

Recherche sur les questions correctionnelles :
Rapport pour spécialistes

**L'ERRRS, la Statique-99R et
la Statique-2002R ajoutent
réciproquement à leur validité
prédictive quant au risque
de récidive des
délinquants sexuels**

2011-02

Kelly M. Babchishin, R. Karl Hanson et Leslie Helmus

Sécurité publique Canada

Sommaire

Les échelles actuarielles fondées sur des données empiriques sont de plus en plus utilisées en psychologie, particulièrement pour évaluer le risque de criminalité et de violence. Bien que les évaluateurs recourent régulièrement à plus d'une échelle, nous savons pas avec certitude comment interpréter les résultats divergents. La présente étude visait à examiner l'exactitude prédictive et la validité ajoutée de trois échelles (ERRRS, Statique-99R et Statique-2002R) pour vingt échantillons distincts de délinquants sexuels ($N = 7\,491$). La Statique-99R et la Statique-2002R se sont révélées supérieures à l'ERRRS pour évaluer le risque de récidive sexuelle, de récidive avec violence et de récidive en général. Aucune différence sur le plan de l'exactitude prédictive n'a été observée entre la Statique-99R et la Statique-2002R. Néanmoins, elles ont toutes trois ajouté l'une l'autre à leur validité prédictive quant au risque de récidive, peu importe sa nature. Le sens des effets différentiels n'était cependant pas toujours positif. Une fois neutralisés les autres instruments de mesure, les scores élevés à l'ERRRS étaient associés à un risque de récidive diminué avec violence et de récidive en général. Par conséquent, les décisions sur l'interprétation des résultats d'échelles d'évaluation multiples doivent être guidées par la validité conceptuelle de ces échelles. Lorsque les échelles évaluent les mêmes domaines de facteurs de risque, il peut être justifié de faire la moyenne des résultats. Cependant, si les échelles ne portent pas sur les mêmes types de facteurs de risque, il faudrait aux évaluateurs un modèle défendable en ce qui concerne 1) les traits psychologiques latents évalués par les échelles et 2) les données empiriques décrivant la manière dont les concepts devraient être pondérés et combinés.

Note des auteurs

Les opinions exprimées dans le présent document sont celles des auteurs et ne traduisent pas nécessairement celles de Sécurité publique Canada. Prière d'acheminer toute correspondance à propos du présent rapport à l'adresse suivante : R. Karl Hanson, Recherche sur les questions correctionnelles, Sécurité publique Canada, 340, rue Laurier Ouest, Ottawa (Ontario), Canada, K1A 0P8.
Courriel : Karl.Hanson@ps-sp.gc.ca.

Remerciements

Nous tenons à remercier les chercheurs suivants de nous avoir accordé la permission d'utiliser leurs données et d'avoir patiemment répondu à nos nombreuses questions : Alfred Allan, Tony Beech, Susanne Bengtson, Jacques Bigras, Sasha Boer, Jim Bonta, Sébastien Brouillette-Alarie, Franca Cortoni, Margretta Dwyer, Reinhard Eher, Doug Epperson, Randolph Grace, Andy Haag, Leigh Harkins, Andreas Hill, Steve Johansen, Ray Knight, Niklas Långström, Terry Nicholaichuk, Kevin Nunes, Jean Proulx, Martin Rettenberger, Rebecca Swinburne Romine, Daryl Ternowski, Robin Wilson et Annie Yessine.

Renseignements sur le produit

Janvier 2011

N° de Cat : PS3-1/2011-2F-PDF

N° ISBN 978-1-100-97023-3

Ottawa

L'ERRRS, la Statique-99R et la Statique-2002R ajoutent réciproquement à leur validité prédictive quant au risque de récidive des délinquants sexuels

La plupart des tests psychologiques sont conçus pour évaluer les traits psychologiques latents, et leurs résultats ont une importance pratique dans la mesure où les traits psychologiques latents sont liés aux variables étudiées. Il n'est pas toujours nécessaire, bien que ce soit souhaitable, de comprendre parfaitement les traits psychologiques latents évalués pour qu'un instrument d'évaluation ait une utilité pratique. En fait, il est rare que ce soit le cas (Cronbach et Meehl, 1955). Même si les spécialistes ne s'entendent toujours pas sur la nature des principaux traits psychologiques (p. ex. bonne santé mentale, intelligence, déviance sexuelle), ils conviennent tout de même de l'utilité pratique des instruments d'évaluation existants au moment de prendre une décision dans un contexte appliqué (p. ex. fin du traitement, inscription dans un établissement d'enseignement, évaluation du risque). Les instruments d'évaluation peuvent avoir de l'importance simplement en raison de leurs relations empiriques avec les variables étudiées (p. ex. Meehl, 1956). Ce type de prédiction empirique est particulièrement pertinent lorsque la préoccupation première de l'évaluateur est de prévoir la valeur (oui ou non) d'une variable discrète (p. ex. rechute de dépression, échec scolaire, récidive sexuelle).

