

**Recherche sur l'efficacité des  
programmes de traitement pour  
délinquants sexuels :  
Lignes directrices du CCDCR  
aux fins d'évaluation**

**Partie 1 : Introduction et aperçu**

**2007-02**

Membres du Comité de collaboration sur les données collectives  
relatives aux résultats (par ordre alphabétique) :

Anthony Beech, Guy Bourgon, R. Karl Hanson, Andrew J. R. Harris,  
Calvin Langton, Janice Marques, Michael Miner, William Murphy,  
Vernon Quinsey, Michael Seto, David Thornton, Pamela M. Yates

N° de cat. : PS4-38/1-2007F-PDF  
N° ISBN : 978-0-662-73902-9

## Préface

Le présent document s'adresse aux personnes qui s'intéressent activement à la recherche sur l'efficacité des programmes de traitement pour délinquants sexuels. Il vise donc trois types de lecteurs, soit les analystes qui désirent effectuer un examen critique des études actuelles sur l'efficacité des programmes de traitement pour délinquants sexuels, dont les rédacteurs en chef de périodiques, les évaluateurs scientifiques et les méta-analystes; les évaluateurs de programmes qui souhaitent déterminer les avantages que procure un programme de traitement en particulier, et les chercheurs qui mettent au point de nouvelles études visant à évaluer l'efficacité des méthodes de traitement qui s'adressent aux délinquants sexuels et à d'autres types de délinquants.

Après avoir examiné les échelles qui servent à évaluer la qualité des études, le Comité a déterminé qu'il en faut une nouvelle. En effet, les échelles existantes ne conviennent pas vraiment aux modèles couramment utilisés pour la recherche sur les délinquants sexuels. Par ailleurs, la plupart des échelles utilisées sont hétérogènes, comportant des éléments liés à la qualité des rapports, à l'éthique et à l'interprétation des données, au lieu de porter sur le biais et la validité interne.

Le présent document traite des délinquants sexuels, mais une bonne partie de l'analyse qu'il contient est également pertinente pour l'évaluation d'autres programmes prolongés et complexes visant à modifier le comportement des délinquants dont l'échec aux programmes peut causer des préjudices à autrui ou passer inaperçu pendant le traitement et parfois pendant des années (violence conjugale et conduite avec facultés affaiblies, par exemple).

Ce document est le fruit d'un travail de groupe. Le Comité de collaboration sur les données collectives relatives aux résultats (CCDCR) a été mis sur pied afin de faire progresser la recherche sur l'efficacité des programmes de traitement pour délinquants sexuels. Les membres du Comité ont été choisis en raison de leur expertise dans l'évaluation de la recherche sur les délinquants sexuels et de leur capacité d'apporter des points de vue différents. Les membres ont exprimé des opinions divergentes au sujet de l'efficacité des traitements, mais le but du projet était d'établir un terrain d'entente. Le Comité ne s'attendait pas et ne cherchait pas à en arriver à un consensus. Il voulait plutôt préciser les hypothèses communes à l'égard des caractéristiques que présentent les études crédibles et moins crédibles. En particulier, il propose un mécanisme plausible et fiable pour évaluer la qualité des études sur l'efficacité des programmes de traitement pour délinquants sexuels (les évaluateurs indépendants s'entendent sur la classification des études), lequel mécanisme a été largement accepté par des chefs de file dans le domaine.

Depuis 1980, il y a eu plus de 20 recensions sur l'efficacité des programmes de traitement pour délinquants sexuels. Même si les groupes expérimentaux accusent en moyenne un taux de récidive inférieur à celui des groupes témoins dans ces études, les évaluateurs ont noté des problèmes, ce qui limite la possibilité de tirer des conclusions rigoureuses. Si le Comité atteint

ses objectifs, les futurs chercheurs pourront présenter des conclusions avec une confiance et une précision accrues et seront plus en mesure d'évaluer la recherche sur l'efficacité des programmes de traitement.

## Table des matières

|  |    |
|--|----|
| Préface.....   | i  |
| Introduction.....  | 1  |
| Définition de la qualité d'une étude.....  | 2  |
| Hypothèses qui sous-tendent le modèle d'évaluation.....  | 2  |
| A) Il est possible et souhaitable d'évaluer la qualité de l'étude .....  | 3  |
| B) Les connaissances sont cumulatives .....  | 3  |
| C) Il faut faire appel à des méthodes multiples.....   | 4  |
| D) L'évaluation des programmes peut et doit contribuer à<br>accroître les connaissances .....                                    | 5  |
| Examen des méthodes employées pour évaluer la qualité des études .....   | 5  |
| Un modèle d'évaluation s'attachant particulièrement à l'efficacité<br>des programmes de traitement pour délinquants sexuels..... | 7  |
| Aperçu des Lignes directrices du CCDCR .....   | 8  |
| Fiabilité .....  | 9  |
| Application des Lignes directrices du CCDCR .....  | 11 |
| Bibliographie.....   | 13 |
| Membres du Comité.....   | 17 |



## Introduction

Les programmes de traitement pour délinquants sexuels donnent-ils des résultats? Il y a eu beaucoup de recherches à cet égard, mais les spécialistes continuent de discuter de l'efficacité des interventions visant à réduire le risque de récidive chez les délinquants sexuels.

Furby, Weinrott et Blackshaw (1989) ont conclu, dans un premier examen qui a fait autorité, que rien ne démontre que les programmes de traitement contribuaient à réduire les taux de récidive. Leurs conclusions allaient à l'encontre d'une étude de la même époque publiée par Solliciteur général Canada (Groupe de travail sur le traitement des délinquants sexuels, 1990), selon laquelle les programmes de traitement contribuaient à faire passer le taux de récidive de 25 % à 10 ou 15 % (p. 19).

