and Clinical Psychology*, 46, 660-672.
- Millbrook, J. M., Farrell, A. D. et Curran, J. P. (1986). Behavioral components of social skills: A look at subject and confederate behaviors. *Behavioral Assessment*, 8, 203-220.
- Minkin, N., Braukmann, C. J., Minkin, B. L., Timbers, G. D., Timbers, B. J., Fixsen, D. L., Philipps, E. L. et Wolf, M. M. (1976). The social validation and training of conversational skills. *Journal of Applied Behavior Analysis*, 9, 127-139.
- Monti, P. M. et Fingeret, A. L. (1987). Social perception and communication skills among schizophrenics and nonschizophrenics. *Journal of Clinical Psychology*, 43, 197-205.
- Mueser, K. T., Bellack, A. S., Douglas, M. S. et Morrison, R. L. (1991). Prevalence and stability of social skills deficits in schizophrenia. *Schizophrenia Research*, 5, 167-176.
- Mueser, K. T., Bellack, A. S., Morrison, R. L. et Wade, J. H. (1990). Gender, social competence, and symptomatology in schizophrenia: A longitudinal analysis. *Journal of Abnormal Psychology*, 99, 138-147.
- Shrout, P. E. et Fleiss, J. L. (1979). Approximate interval estimation for a certain intraclass correlation coefficient. *Psychometrika*, 43, 259-262.
- Storey, K. et Horner, R. H. (1991). An evaluative review of social validation research involving persons with handicaps. *The Journal of Special Education*, 25(3), 352-401.
- Strupp, H. H. et Hadley, S. (1977). A tripartite model of mental health and therapeutic outcomes; with special reference to negative effects in psychotherapy. *American Psychologist*, 32, 187-196.
- Tinsley, A. T. et Weiss, D. J. (1975). Interrater, reliability and agreement of subjective judgements. *Journal of Counseling Psychology*, 22, 358-376.

- Tisdelle, D. A. et St-Lawrence, J. S. (1986). Interpersonal problem-solving competency: Review and critique of the literature. *Clinical Psychology Review*, 6, 337-356.
- Van Houten, R. (1979). Social validation: The evolution of standards of competency for target behaviors. *Journal of Applied Behavior Analysis*, 12, 581-591.
- Vézina, J. et Ladouceur, R. (1984). La validation sociale: Considérations théoriques et pratiques. *Revue de modification du comportement*, 14, 66-76.
- Wolf, M. W. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding it's heart. *Journal of Applied Behavior Analysis*, 11, 203-214.
- Yeaton, W. H. et Sechrest, L. (1981). Critical dimensions in choice and maintenance of successful treatments: strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology*, 49, 156-167.

### Abstract

Social validation is a useful research method to answer clinical questions. This includes two strategies. In a social comparison study, a clinical group is compared to a non-clinical group, either before or after treatment. In a subjective evaluation study, members of the living community of the clinical sample are asked to evaluate the competence level of the clinical group. This paper discusses methodological criteria needed to execute a social validation study. Selection of the non-clinical group, homogeneity of the clinical one, and the choice of measurement devices are the most important aspects to be taken in account in a social validation study.