Un des domaines dans lesquels les prédictions empiriques se sont imposées au cours des dernières années est celui de l'évaluation du risque de criminalité ou de violence (Hanson, 2005, 2009; Quinsey, Harris, Rice et Cormier, 2006). Aux États-Unis, c'est le test de Daubert (*Daubert c. Merrell Dow Pharmaceuticals, Inc.*, 1993) qui est le plus souvent utilisé pour déterminer si les données scientifiques (p. ex. les facteurs de risque) sont recevables devant les tribunaux (Monahan et Walker, 2010). Selon le test de Daubert, le témoignage doit être fondé sur des données empiriques, mais le spécialiste n'a pas besoin de convaincre la cour qu'elle doit avoir une « compréhension globale » (*Daubert c. Merrell Dow Pharmaceuticals, Inc.*, 1993, par. 43) des questions dont elle est saisie. Ainsi les cours américaines acceptent régulièrement des données empiriques sur les facteurs de risque de criminalité ou de violence sans nécessairement comprendre les liens de causalité inhérents.

Bien qu'il existe un consensus quant au fait que les facteurs de risque doivent être fondés sur des données empiriques (p. ex. Kraemer et coll., 1997), les évaluateurs ne s'entendent pas sur la meilleure façon de combiner les facteurs de risque en vue d'effectuer une évaluation globale. La recherche a constamment révélé que les évaluations structurées du risque sont plus précises que les jugements professionnels non structurés (Gendreau, Goggin et Law, 1997; Grove, Zald, Lebow, Snitz et Nelson, 2000; Hanson et Morton-Bourgon, 2009), mais il n'y a pas de consensus sur la façon de structurer les évaluations du risque.

Dans le domaine de l'évaluation du risque de violence, la plupart des évaluateurs recourent à une certaine forme de jugement professionnel structuré ou JPS (Archer, Buffington-Vollum, Stredny et Handel, 2006). Dans ce type d'évaluation, les facteurs de risque sont choisis d'avance en fonction de leur lien avec la variable étudiée. Le regroupement de ces facteurs dans une évaluation globale est toutefois laissé au jugement de l'évaluateur (Douglas et Kropp, 2002). Contrairement au JPS, les instruments d'évaluation *structurés* précisent à l'avance les facteurs et fournissent des méthodes explicites sur la façon de combiner les facteurs pour arriver à un score total (Grove et coll., 2000). Lorsque les instruments d'évaluation structurés fournissent également des estimations de probabilité tirées d'observations empiriques à propos d'une variable étudiée en particulier, il est alors question d'*échelles actuarielles* (Dawes, Faust et Meehl, 1989; Meehl, 1954).

L'utilisation d'échelles actuarielles est courante dans le cas de certaines évaluations du risque lorsque l'enjeu est considérable. Dans le cas des procédures visant l'internement civil des délinquants sexuels, par exemple, 95 % des évaluateurs disent toujours ou presque toujours utiliser la Statique-99 (échelle actuarielle d'évaluation du risque de récidive sexuelle; Jackson et Hess, 2007). Cette façon de faire tranche avec celle en cours dans le domaine de la psychologie clinique générale, où la majorité des psychologues (68 %) compte sur le jugement clinique non structuré (Vrieze et Grove, 2009).

Recourir aux instruments d'évaluation structurés et aux échelles actuarielles comporte des avantages clairs (p. ex. biais réduit, fiabilité élevée; Garb, 2003), mais l'utilisation routinière présente des obstacles. Souvent lorsqu'il faut prendre une décision dans un contexte appliqué, les instruments d'évaluation validés n'existent tout simplement pas (Vrieze et Grove, 2009). La présente étude porte toutefois sur le problème inverse : que doivent faire les évaluateurs lorsqu'il existe plusieurs instruments d'évaluation du risque?

Utilisation d'instruments d'évaluation multiples

Lorsqu'ils doivent évaluer le risque de récidive des délinquants sexuels, les évaluateurs ont l'embarras du choix. Ils disposent de divers instruments, dont les suivants : Échelle spécialisée pour délinquants sexuels Minnesota – Révisée (MnSOST-R; Epperson et coll., 1998), Évaluation rapide du risque de récidive sexuelle (ERRRS; Hanson, 1997), Risque de violence sexuelle-20 (SVR-20; Boer, Hart, Kropp et Webster, 1997), Le guide d'évaluation du risque chez les délinquants sexuels (SORAG; Quinsey et coll., 2006), Statique-99 (Hanson et Thornton, 2000) et Statique-2002 (Hanson et Thornton, 2003). Les facteurs de risque de ces instruments se recoupent considérablement (caractéristiques démographiques, antécédents criminels) et le degré d'exactitude prédictive de chacun de ces instruments est semblable en ce qui a trait à leur capacité de distinguer parmi les délinquants sexuels les récidivistes des non récidivistes (Hanson et Morton-Bourgon, 2009).

Bien que les évaluateurs utilisent souvent plus d'un instrument d'évaluation (Jackson et Hess, 2007), ils ne savent pas avec certitude comment interpréter les résultats divergents. Malheureusement, les résultats divergents sont communes (p. ex. Mills et Kroner, 2006). Barbaree, Langton et Peacock (2006) ont constaté que moins de 8 % ($n = 20$) des délinquants sexuels échantillonnés ($N = 262$) ont été constamment déclarés délinquants à risque élevé ou à faible risque par cinq échelles actuarielles couramment utilisées (soit le Guide d'évaluation du risque de violence [VRAG; Quinsey et coll., 2006], le SORAG, la Statique-99, l'ERRRS et le MnSOST-R). Par conséquent, les évaluateurs qui veulent faire une évaluation actuarielle du risque chez les délinquants sexuels doivent décider de l'instrument d'évaluation à utiliser et s'ils en utilisent plus d'un, de quelle façon ils interpréteront les résultats divergents.