Ces deux positions n'ont pas beaucoup changé au cours des quinze dernières années. Hall (1995), dans le cadre d'une méta-analyse englobant douze études publiées après l'examen mené par Furby et coll. (1989), a noté que les traitements cognitivo-comportementaux et hormonaux avaient un effet positif général. Or, l'étude de Hall (1995) a été critiquée à son tour, car elle englobait des études qui n'étaient pas suffisamment rigoureuses (Harris, Rice et Quinsey, 1998). Gallagher, Wilson, Hirschfield, Coggeshall et Mackenzie (1999) ont aussi effectué une nouvelle méta-analyse portant sur 22 études, tentant d'utiliser seulement les meilleures études possible. Ils ont cependant utilisé certaines études jugées imparfaites par Harris et coll. (1998) de même que les versions préliminaires de plusieurs études dont les résultats ont changé par suite d'une analyse plus poussée.

Le Comité de collaboration sur les données collectives relatives aux résultats a été formé en 1997 afin d'organiser les études sur l'efficacité des programmes de traitement pour délinquants sexuels et de promouvoir l'exécution d'évaluations de grande qualité. Dans son premier rapport, qui portait sur une méta-analyse regroupant 43 études (Hanson et coll., 2002), le Comité a conclu que les programmes de traitement actuels avaient un effet positif modeste, mais qu'il fallait multiplier et améliorer la recherche pour tirer des résultats concluants. Les résultats du Comité ont été reproduits dans le cadre d'un vaste examen entrepris par Lösel et Schmucker (2005), qui ont aussi constaté que les programmes de traitement cognitivo-comportementaux donnaient des résultats significatifs.

Comme il fallait s'y attendre, ces méta-analyses ont fait l'objet de critiques. En réponse au rapport de Hanson et coll. (2002), Rice et Harris (2003) ont fait remarquer que les résultats observés pouvaient facilement être expliqués par la possibilité d'un biais dans l'affectation des sujets aux groupes expérimentaux et aux groupes témoins. En outre, les traitements n'avaient eu aucun effet général selon les « meilleures » études définies par Rice and Harris (2003). De même, un examen de neuf études sur l'efficacité des programmes de traitement pour délinquants sexuels mené pour la Cochrane Collaboration n'a révélé aucun effet (Kenworthy, Adams, Bilby, Brooks-Gordon et Fenton, 2004; Brooks-Gordon, Bilby et Wells, 2006). Enfin, l'examen dirigé par Kenworthy et coll. (2004) est aussi digne de mention puisqu'il ne visait que

des études répondant à des critères bien établis par les chercheurs médicaux (comme l'affectation aléatoire).

Or, en ce qui a trait à la recherche sur les délinquants sexuels, les études exemplaires recensées par Rice et Harris (2003), par Kenworthy et coll. (2004) et par Hanson et coll. (2002) sont toutes différentes. Le problème n'était pas qu'un groupe de chercheurs était plus tolérant ou strict que l'autre à l'égard de la qualité des études, mais que la plupart des études jugées crédibles par les uns étaient considérées comme teintées d'erreurs systématiques par les autres. Par exemple, Kenworthy et coll. (2004) ont utilisé des études fondées sur des critères de résultats mesurant les changements rapportés par les participants à l'égard des caractéristiques psychologiques, tandis que Rice et Harris (2003) et Hanson et coll. (2002) ont exclu ces études, estimant que les mesures intermédiaires n'étaient pas suffisamment valides pour tirer des conclusions nettes. Une seule étude figurait parmi les « meilleures » études utilisées dans le cadre des trois examens : le Sex Offender Treatment and Evaluation Project de la Californie (SOTEP; Marques, Wiederanders, Day, Nelson et van Ommeren, 2005). Cette étude portant sur le traitement et l'évaluation des délinquants sexuels était particulière, car elle faisait appel à une méthodologie de recherche rigoureuse (affectation aléatoire) afin d'évaluer un programme de traitement crédible (cognitivo-comportemental) s'adressant à des délinquants sexuels adultes.

#### Définition de la qualité d'une étude

Pour évaluer la qualité d'une étude, il est nécessaire de définir ce qu'on évalue. Conformément aux recommandations du Potsdam Panel (Cook, Sackett et Spitzer, 1995), nous considérons comme de bonnes études celles qui réduisent au minimum le biais. Dans le cadre de l'étude parfaite, l'ampleur de l'effet calculée est entièrement attribuable aux différences dans le traitement (auxquels s'ajoute l'erreur aléatoire). Le biais constitue le principal critère permettant d'évaluer la qualité d'une étude. Cependant, il faut aussi tenir compte de la *confiance* qu'inspirent les résultats. Une étude reposant sur une affectation aléatoire, par exemple, ne devrait pas produire de différences systématiques entre les groupes. Néanmoins, il est possible de prêter foi davantage aux résultats lorsque les chercheurs examinent les différentes menaces à la validité des résultats et peuvent démontrer que l'étude a été mise en œuvre conformément aux objectifs. Par conséquent, une étude de grande qualité se définit comme une étude pour laquelle il est possible de poser avec *beaucoup de confiance* un jugement de *biais minimal*.

#### Hypothèses qui sous-tendent le modèle d'évaluation

Les Lignes directrices sur l'évaluation des recherches sur l'efficacité des programmes de traitement pour délinquants sexuels du Comité de collaboration sur les données collectives relatives aux résultats (Lignes directrices du CCDCR) se fondent sur les hypothèses ci-dessous.