Le recours à des instruments d'évaluation multiples est une pratique courante dans de nombreux domaines de psychologie appliquée, comme dans le cas des évaluations neurologiques et cognitives (Brooks, Strauss, Sherman, Iverson et Slick, 2009; Malloy et coll., 1997). Par exemple, il s'est révélé que l'utilisation de plusieurs instruments (plutôt que d'un seul) augmente la précision des décisions au sujet de l'état de santé autodéclaré des patients cancéreux (Cella et coll., 1995), du rendement au travail (Avis, Kudisch et Fortunato, 2002) et des traitements contre le tabagisme (Sledjeski et coll., 2006). Lorsque plusieurs instruments d'évaluation sont utilisés, certains principes généraux de psychométrie éclairent leur utilisation et l'interprétation des résultats obtenus (Weiner, 2003). En général, les évaluateurs devraient privilégier les instruments d'évaluation 1) dont les items peuvent être codés de manière fiable (coefficient d'objectivité raisonnablement élevé), 2) qui comportent des données normatives pertinentes et 3) qui permettent des inférences statistiques valides (exactitude raisonnable des prévisions).

Lorsque ces critères généraux sont appliqués à l'évaluation du risque de récidive sexuelle, toutefois, aucun instrument n'est supérieur à un autre. En particulier, toutes les échelles actuarielles d'évaluation du risque de récidive sexuelle ont un coefficient d'objectivité similaire et acceptable (Barbaree, Seto, Langton et Peacock, 2001; G. T. Harris et coll., 2003) et l'exactitude prédictive diffère peu dans l'ensemble (Hanson et Morton-Bourgon, 2009; Rettenberger, Matthes, Boer et Eher, 2010).

Comme aucun instrument d'évaluation ne se démarque clairement, l'utilisation de plusieurs instruments est justifiée par la théorie psychométrique. Selon la théorie classique des tests, on peut réduire l'erreur expérimentale en augmentant le nombre d'items (« plus il y en a, mieux c'est »). En particulier, un score observé à un test (ou à un item) comporte deux éléments : la note vraie (ou selon la théorie des réponses aux items, l'aptitude ou le trait psychologique de la personne évaluée) et l'erreur de mesure (voir l'analyse de Rust et Golombok, 2009). À ce titre, augmenter le nombre d'items ou d'instruments d'évaluation devrait réduire l'erreur de mesure. Comme la théorie veut que l'erreur soit aléatoire, les

erreurs devraient s'annuler l'une l'autre au fil des observations (Nunnally et Bernstein, 1994). Par conséquent, ajouter des items aux instruments d'évaluation devrait se traduire par une exactitude accrue des prévisions. Certes, si les items ajoutés sont considérablement moins bons (moins prédictifs) que ceux déjà envisagés, l'exactitude de l'ensemble des prévisions sera amoindrie.

Validité ajoutée

Lorsque plusieurs échelles sont utilisées dans l'évaluation du risque, la validité ajoutée est une préoccupation fondamentale. Plus particulièrement, la validité ajoutée est la mesure dans laquelle les nouveaux éléments d'information améliorent l'exactitude d'une prévision par rapport à celle établie à l'aide d'un ou de plusieurs autres instruments d'évaluation. Sur le plan conceptuel, si un instrument fournit de nouveaux éléments d'information permettant de mieux comprendre le risque que représente un délinquant, il fournit de l'information ajoutée. Par exemple, de l'information ajoutée à propos du comportement antisocial aiderait à mieux comprendre le risque de récidive d'un délinquant que l'information fournie par un instrument d'évaluation particulier qui portant que sur les problèmes de santé mentale.

Il y a unanimité sur le fait que certains items ou certains types de facteurs de risque ajoutent à la validité prédictive des instruments d'évaluation du risque de violence ou de criminalité. De fait, la conception de la plupart des échelles actuarielles tient compte de la validité ajoutée des items retenus dans l'échelle finale (p. ex. Level of Service Inventory-Revised [LSI-R; Andrews et Bonta, 1995], Statique-99, Statique-2002 et VRAG). Avoir à notre disposition ces instruments d'évaluation reconnus, la question est de savoir dans quelle mesure les instruments finaux ont échantillonné et pondéré efficacement les variables pertinentes. C'est-à-dire, parmi les instruments actuellement utilisés et destinés aux évaluations globales du risque, dans quelle mesure est-il possible d'établir d'autres variables ou échelles qui ajoutent à la validité prédictive de ces instruments?

Les résultats des travaux de recherche sur la validité ajoutée d'instruments d'évaluation du risque couramment utilisés divergent. D'une part, Seto (2005) a constaté que certaines échelles couramment utilisées (c.-à-d. l'ERRRS, la Statique-99, le SORAG et le VRAG) n'ajoutaient pas l'une l'autre à leur validité prédictive sur la récidive sexuelle. Ces constatations donnent à penser que le recours à des instruments d'évaluation multiples ne vaut pas la peine. Cette étude, toutefois, était limitée de par la faible taille de l'échantillon de délinquants sexuels ($N = 215$). En outre, parmi les instruments qu'il a analysés, Seto (2005) a conclu que l'ERRRS prédisait le mieux la récidive sexuelle. Cependant, selon la plupart des études, l'ERRRS est moins bon que d'autres instruments d'évaluation existants, comme la Statique-99 (Hanson et Morton-Bourgon, 2009). Bref, il est difficile de suivre la recommandation de Seto (2005) selon laquelle il faut choisir le meilleur instrument, car d'une part, il n'existe aucun consensus scientifique à propos du « meilleur » instrument d'évaluation du risque de récidive sexuelle et, d'autre part, certains instruments peuvent mieux ou moins bien convenir à la prise de décision selon le pays ou échantillons.

Lloyd (2008) a examiné un grand nombre d'échelles actuarielles (MNSOST-R, Risk Matrix 2000 [Thornton et coll., 2003], ERRRS, SORAG, Statique-99), de directives cliniques sur le jugement professionnel structuré (Structured Risk Assessment - Need Assessment [SRA - Thornton, 2002], SVR-20) et d'autres variables hypothétiques sur le risque de récidive sexuelle (p. ex. nombre de victimes de sexe masculin) d'un groupe de délinquants sexuels ($N = 391$). Lloyd (2008) a constaté que l'utilisation combinée d'échelles d'évaluation du risque concourait à mieux prévoir le risque de récidive sexuelle et ajoutait à la validité prédictive réciproque de ces échelles (y compris le SORAG, le MNSOST-R, la SRA [score sur le plan socioaffectif] et le SVR-20). Bien qu'un certain surajustement soit possible à cause du grand nombre de variables incluses dans l'équation de régression, l'étude démontre que les échelles existantes peuvent ajouter l'une l'autre à leur validité prédictive quant à la récidive sexuelle.

Mills et Kroner (2006) ont poussé plus loin l'examen de la validité ajoutée en se penchant sur l'incidence des divergences entre les instruments d'évaluation du risque. Ils ont examiné la validité ajoutée de l'Échelle d'information statistique générale sur la récidive (ISGR; Nuffield, 1982), du LSI-R et du VRAG dans la prédiction du risque de récidive en général et de récidive avec violence des délinquants (environ

75 % de délinquants violents). De plus, ils ont séparé les délinquants selon que les résultats des instruments d'évaluation du risque étaient peu discordants (c.-à-d. que les différences normalisées moyennes entre les instruments étaient petites, ce qui semblait indiquer que les résultats de ces derniers étaient uniformes quant à l'évaluation du risque relatif) ou très discordants (c.-à-d. que les différences normalisées moyennes entre les instruments étaient grandes, ce qui semblait indiquer que les résultats de ces derniers n'étaient pas uniformes quant à l'évaluation du risque relatif). Mills et Kroner (2006) ont constaté que les échelles aux résultats peu discordants ajoutaient à la validité prédictive quant au risque de récidive en général et de récidive avec violence des délinquants ($n = 140$), mais pas les échelles aux résultats discordants ($n = 69$). Étant donné la faible taille de l'échantillon du groupe pour lequel les résultats des instruments sont discordants, une explication plausible de ces constatations contraires est le manque de puissance statistique nécessaire pour tester ces hypothèses.

Welsh, Schmidt, McKinnon, Chattha et Meyers (2008) ont examiné la validité ajoutée de l'Inventaire du niveau de service et de la gestion des cas pour les jeunes (YLS/CMI; Hoge et Andrews, 2002), de l'Évaluation structurée du risque de violence chez les jeunes (SAVRY; Borum, Bartel et Forth, 2002) et de l'Échelle de psychopathie de Hare (version révisée pour les jeunes) (PCL:YV; Forth, Kosson et Hare, 2003) sur un échantillon de délinquants juvéniles ($N = 105$) pour prévoir le risque de récidive en général et de récidive avec violence. Malgré la petite taille de l'échantillon, Welsh et coll. (2008) ont constaté que l'échelle SAVRY ajoutait à la validité prédictive des échelles PCL:YV et YLS/CMI sur le risque de récidive, tant avec violence qu'en général. Ils ont en outre constaté que l'échelle PCL:YV ajoutait à la validité prédictive de l'échelle YLS/CMI, mais que cette dernière n'ajoutait pas à la validité prédictive des deux autres échelles.

En somme, il existe peu d'études portant sur la validité ajoutée d'échelles d'évaluation non modifiées (c.-à-d. dont aucun item n'a été supprimé) sur le risque de criminalité et de violence et la plupart des études existantes sont limitées de par la faible taille de l'échantillon. Dans l'ensemble, la recherche laisse supposer que les instruments d'évaluation du risque pourraient ajouter l'un l'autre à leur validité prédictive s'ils sont utilisés concurremment. D'autres travaux de recherche sur des échantillons plus importants sont toutefois nécessaires pour mieux comprendre si le recours à plusieurs instruments d'évaluation a une utilité pratique.