### A) Il est possible et souhaitable d'évaluer la qualité de l'étude

Hypothèse initiale : la mesure dans laquelle les études peuvent aider à répondre aux questions des chercheurs varie, et il faut accorder plus de poids aux études bien faites qu'aux études de qualité inférieure. Or, cette hypothèse ne rallie pas tout le monde. Greenland (1994a, 1994b) affirme que l'évaluation de la qualité des études introduit un élément subjectif et a peu à voir avec les résultats (Greenland et O'Rourke, 2001). Il est difficile d'établir une mesure interne constante (un seul facteur) de la qualité des études, et il est encore plus difficile d'inférer une telle dimension dans le cas de rapports publiés. Au lieu d'utiliser des mesures globales de la qualité des études, Greenland recommande d'examiner les effets des éléments de qualité (éléments cotés; Greenland, 1994a; Greenland et O'Rourke, 2001). Par exemple, les méta-analyses pourraient évaluer si les études menées sur une longue période donnent des résultats différents de celles portant sur une courte période.

La position adoptée par Greenland mérite d'être étudiée sérieusement, puisque les méta-analyses peuvent donner des résultats différents selon les divers modèles d'évaluation de la qualité des études (Juni, Witschi, Bloch et Egger, 1999). Cependant, les chercheurs et les évaluateurs doivent poser un certain jugement quant à la qualité des études, même s'ils s'en tiennent à la décision dichotomique de déterminer si une étude en particulier doit être prise en compte ou non. En utilisant un nombre important d'études dont les attributs comportent des variantes sans corrélation, il est sans doute possible de traduire de manière empirique les effets des attributs de l'étude sur les résultats. Par contre, lorsqu'un faible nombre d'études comportent des caractéristiques corrélées, il est peu probable que le modèle statistique proposé par Greenland apporte beaucoup d'information.

Nous convenons avec les rédacteurs en chef de revues scientifiques que les évaluateurs et les chercheurs peuvent et doivent émettre un jugement concernant la qualité des études. Les lignes directrices relatives à l'évaluation de la qualité des études sont utiles, car elles permettent d'évaluer la recherche actuelle et peuvent aussi motiver les chercheurs à entreprendre de nouvelles études qui apporteront le plus d'information possible.

### B) Les connaissances sont cumulatives

Le milieu de la recherche – non unanime – cherche à déterminer si une seule étude définitive ou l'accumulation des résultats de plusieurs études mineures répondent le mieux aux questions posées. En médecine, l'étude définitive constitue souvent un essai clinique aléatoire regroupant plusieurs sites et s'étendant à des milliers de patients. Les résultats de ces études sont souvent convaincants, mais ces études sont longues et coûteuses. Elles ne sont menées que si des études mineures précédentes montrent que le projet a des chances raisonnables de réussir. Ironiquement, une méta-analyse

cumulative des études précédentes de portée réduite apporte souvent la même réponse que l'étude définitive, suscitant un débat sur la nécessité des essais cliniques d'envergure (Lau, Schmid et Chalmers, 1995). En fait, les deux types d'études sont nécessaires. Compte tenu des désaccords, l'essai clinique d'envergure est jugé plus convaincant que la synthèse d'études mineures diverses (LeLorier, Grégoire, Benhaddad, Lapierre et Derderian, 1997).

Or, il est peu probable qu'il y ait un jour une étude définitive sur les programmes de traitement pour délinquants sexuels, même si c'est très souhaitable. La complexité des interventions et les longues périodes nécessaires pour connaître les résultats définitifs (récidive) constituent d'importants obstacles techniques, et ce, même s'il existait une volonté sociale et politique d'investir des sommes importantes dans la recherche sur les délinquants sexuels. Par ailleurs, la population des délinquants sexuels étant hétérogène, il n'est pas possible de trouver les réponses dans une seule étude. Par conséquent, l'avenir de la recherche sur l'efficacité des programmes pour délinquants sexuels repose sur l'accumulation d'éléments de preuve tirés de petites études.

### C) Il faut faire appel à des méthodes multiples

La recherche est la mise en œuvre de bon sens et de l'intelligence. Il n'existe pas une seule méthode pour découvrir la vérité. Cependant, certaines solutions normalisées aux problèmes communs liés à la recherche sont passées dans la pratique au cours des cent dernières années. En particulier, l'affectation aléatoire a été reconnue comme approchant la perfection lorsqu'il s'agit de minimiser les différences préexistantes entre les groupes expérimentaux et les groupes témoins. L'affectation aléatoire n'élimine pas les différences. Toutefois, lorsqu'on y procède correctement, on peut s'attendre à ce que l'influence de ces différences donne une moyenne de zéro. Les études fondées sur l'affectation aléatoire ont fait l'objet de critiques, car elles peuvent avoir pour effet d'exclure la clientèle dangereuse. Cependant, l'affectation aléatoire constitue probablement l'approche la plus éthique en matière d'affectation aux programmes de traitement lorsque la demande excède les ressources.

Dans le cas d'interventions sociales complexes, les études prévoyant une affectation aléatoire se heurtent à d'importantes difficultés sur les plans conceptuel et pratique. Par conséquent, les chercheurs ont mis au point une gamme de modèles de rechange pour évaluer les problèmes sociaux (Cook et Campbell, 1979; Shadish, Cook et Campbell, 2002). Ces modèles sont souvent désignés par le terme « quasi-expériences », car le chercheur ne commande pas complètement le choix des personnes soumises à l'intervention. Il est généralement admis que les modèles quasi-expérimentaux peuvent contribuer de manière importante à élargir les connaissances, mais qu'il faut porter une attention particulière à leur conception, mise en œuvre et interprétation.

Les études portant sur des affectations aléatoires ont certes du mérite, et les chercheurs doivent trouver des occasions de faire appel à ces modèles. Ils ne constituent pas, par contre, la seule source d'information. Il importe d'utiliser diverses approches pour répondre aux différentes questions posées par la recherche. Dans le domaine du traitement psychique en général, les limites et les conséquences non intentionnelles des essais cliniques aléatoires sont de plus en plus reconnues (Haaga, 2004; Westen, Novotny et Thompson-Brenner, 2004). À notre avis, les connaissances sur l'efficacité des programmes de traitement pour délinquants sexuels reposeront sur l'utilisation de différentes méthodes de recherche. Aucun modèle pris seul ne peut être concluant, mais les résultats cumulatifs de différentes études contribueront à limiter de plus en plus les différentes interprétations plausibles.