Présente étude

La présente étude visait à comparer la validité prédictive de trois instruments d'évaluation couramment utilisés pour prévoir le risque de récidive des délinquants sexuels, soit l'ERRRS, la Statique-99R et la Statique-2002R. Plus particulièrement, elle examine 1) si chacun des instruments (ERRRS, Statique-99R et Statique-2002R) prédit mieux le risque de récidive sexuelle, de récidive avec violence et de récidive en général que les deux autres et 2) si les trois instruments ajoutent l'un l'autre à leur validité prédictive sur les trois types de risque de récidive. Les trois échelles à l'étude se ressemblent en ce sens qu'elles ont le même objectif (évaluer le risque de récidive sexuelle) et sont fondées sur des variables similaires quant aux caractéristiques démographiques et aux antécédents criminels. Si l'un des instruments était clairement supérieur aux deux autres sur le plan de l'exactitude prédictive et qu'aucune autre échelle n'ajoutait à sa validité prédictive, les évaluateurs seraient justifiés de n'utiliser que le « meilleur » instrument d'évaluation. Le choix des instruments serait moins clair, cependant, si l'exactitude prédictive d'aucun d'entre eux n'était supérieure à celle des autres ou s'il s'avérait qu'ils ajoutaient l'un l'autre à leur validité prédictive.

Méthodologie

Instruments d'évaluation

Évaluation rapide du risque de récidive sexuelle (ERRRS)

L'ERRRS (Hanson, 1997) est une échelle actuarielle conçue pour mesurer le risque de récidive sexuelle. Les scores vont de 0 à 6, les scores élevés signifiant un risque élevé de récidive sexuelle. Elle comporte quatre items : 1) infractions sexuelles antérieures, 2) au moins une victime sans lien de parenté, 3) au moins une victime de sexe masculin et 4) âge inférieur à 25 ans. Dans le cadre de la présente étude, les items de la Statique-99 ont été utilisés pour calculer l'ERRRS. Les règles de codage des items de l'ERRRS et de la Statique-99 sont identiques, exception faite des infractions sexuelles antérieures. En particulier, contrairement à celles de l'ERRRS, les règles de codage de la Statique-99 ne comptent pas la pseudorécidive parmi les infractions sexuelles antérieures. La pseudorécidive touche, selon les estimations, environ 5 % des délinquants (Phenix, Doren, Helmus, Hanson et Thornton, 2009) et, par conséquent, la différence entre la notation des items selon les règles de codage de la Statique-99 plutôt que celles de l'ERRRS devrait être minimale.

Dans l'étude d'élaboration, l'ERRRS a distingué les délinquants sexuels récidivistes de ceux qui ne le sont pas suivant une aire sous la courbe (ASC) de 0,71 (Hanson, 1997). Lors d'une méta-analyse récente, Hanson et Morton-Bourgon (2009) ont constaté que l'ERRRS a révélé des effets similaires, quoique de moindre importance, lorsqu'on a fait la moyenne de ses résultats pour 34 études de suivi diverses (moyenne pondérée $d = 0,60$; IC à 95 % de 0,54 à 0,65; $N = 11\ 031$; $k = 34$; ce qui se traduit par une ASC de 0,66, IC à 95 % de 0,65 à 0,68).

Statique-99R

L'échelle Statique-99R est une échelle actuarielle de 10 items qui évalue le risque de récidive des délinquants sexuels de sexe masculin adultes. Les items sont identiques à ceux de la Statique-99 (Hanson et Thornton, 2000), exception faite de la pondération révisée du facteur âge (voir Helmus, Thornton, Hanson et Babchishin, 2010). La Statique-99 est l'échelle actuarielle la plus couramment utilisée au Canada et aux États-Unis pour prévoir le risque de récidive sexuelle (Archer et coll., 2006; Jackson et Hess, 2007; McGrath, Cumming, Burchard, Zeoli et Ellerby, 2010). Elle est couramment utilisée pour la planification des traitements (McGrath et coll., 2010; Jackson et Hess, 2007), la surveillance dans la collectivité (Interstate Commission for Adult Offender Supervision, 2007) et l'évaluation aux fins de l'internement civil (Jackson et Hess, 2007).

La Statique-99R comporte tous les items de l'ERRRS ainsi que d'autres items portant sur les relations antérieures (un item), les infractions sexuelles antérieures (victimes qui sont des inconnus, infractions sexuelles sans contact) et les antécédents criminels en général (nombre de prononcés de peine antérieurs, infraction pour violence non sexuelle avant l'infraction répertoriée, infraction pour violence non sexuelle au moment de l'infraction répertoriée; voir tableau 1). Selon une méta-analyse récente, il existe une relation modérée entre la Statique-99 et la récidive sexuelle (moyenne pondérée $d = 0,67$; IC à 95 % de 0,62 à 0,72; $N = 20\ 010$; $k = 63$; ce qui se traduit par une ASC de la fonction d'efficacité du récepteur [FER] de 0,68; IC à 95 % de 0,67 à 0,70; Hanson et Morton-Bourgon, 2009). Pour un aperçu de la recherche sur la Statique-99, voir Anderson et Hanson (2010).

Tableau 1 – *Items dans l'ERRRS, la Statique-99R et la Statique-2002R.*

| Remarque | ERRRS | STATIQUE-99/STATIQUE-99R | STATIQUE-2002/STATIQUE-2002R |
|----------|--|--|---|
| a | Âge à la mise en liberté | Âge à la mise en liberté | Âge à la mise en liberté |
| b | Nombre d'accusations et de condamnations antérieures relatives à une infraction sexuelle | Nombre d'accusations et de condamnations antérieures relatives à une infraction sexuelle | Prononcés de peine antérieurs pour infractions sexuelles |
| c | Au moins une victime d'agression sexuelle sans lien de parenté | Au moins une victime d'agression sexuelle sans lien de parenté | Au moins une victime d'agression sexuelle sans lien de parenté |
| c | Au moins une victime d'agression sexuelle de sexe masculin | Au moins une victime d'agression sexuelle de sexe masculin | Au moins une victime d'agression sexuelle de sexe masculin |
| d | | Condamnations relatives à des infractions sexuelles sans contact | Condamnations relatives à des infractions sexuelles sans contact |
| d | | Au moins une victime d'agression sexuelle qui est un inconnu | Au moins une victime d'agression sexuelle qui est un inconnu |
| a | | Nombre de prononcés de peine antérieurs | Prononcés de peine antérieurs pour n'importe quelle infraction |
| e | | Infraction pour violence non sexuelle avant l'infraction répertoriée | Condamnations antérieures pour infractions pour violence non sexuelle |
| f | | Infraction pour violence non sexuelle au moment de l'infraction répertoriée | Démêlés antérieurs avec le système de justice pénale |
| f | | Cohabitation avec un conjoint pendant deux années consécutives | Jeunes victimes sans lien de parenté |
| f | | | Fréquence des infractions sexuelles |
| f | | | Violation des conditions de la surveillance dans la collectivité |
| f | | | Arrestations pour infraction sexuelle comme personne d'âge adulte et d'âge mineur |
| f | | | Nombre d'années sans infraction avant l'infraction répertoriée |

Remarques : Adapté d'A. J. R. Harris et Hanson (2010). Les échelles Statique-99 et Statique-2002 sont identiques à leur version « R » (révisée), exception faite des seuils de démarcation et des coefficients de pondération du facteur âge.

^a Même définition, mais seuils de démarcation et coefficients de pondération différents.

^b Les échelles Statique-99 et ERRRS définissent les infractions sexuelles antérieures de la même façon et leur accordent le même coefficient de pondération, mais la notation de la Statique-99 comprend le concept de « pseudorécidive », alors que ce n'est pas le cas de l'ERRRS. La Statique-2002 les définit autrement.

^c Items identiques dans les trois instruments d'évaluation.

^d Items identiques dans les échelles Statique-99 et Statique-2002.

^e Concepts similaires, mais définitions différentes.

^f Items différents (aucun équivalent dans l'autre échelle).

Statique-2002R

L'échelle Statique-2002 (Hanson et Thornton, 2003) avait pour objectif d'améliorer la Statique-99. La Statique-2002R est une échelle actuarielle de 14 items qui évalue le risque de récidive des délinquants sexuels de sexe masculin adultes. Les items sont identiques à ceux de la Statique-2002 (Hanson et Thornton, 2003), exception faite de la pondération révisée du facteur âge (voir Helmus et coll., 2010). La Statique-2002 diffère de la Statique-99 en ce qu'elle a ajouté des items, en a modifié certains autres, les a regroupés dans des sous-échelles significatives afin de faciliter l'interprétation des résultats, et qu'elle comporte davantage de règles de codage normalisées. Il existe une relation modérée entre la Statique-2002 et la récidive sexuelle (moyenne pondérée $d = 0,70$; IC à 95 % de 0,59 à 0,81; $N = 3\ 330$; $k = 8$; ce qui se traduit par une ASC de la FER de 0,69; IC à 95 % de 0,66 à 0,72; Hanson et Morton-Bourgon, 2009). Selon la recherche antérieure, la Statique-2002 prédit mieux que la Statique-99 la récidive sexuelle, la récidive avec violence ou la récidive en général (Hanson, Helmus et Thornton, 2010; Stalans, Hacker et Talbot, 2010).

Le tableau 1 présente la liste des items de l'ERRRS, de la Statique-99R et de la Statique-2002R. Pour plus d'information sur les échelles Statique-99R et Statique-2002R, consulter le site Web à l'adresse <http://www.static99.org> (en anglais seulement).

Échantillons

De multiples échantillons de divers pays ont été utilisés. Le tableau 2 présente les principales caractéristiques de chacun des échantillons ($k = 20$; $N = 7\ 491$). Les vingt échantillons avaient les scores de l'ERRRS et de la Statique-99R, mais seulement sept d'entre eux avaient les scores de la Statique-2002R. La plupart des échantillons provenaient du Canada ($k = 10$) ou des États-Unis ($k = 4$), et un échantillon provenait de chacun des pays suivants : Allemagne, Autriche, Danemark, Nouvelle-Zélande, Royaume-Uni et Suède. La présente étude a examiné trois types de récidive : sexuelle, avec violence (y compris la récidive sexuelle) et en général. Des vingt échantillons, quatre ne portaient que sur la récidive sexuelle, deux portaient sur la récidive sexuelle et la récidive avec violence et quatorze portaient sur tous les types de récidive.