#### D) L'évaluation des programmes peut et doit contribuer à accroître les connaissances

La plupart des études sur les programmes de traitement pour délinquants sexuels consistent en des évaluations de programmes, et non en des expériences scientifiques. Dans le cas d'expériences scientifiques, la recherche vise à répondre à des questions d'intérêt scientifique. Les résultats et les conséquences de l'expérience sont importants. Ce qui s'est produit dans le cadre de l'expérience n'est qu'un moyen d'actualiser les connaissances. À l'inverse, les évaluations de programme portent sur le déroulement d'un programme en particulier. Les administrateurs veulent savoir si un programme donné fonctionne (et non les programmes en général), et les décisions de financement reposent souvent sur les résultats de ces évaluations.

Nous sommes d'avis que les évaluations de programmes qui sont bien conçues peuvent contribuer à élargir les connaissances. Même si l'évaluation n'est pas conçue à titre d'étude de recherche, il est possible pour les évaluateurs de recueillir des renseignements qui aideront à répondre aux questions concernant l'efficacité de ce programme en particulier ou de programmes semblables. Par ailleurs, dans les cas où l'efficacité des programmes pour délinquants sexuels est discutable, les administrateurs qui parrainent les programmes de traitement se doivent d'évaluer leurs programmes et de contribuer à l'accumulation de connaissances sur les traitements qui sont efficaces chez les délinquants sexuels. Par conséquent, les Lignes directrices du CCDCR accordent une grande importance à la manière de tirer le meilleur parti des évaluations de programmes dans le but d'élargir les connaissances.

#### Examen des méthodes employées pour évaluer la qualité des études

Les évaluations formelles de la qualité des études sont une pratique relativement récente. Cependant, un nombre important d'échelles et de listes de contrôle ont été mises au point dans le domaine médical afin d'évaluer la qualité des essais aléatoires et cliniques (voir Juni et coll., 1999). Les évaluations de la qualité des études ont servi dans le cadre d'examen systématiques et de méta-analyses. De plus, les praticiens sont encouragés à utiliser les lignes directrices sur la

qualité des études pour évaluer de manière critique les études de recherche, et ce, dans le but d'améliorer les traitements offerts aux patients.

Moher et coll. (1995) ont recensé 25 échelles et neuf listes de contrôle, notant des dissemblances considérables entre celles-ci. Ainsi, le nombre de points évalués variaient de 3 à 34, et les caractéristiques méthodologiques et de présentation de rapports étaient aussi différentes. La plupart des échelles consacraient au moins un point à l'évaluation de l'affectation des patients, aux procédures de masquage et aux analyses statistiques. De plus, nombre d'entre elles comportaient des points visant à évaluer la qualité des rapports, les questions éthiques (consentement obtenu, par exemple) et l'interprétation des résultats.

Toutefois, ces échelles n'utilisaient pas une définition commune du terme « qualité ». Juni et coll. (1999) ont constaté que le choix de l'échelle de l'évaluation de la qualité avait une incidence sur les résultats de la méta-analyse. Ces chercheurs ont coté 17 études comparant l'efficacité de l'héparine de faible masse moléculaire à celle de l'héparine ordinaire dans la prévention de la thrombose postopératoire en utilisant 25 échelles d'évaluation de la qualité. Les études de « grande qualité » recensées à l'aide de ces échelles n'étaient pas les mêmes, et les meilleures études ainsi relevées donnaient des résultats différents. Les chercheurs en ont conclu que les échelles d'évaluation de la qualité étaient hétérogènes et ont remarqué que nombre des points évalués portaient sur la qualité des rapports, les questions éthiques et l'interprétation des données au lieu de traiter du biais et de la validité interne.

En criminologie, l'échelle de Maryland constitue l'une des échelles d'évaluation les plus influentes (Sherman et coll., 1997). Mise au point initialement afin d'aider à relever les programmes de prévention du crime prometteurs (voir aussi Aos, Phipps, Barnoski et Liebe, 1999), cette échelle a été utilisée par Lösel et Schmucker (2005) dans le cadre de leur méta-analyse portant sur des études de l'efficacité des programmes de traitement pour délinquants sexuels.

Les évaluateurs qui utilisent l'échelle de Maryland prennent en compte sept éléments relatifs à la « rigueur méthodologique » avant d'attribuer une cote générale. Ces éléments sont les suivants :

- 1) taille de l'échantillon;
- 2) type de groupes témoins;
- 3) utilisation de variables de contrôle pour expliquer les différences initiales entre les groupes;
- 4) caractère approprié des variables évaluées;
- 5) attrition;

6) durée du suivi et recours ou non à des tests statistiques.

Le barème variait de un à cinq points, cinq représentant le niveau le plus rigoureux.

La cote générale repose sur trois aspects importants :

- 1) la capacité de l'étude de limiter les variables parasites;
- 2) la marge d'erreur de mesure prévue;
- 3) l'efficacité statistique.

L'échelle de Maryland a l'avantage de s'appliquer à un large éventail d'études relatives à des interventions dans le domaine de la justice pénale. Par contre, comme la plupart des échelles d'évaluation utilisées dans le domaine médical, cette échelle omet d'utiliser une définition cohérente de la qualité et fait l'objet de réserves quant à l'efficacité statistique et le biais. L'échelle de Maryland suppose que les évaluateurs veulent connaître les conclusions des différentes études au lieu de recueillir des données aux fins d'une analyse secondaire. Or, dans l'évaluation des études aux fins de méta-analyse, les inquiétudes à l'égard de l'efficacité statistique ou de l'erreur de mesure cèdent le pas à celles liées au biais.