Chaque série de données a été vérifiée pour relever les incohérences internes (p. ex. calcul erroné des scores totaux ou score des items contraire à d'autre d'information de la série de données). Les erreurs relevées ont été corrigées dans la mesure du possible, sinon, le cas était exclu de l'analyse. Les cas étaient également exclus s'il manquait de l'information sur le suivi, s'il manquait un item de la Statique-99R autre que « cohabitation avec un conjoint » (item 2), s'il manquait plus d'un item de la Statique-2002R, si le délinquant avait moins de 18 ans au moment de la mise en liberté ou moins de 16 ans au moment de l'infraction répertoriée ou s'il s'agissait d'une délinquante. Les critères d'exclusion liés à l'âge et au sexe sont précisés dans les règles de codage de la Statique-99 (A. J. R. Harris, Phenix, Hanson et Thornton, 2003) et de la Statique-2002 (Phenix et coll., 2009). L'item « âge » de la Statique-99R et de la Statique-2002R a été calculé à partir des séries de données pour chacun des échantillons.

Le nombre de participants dans ces échantillons était inférieur à ce qui a été établi antérieurement (p. ex. Helmus, 2009), car 1) la date de naissance ou l'âge du délinquant à la mise en liberté était nécessaire pour coder le nouveau facteur âge de la Statique-99R et de la Statique-2002R, et 2) la série de données devait comporter les scores totaux d'au moins deux des échelles étudiées (p. ex. les scores des items de la Statique-99 étaient nécessaires au calcul des scores totaux de l'ERRRS). On trouve une description détaillée des échantillons dans Helmus (2009) à l'adresse <http://www.static99.org> (en anglais seulement).

Présentation générale des analyses

Toutes les analyses ont été effectuées séparément par le premier et le troisième auteur par souci d'exactitude.

Exactitude prédictive

Dans la première série d'analyses, des méta-analyses à effets fixes et à effets aléatoires ont été utilisées pour calculer les aires sous la courbe de la fonction d'efficacité du récepteur (ASC de la FER) pondérées et les intervalles de confiance à 95 % pour chacun des instruments d'évaluation du risque. L'ASC est une mesure du risque relatif et peut être interprétée comme la probabilité qu'un récidiviste choisi de façon aléatoire obtienne un score supérieur sur l'échelle d'évaluation du risque qu'un non-récidiviste choisi de façon aléatoire. L'ASC est utile pour comparer les résultats de divers échantillons, car elle n'est pas influencée par les taux de récidive de base (Rice et Harris, 1995). Elle est toutefois influencée par la variance dans la distribution des scores utilisés pour prévoir la récidive (Hanson, 2008; Humphreys et Swets, 1991).

Les estimations à effets fixes des ASC et des erreurs types ont été calculées au moyen de la formule et des méthodes présentées dans Hedges (1994). Les analyses à effets fixes ont l'avantage de fournir une estimation de la variabilité inter-études (c.-à-d. la valeur Q de Cochran; Hedges et Olkin, 1985). Une valeur Q de Cochran significative indique que la variabilité des résultats entre les études est supérieure à ce à quoi on pourrait normalement s'attendre par l'effet du hasard (la statistique Q suit la loi du khi carré, à $k - 1$ degrés de liberté). La méta-analyse à effets aléatoires introduit un terme d'erreur pour tenir compte de la variabilité inter-études, ce qui donne des intervalles de confiance plus grands (et souvent plus réalistes) (Schmidt, Oh et Hayes, 2009). Par conséquent, les résultats des modèles à effets aléatoires et à effets fixes convergent à mesure que le degré de variabilité diminue (lorsque $Q <$ degrés de liberté, les deux approches donnent des résultats identiques). Les estimations du modèle à effets aléatoires ont été calculées au moyen des formules 10, 12 et 14 présentées dans Hedges et Vevea (1998).

Le test de corrélation entre les ASC de la FER de Hanley et McNeil (1983) a été utilisé pour vérifier si le degré d'exactitude prédictive des instruments d'évaluation du risque différait. Ce test nécessite ce qui suit : 1) l'ASC moyenne des deux instruments d'évaluation du risque comparés et 2) la corrélation moyenne entre les deux instruments comparés, calculée séparément en fonction des récidivistes et des non-récidivistes. Les ASC et les corrélations moyennes ont été calculées pour chacun des trois types de récidive (sexuelle, avec violence [y compris la récidive sexuelle] et en général). Hanley et McNeil (1983) ont proposé de recourir au coefficient de corrélation de Kendall (τ [tau]) plutôt qu'au coefficient de corrélation de Bravais-Pearson. Le coefficient τ est une mesure de corrélation de rang qui représente la relation entre le classement des données selon l'un et l'autre des deux instruments d'évaluation (soit pour les données ordinales). Le coefficient τ permet donc un test plus prudent que ne le ferait le coefficient de corrélation de Bravais-Pearson, lequel suppose des données d'intervalle. Le tableau 1 (dans Hanley et McNeil, 1983, p. 841) associe une corrélation globale basée sur l'ASC moyenne (des deux instruments d'évaluation du risque comparés) et le coefficient τ moyen (entre les instruments, selon qu'il s'agit des récidivistes ou des non-récidivistes). Cette nouvelle corrélation dérivée du tableau 1 (de Hanley et McNeil, 1983, p. 841) sera appelée la moyenne générale r . Les erreurs types des différences entre les deux ASC ($A_1 - A_2$) sont basées sur la formule 3 de Hanley et McNeil (1983) :

$$\hat{ET}_{A_1 - A_2} = \sqrt{ET_1^2 + ET_2^2 - 2rET_1ET_2}$$

où r est la moyenne générale et ET est l'erreur type respective de l'ASC de chacun des instruments d'évaluation. Si la valeur zéro était comprise dans l'intervalle de confiance à 95 % de la différence entre les deux instruments, cette différence ne serait pas statistiquement significative.