#### Un modèle d'évaluation s'attachant particulièrement à l'efficacité des programmes de traitement pour délinquants sexuels

Malgré les efforts considérables déployés afin de mettre au point des modèles d'évaluation de la qualité des études, aucune des échelles actuelles ne convient bien lorsqu'il s'agit de mesurer la qualité des études sur l'efficacité des programmes de traitement pour délinquants sexuels. Certains éléments communs sont utiles à la plupart des études, mais les menaces graves à la validité varient en fonction des questions. Par exemple, la masse corporelle, le régime alimentaire et l'exercice constituent des variables importantes dans le cas d'une étude sur l'efficacité du traitement du diabète. Or, les chercheurs qui s'intéressent aux délinquants sexuels voudront tenir compte d'autres variables, comme l'état matrimonial, l'impulsivité en ce qui a trait au mode de vie et l'équivalence des critères de la récidive. Pour évaluer la qualité des études, il faut une connaissance du problème à l'étude.

Les Lignes directrices du CCDRCR mettent l'accent sur les préoccupations particulières associées à la conception et à la mise en œuvre des études sur l'efficacité des programmes de traitement pour délinquants sexuels. Elles ont été conçues pour évaluer les études comparant un groupe expérimental de délinquants sexuels ayant suivi le traitement à un groupe témoin (ou normes) et utilisant la récidive comme critère de mesure des résultats. Étant donné que les critères importants de résultats peuvent être mesurés seulement bien des années après la fin du traitement, certains modèles sont difficiles à mettre en œuvre (p. ex., groupe témoin placé sur la liste d'attente, discontinuité de la régression) et ne font pas partie de la discussion. Les Lignes directrices sont plutôt axées sur les décisions qui doivent couramment être prises dans le cadre

d'études sur l'efficacité des programmes de traitement pour délinquants sexuels tels que le choix des variables de contrôle, les critères relatifs à la récidive et la taille de l'échantillon.

### Aperçu des Lignes directrices du CCDCR

Les Lignes directrices reposent sur un examen des échelles existantes d'évaluation de la qualité des études (p. ex., Cowley, 1995; Downs et Black, 1998; Gibbs, 1989; Miner, Murphy et Yates, 2002; Reisch, Tyson et Mize, 1989; Sherman et coll., 1997; Thomas, Ciliska, Dobbins et Micucci, 2004; Wortman, 1994; Zaza et coll., 2000) et des observations particulières soulevées au sujet de la recherche sur les délinquants sexuels (p. ex., Rice et Harris, 2003). Le contenu et la structure des Lignes directrices du CCDCR sont, en grande partie, tirés d'une analyse des méthodes employées par les membres du CCDCR pour décrire les forces et les faiblesses des études de recherche individuelles. Nous avons envisagé au départ d'établir des critères distincts en fonction de modèles particuliers (p. ex., affectation aléatoire, cohorte), mais les observations soulevées pour les différents modèles étaient remarquablement semblables. Par conséquent, les Lignes directrices du CCDCR prévoient des critères généraux à l'égard de l'évaluation des études sur l'efficacité des programmes de traitement pour délinquants sexuels et présentent quelques rares questions distinctes relatives à des modèles de recherche particuliers.

Les Lignes directrices du CCDCR comportent 20 points répartis entre sept catégories :

- 1) contrôle administratif de la variable explicative;
- 2) attentes de l'expérimentateur;
- 3) taille de l'échantillon;
- 4) attrition;
- 5) équivalence des groupes;
- 6) variables des résultats;
- 7) comparaison correcte.

Elles prévoient aussi un point à évaluer seulement dans le cas de modèles englobant plusieurs établissements (taille de l'échantillon des établissements), et trois listes de contrôle visant à aider à évaluer le point 13 (équivalence des groupes *a priori*) pour certains modèles (affectation aléatoire, gamme de risque comparée à la norme, cohorte). Elles contiennent aussi un diagramme adapté de celui de Zaza et coll. (2000), afin d'aider les évaluateurs à déterminer la catégorie des études.

Les 20 (21) points à évaluer visent à déterminer la mesure dans laquelle les caractéristiques de l'étude introduisent un élément de biais dans l'évaluation de l'efficacité du traitement ou

influencent la confiance que l'on peut accorder aux conclusions de l'étude. S'ils disposent de peu de renseignements, les évaluateurs sont encouragés à en obtenir de nouveaux et à réévaluer le point en question.

Le jugement général posé à l'égard de la qualité de l'étude constitue une sorte de jugement structuré. Après avoir évalué chaque point, les évaluateurs sont invités à émettre des jugements globaux sur la place inhérente du biais dans le modèle de recherche et sur la « confiance » à accorder à la cote du biais. Les choix sont les suivants : a) aucun biais ou biais minime; b) certain biais; c) biais considérable. L'évaluation de la confiance repose aussi sur une échelle à trois points : a) aucune confiance ou très peu de confiance dans les résultats; b) certain degré de confiance; c) confiance dans les résultats rapportés. Selon les cotes attribuées pour la confiance et le biais, les études sont classées dans l'une ou l'autre des quatre catégories ci-dessous :

1) DÉTERMINANTE

Confiance élevée que le biais est minime dans l'évaluation de l'efficacité du programme de traitement pour délinquants sexuels. Il s'agit d'études bien conçues et bien exécutées qui donnent des résultats convaincants. Elles comportent des défauts mineurs qui sont toutefois peu susceptibles d'influencer les conclusions principales ou de changer la direction des effets observés.

2) BONNE

Confiance élevée que le biais est faible (cote intermédiaire). Des efforts raisonnables ont été déployés afin d'éliminer les menaces à la validité, mais il reste beaucoup d'impondérables.

3) FAIBLE

Certain degré de confiance que le biais est faible (cote intermédiaire). Ces études comportent de graves lacunes, mais peuvent être pertinentes en ce qui concerne la question de l'efficacité du traitement. Au plus, elles apportent une preuve non concluante.

4) REJETÉE

Faible confiance à l'égard des résultats *ou* biais considérable. Ces études comportent un grand nombre de lacunes graves. La démarche utilisée a pour effet d'introduire un biais grave, ou encore l'étude ne donne pas les renseignements importants nécessaires pour éliminer les autres explications plausibles des résultats.

### Fiabilité

Deux étudiantes de premier cycle (3<sup>e</sup> année de criminologie; 4<sup>e</sup> année de psychologie) ont reçu une formation de cinq jours sur l'utilisation des Lignes directrices du CCDCR<sup>1</sup>. Durant cette

---

<sup>1</sup> Leslie Helmus, Shannon Hodgson.

formation, elles ont surtout procédé à l'évaluation de huit études de pratique avec un formateur<sup>2</sup>. Les deux évaluatrices ont ensuite évalué de façon autonome dix études.

En ce qui a trait à l'évaluation globale, les évaluatrices ont obtenu des résultats concordants pour neuf des dix études (coefficient de corrélation intraclasse ou CCI = 0,95). Le degré de concordance était de 100 % en ce qui concerne la confiance globale (*Kappa* = 1,0; CCI = 1,0), de 90 % en ce qui a trait au biais global (CCI = 0,69; impossibilité d'effectuer le calcul pour *Kappa*), et de 70 % pour ce qui est de la direction globale du biais (impossibilité d'effectuer le calcul pour *Kappa*). Le degré de concordance pour chacun des points était aussi élevé, la médiane étant de 1,0 pour la plupart des catégories.

Une deuxième étude de fiabilité a été menée en faisant appel à douze spécialistes dans le domaine de la recherche sur les délinquants sexuels<sup>3</sup>. Dix études hypothétiques ont servi à déterminer la fiabilité des évaluations auxquelles s'étaient livrés les spécialistes. Il n'était pas question d'utiliser des études réelles, sur lesquelles les spécialistes se seraient déjà sûrement formés une opinion en tant qu'auteurs ou évaluateurs. Les spécialistes n'ont pas reçu de formation particulière sur l'application des Lignes directrices du CCDRC, mais la moitié d'entre eux les connaissaient en tant que membres du CCDRC.

Le degré de concordance parmi les spécialistes était faible. Ils ont obtenu un taux de concordance moyen pour les points individuels, mais ils en sont arrivés à des résultats différents pour ce qui est des évaluations globales. Ils ont attribué la même cote à seulement trois des dix études (tous des rejets). Certaines différences étaient attribuables à des erreurs qu'il aurait été possible de corriger en assurant une formation adéquate et en faisant un peu plus attention (défaut de remarquer certaines caractéristiques des études, mauvaise interprétation des règles). Il est cependant ressorti des discussions tenues par la suite par les spécialistes que ceux-ci ne s'entendaient pas en principe sur ce qui caractérisait, au minimum, une étude « suffisamment » bonne.

Ils ont convenu en majorité par la suite que les caractéristiques définies par les Lignes directrices du CCDRC constituaient des indices importants de la qualité des études, mais ils ont exprimé des opinions divergentes quant à l'importance relative de ces caractéristiques par rapport à la qualité générale des études.

Essentiellement, il est ressorti des études sur la fiabilité qu'il est possible de former des évaluateurs novices pour qu'ils puissent utiliser de manière fiable les Lignes directrices du CCDRC. Par contre, l'exposition à ces Lignes directrices ne suffit pas en soi à changer des convictions bien ancrées quant à la méthodologie appropriée dans l'évaluation de l'efficacité des programmes pour délinquants sexuels (pour des résultats semblables à l'égard de la recherche médicale, voir Schroter et coll., 2004). Nous avons supposé au départ que les

---

<sup>2</sup> Guy Bourgon.

<sup>3</sup> Guy Bourgon, Andrew Harris, Grant Harris, Niklas Långström, Roxanne Lieb, Ruth Mann, Robert McGrath, William Murphy, Vernon Quinsey, Marnie Rice, David Thornton, Pamela Yates.

principes généraux de la conception des recherches s'imposeraient de manière intuitive aux spécialistes, mais nous nous sommes trompés à cet égard.

Bon nombre de spécialistes ont soutenu avec passion une définition différente d'une « bonne » étude. Néanmoins, tous les spécialistes ont convenu que les caractéristiques utilisées dans les Lignes directrices du CCDCR étaient importantes pour évaluer la qualité des études.

### Application des Lignes directrices du CCDCR

Les Lignes directrices du CCDCR seront sûrement utiles à trois égards :

- 1) examiner les études existantes;
- 2) évaluer les programmes existants;
- 3) concevoir de nouvelles études sur l'efficacité des programmes de traitement.

Les évaluateurs peuvent utiliser les lignes directrices à titre de critères de sélection dans le cadre de synthèses narratives ou quantitatives des preuves de l'efficacité des programmes de traitement pour délinquants sexuels. Elles seront également utiles aux chefs de rédaction de revues professionnelles (et aux évaluateurs), qui pourront s'en servir pour évaluer la qualité des études et donner des conseils sur les améliorations à apporter. Enfin, elles proposeront aux concepteurs de programmes des caractéristiques visant à faciliter les évaluations futures (p. ex., recueillir périodiquement des renseignements sur les personnes qui n'ont pas été admises dans le cadre du programme).

Les évaluateurs de programmes doivent souvent déterminer l'efficacité du traitement dans des conditions qui sont loin d'être parfaites. Ils peuvent toutefois utiliser les Lignes directrices pour prendre des décisions relatives à la conception de manière à tirer le plus d'informations possible à un coût minime. Par exemple, ils peuvent axer les ressources limitées consacrées à l'évaluation sur les variables de risque pertinentes et les échelles de risque établies; ils peuvent bien préciser les critères d'inclusion et ils peuvent faire en sorte que les analyses englobent tous les délinquants qui doivent suivre le traitement (intention de traitement).

Nous recommandons aux chercheurs qui ont l'occasion de mettre sur pied de nouvelles études de recherche d'utiliser des modèles de recherche solides tels que l'affectation aléatoire. Nous recommandons également de déterminer, avant l'affectation au traitement, que les délinquants posent un risque semblable. Les études reposant sur une affectation aléatoire ne sont pas populaires sur le plan politique et sont difficiles à mettre en œuvre, mais elles comportent de tels avantages que les chercheurs devraient les préconiser chaque fois que cela est possible. Les chercheurs qui décident de procéder à des études reposant sur une affectation aléatoire doivent toutefois s'attendre à une détérioration du processus de randomisation. Par conséquent, nous recommandons que tous les participants (groupes expérimentaux et groupes témoins) soient soumis avant leur traitement à une évaluation des facteurs de risque pertinents, et que les

chercheurs se montrent vigilants à l'égard de problèmes tels que l'intégrité du traitement, l'attrition et les chevauchements (groupes témoins recevant des services équivalents). En outre, les études spécialisées doivent porter sur un traitement qui devrait, selon des attentes raisonnables, être efficace (p. ex., Andrews et Bonta, 2006, chapitre 10). Il serait en effet inapproprié de représenter une étude sur le traitement des délinquants sexuels si l'intervention n'est pas crédible selon les normes actuelles.

## Bibliographie

- ANDREWS, D.A. et J.S. BONTA. *The psychology of criminal conduct* (4th edition). Cincinnati, Ohio : Anderson, 2006
- AOS, S., P. PHIPPS, R. BARNOSKI et R. LIEB. *The comparative costs and benefits of programs to reduce crime: A review of national research findings with implications for Washington State* (Document n° 99-05-1202). Olympia, Washington : Washington State Institute for Public Policy, 1999.
- BROOKS-GORDON, B., C. BILBY et H. WELLS. « A systematic review of psychological interventions for sexual offenders I: Randomised control trials », *Journal of Forensic Psychiatry and Psychology*, 17, p. 442-466, 2006.
- COOK, D.J., D.L. SACKETT et W.O. SPITZER. « Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam consultation on meta-analysis », *Journal of Clinical Epidemiology*, 48, p. 167-171, 1995.
- COOK, T.D. et D.T. CAMPBELL. *Quasi-experimentation: Design and analysis issues for field settings*. Boston : Houghton Mifflin, 1979.
- COWLEY, D.E. « Prostheses for primary total hip replacement: A critical appraisal of the literature », *International Journal of Technology Assessment in Health Care*, 11, p. 770-778, 1995.
- DOWNS, S.H. et N. BLACK. « The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions », *Journal of Epidemiology and Community Health*, 52, p. 377-384, 1998.
- FURBY, L., M.R. WEINROTT et L. BLACKSHAW. « Sex offender recidivism: A review », *Psychological Bulletin*, 105, p. 3-30, 1989.
- GALLAGHER, C. A., D.B. WILSON, P. HIRSCHFIELD, M.B. COGGESHALL et D.L. MACKENZIE. « A quantitative review of the effects of sex offender treatment on sexual reoffending », *Corrections Management Quarterly*, 3, p. 19-29, 1999.
- GIBBS, L.E. « Quality of study rating form: An instrument for synthesizing evaluation studies », *Journal of Social Work Education*, 25, p. 55-67, 1989.
- GREENLAND, S. « Invited commentary: A critical look at some popular meta-analytic methods », *American Journal of Epidemiology*, 140, p. 290-296, 1994a.

- GREENLAND, S. « Quality scores are useless and potentially misleading. Reply to “Re: A critical look at some popular meta-analytic methods” », *American Journal of Epidemiology*, 140, p. 300-301, 1994b.
- GREENLAND, S. et K. O’ROURKE. « On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions », *Biostatistics*, 2, p. 463-471, 2001.
- GROUPE DE TRAVAIL SUR LE TRAITEMENT DES DÉLINQUANTS SEXUELS. *Gestion et traitement des délinquants sexuels*. Ottawa : Solliciteur général Canada, 1990.
- HAAGA, D.A.F. « A healthy dose of criticism for randomized trials: Comment on Westen, Novotny, and Thompson-Brenner (2004) », *Psychological Bulletin*, 130, p. 674-676, 2004.
- HALL, G.C.N. « Sexual offender recidivism revisited: A meta-analysis of recent treatment studies », *Journal of Consulting and Clinical Psychology*, 63, p. 802-809, 1995.
- HANSON, R. K., A. GORDON, A.J.R. HARRIS, J.K. MARQUES, W. MURPHY, V.L. QUINSEY et M.C. SETO. « First report of the Collaborative Outcome Data Project on the effectiveness of psychological treatment of sex offenders », *Sexual Abuse: A Journal of Research and Treatment*, 14, p. 169-194, 2002.
- HARRIS, G.T., M.E. RICE et V.L. QUINSEY. « Appraisal and management of risk in sexual aggression: Implications for criminal justice policy », *Psychology, Public Policy, and Law*, 4, p. 73-115, 1998.
- JUNI, P., A. WITSCHI, R. BLOCH et M. EGGER. « The hazards of scoring the quality of clinical trials for meta-analysis », *Journal of the American Medical Association*, 282, p. 1054-1060, 1999.
- KENWORTHY, T., C.E. ADAMS, B. BROOKS-GORDON et M. FENTON. « Psychological interventions for those who have sexually offended or are at risk of offending (Cochrane Review) », *Cochrane Library*, Issue 3. Chichester, UK : John Wiley & Sons, 2004.
- LAU, J., C.H. SCHMID et T.C. CHALMERS. « Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care », *Journal of Clinical Epidemiology*, 48, p. 45-57, 1995.
- LELORIER, J., G. GRÉGOIRE, A. BENHADDAD, J. LAPIERRE et F. DERDERIAN. « Discrepancies between meta-analyses and subsequent large randomized, controlled trials », *The New England Journal of Medicine*, 337, p. 536-542, 1997.
- LÖSEL, F. et M. SCHMUCKER. « The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis », *Journal of Experimental Criminology*, 1, p. 117-146, 2005.

- MARQUES, J.K., M. WIEDERANDERS, D.M. DAY, C. NELSON et A. VAN OMMEREN. « Effects of a relapse prevention program on sexual recidivism: Final results from California's Sex Offender Treatment and Evaluation Project (SOTEP) ». *Sexual Abuse: A Journal of Research and Treatment*, 17, p. 79-107, 2005.
- MINER, M., W. MURPHY et P.M. YATES. *Research criteria for ATSA Collaborative Data Project: Subcommittee Report*. Unpublished manuscript, 2002.
- MOHER, D., A.R. JADAD, G. NICHOL, M. PENMAN, P. TUGWELL et S. WALSH. « Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists », *Controlled Clinical Trials*, 16, p. 62-73, 1995.
- REISCH, J.S., J.E. TYSON et S.G. MIZE. « Aid to the evaluation of therapeutic studies », *Pediatrics*, 84, p. 815-824, 1989.
- RICE, M.E. et G.T. HARRIS. « The size and sign of treatment effects in sex offender therapy », *Annals of the New York Academy of Sciences*, 989, p. 428-440, 2003.
- SCHROTER, S., N. BLACK, S. EVANS, J. CARPENTER, F. GODLEE et R. SMITH. « Effects of training on quality of peer review: Randomized controlled trial », *British Medical Journal*, 328, p. 673-675, 2004.
- SHADISH, W.R., T.D. COOK et D.T. CAMPBELL. *Experimental and quasi-experimental designs for generalized causal inference*. Boston : Houghton Mifflin, 2002.
- SHERMAN, L.W., D. GOTTFREDSON, D. MACKENZIE, J. ECK, P. REUTER et S. BUSHWAY. *Preventing crime: What works, what doesn't, what's promising. A report to the United States Congress*. College Park, Maryland : University of Maryland, Department of Criminology and Criminal Justice, 1997.
- THOMAS, H., D. CILISKA, M. DOBBINS et S. MICUCCI. « A process for systematically reviewing the literature: Providing the research evidence for public health nursing interventions », *Worldviews on Evidence-Based Nursing*, 2, p. 91-99, 2004.
- WESTEN, D., C.M. NOVOTNY et H. THOMPSON-BRENNER. « The empirical status of empirically supported psychotherapies: Assumptions, findings and reporting in controlled clinical trials ». *Psychological Bulletin*, 130, p. 631-663, 2004.
- WORTMAN, P.M. « Judging research quality ». In H. Cooper & L.V. Hedges, (Eds.), *The handbook of research synthesis* (p. 97-109). New York : Russell Sage Foundation, 1994.
- ZAZA, S., L.K. WRIGHT-DE AGÜERO, P.A. BRISS, B.I. TRUMAN, D.P. HOPKINS, M.H. HENNESSY et al. « Data collection instrument and procedure for systematic reviews in the Guide to Community Preventive Services », *American Journal of Preventive Medicine*, 18, p. 44-74, 2000.



## Membres du Comité

Anthony R. Beech, Ph.D., est professeur de psychologie criminelle au Centre for Forensic and Family Psychology, à l'École de psychologie de l'Université de Birmingham, au Royaume-Uni, et membre de la Société britannique de psychologie. a.r.beech@bham.ac.uk

Guy Bourgon, Ph.D., est chercheur à Sécurité publique Canada et professeur auxiliaire au Département de psychologie de l'Université Carleton, à Ottawa, au Canada. Guy.Bourgon@sp-ps.gc.ca

R. Karl Hanson, Ph.D., est chercheur principal à Sécurité publique Canada et professeur auxiliaire au Département de psychologie de l'Université Carleton, à Ottawa, au Canada. Karl.Hanson@sp-ps.gc.ca

Andrew J.R. Harris, Ph.D., est gestionnaire principal de recherche, à la Direction de la recherche, au Service correctionnel du Canada. Il n'est pas professeur. HarrisAJ@csc-scc.gc.ca

Calvin M. Langton, Ph.D., est professeur auxiliaire au Département de psychiatrie de l'Université de Toronto, au Canada, et membre honoraire de la School of Community Health Sciences, de l'Université de Nottingham, au Royaume-Uni. calvin.langton@utoronto.ca

Janice Marques, Ph.D., est une psychologue consultante qui a pris sa retraite récemment du Département de la santé mentale de la Californie. Elle a été présidente de l'ATSA lorsque ce projet collectif a été lancé en 1998. jkmarques@sbcglobal.net

Michael H. Miner, Ph.D., est professeur associé au Programme de la sexualité humaine, du Département de médecine familiale et de santé communautaire de l'Université du Minnesota, à Minneapolis, au Minnesota. miner001@umn.edu

William Murphy, Ph.D., est professeur au Département de psychiatrie au Centre des sciences de la santé de l'Université du Tennessee, à Memphis, au Tennessee. wmurphy@utm.edu

Michael Seto, Ph.D., est psychologue au Programme de droit et de santé mentale du Centre for Addiction and Mental Health, et professeur associé au Département de psychiatrie et au Centre de criminologie de l'Université de Toronto. Michael\_Seto@camh.net.

Vernon Quinsey, est professeur et chef du Département de psychologie à l'Université Queen, à Kingston, en Ontario. vernon.quinsey@queensu.ca

David Thornton, Ph.D., est directeur clinique au Sand Ridge Secure Treatment Center, à Mauston, au Wisconsin. thorndm@dhfs.state.wi.us

Pamela M. Yates, Ph.D. est psychologue au Service correctionnel du Canada et spécialiste dans le traitement des délinquants sexuels. YatesPM@csc-scc.gc.ca