Tableau 2 – Description des échantillons.

| Étude | N | Période de libération | Suivi (ÉT) ^a | Taux de récidive | | | Âge (ÉT) | ERRRS M (ÉT) | Statique-99R M (ÉT) | Statique-2002R M (ÉT) |
|-------------------------------------|-------|-----------------------|-------------------------|------------------|----------------------------|------------|-------------|--------------|---------------------|-----------------------|
| | | | | Sexuelle | Avec violence ^b | En général | | | | |
| Allan et coll. (2007) | 492 | 1990-2000 | 5,7 (2,9) | 9,6 | 16,5 | 25,2 | 42,3 (12,2) | 1,4 (1,4) | 1,8 (2,3) | - |
| Bengtson (2008) | 308 | 1978-1995 | 16,2 (4,2) | 34,1 | 52,3 | 64,6 | 32,5 (10,4) | 1,8 (1,2) | 3,8 (2,4) | 4,6 (2,4) |
| Bigras (2007) | 457 | 1995-2004 | 4,6 (1,9) | 5,7 | 14,7 | 23,4 | 42,8 (12,0) | 1,3 (1,3) | 2,1 (2,4) | 3,5 (2,5) |
| Boer (2003) | 296 | 1976-1994 | 13,3 (2,1) | 8,8 | 23,3 | 48,3 | 41,2 (12,5) | 1,4 (1,2) | 2,8 (2,8) | 3,9 (2,7) |
| Bonta et Yessine (2005) | 133 | 1992-2004 | 5,5 (2,4) | 15,8 | 33,8 | 48,9 | 39,8 (9,6) | 2,7 (1,3) | 5,0 (2,1) | - |
| Brouillette-Alarie et Proulx (2008) | 228 | 1979-2006 | 9,9 (4,5) | 20,2 | 30,7 | - | 36,0 (10,2) | 2,1 (1,4) | 3,9 (2,3) | - |
| Cortoni et Nunes (2007) | 73 | 2001-2004 | 4,6 (0,6) | 0,0 | 8,2 | 12,3 | 41,6 (12,3) | 1,2 (1,0) | 2,2 (2,1) | - |
| Eher et coll. (2008) | 706 | 2000-2005 | 3,9 (1,1) | 4,0 | 14,7 | 26,2 | 40,7 (12,6) | 1,2 (1,0) | 2,3 (2,3) | - |
| Epperson (2003) | 177 | 1989-1998 | 7,9 (2,5) | 14,1 | - | - | 37,2 (13,2) | 1,5 (1,2) | 2,5 (2,6) | - |
| Haag (2005) | 190 | 1995 | 7,0 (0,0) | 24,7 | - | - | 36,7 (9,7) | 2,0 (1,4) | 4,1 (2,2) | 5,7 (2,3) |
| Hanson et coll. (2007) | 702 | 2001-2005 | 3,4 (1,0) | 8,1 | 16,4 | 27,9 | 41,6 (13,2) | 1,5 (1,2) | 2,4 (2,4) | 3,5 (2,5) |
| Harkins et Beech (2007) | 190 | 1994-1998 | 10,4 (1,1) | 14,2 | 21,1 | 36,3 | 43,3 (12,5) | 1,5 (1,3) | 2,2 (2,6) | 3,7 (2,8) |
| Hill et coll. (2008) | 86 | 1971-2002 | 12,6 (6,6) | 15,1 | 29,1 | 61,6 | 39,4 (11,1) | 1,9 (1,0) | 4,7 (2,0) | - |
| Johansen (2007) | 273 | 1994-2000 | 9,1 (1,1) | 7,7 | 20,5 | 53,5 | 37,8 (10,8) | 1,8 (1,2) | 2,9 (2,3) | - |
| Knight et Thornton (2007) | 466 | 1957-1986 | 8,6 (2,6) | 26,2 | 36,9 | 53,0 | 36,1 (11,4) | 2,4 (1,3) | 4,6 (2,4) | 6,1 (2,5) |
| Långström (2004) | 1 278 | 1993-1997 | 8,9 (1,4) | 7,5 | 21,4 | - | 41,5 (12,0) | 0,8 (0,9) | 2,0 (2,4) | - |
| Nicholaichuk (2001) | 281 | 1983-1998 | 6,4 (4,0) | 18,5 | - | - | 34,8 (9,4) | 2,4 (1,4) | 4,8 (2,4) | - |
| Swinburne Romine et coll. (2008) | 680 | 1977-2007 | 16,8 (7,8) | 13,8 | - | - | 38,2 (12,3) | 1,2 (1,1) | 1,7 (2,2) | - |
| Ternowski (2004) | 247 | 1994-1998 | 7,5 (1,0) | 8,1 | 15,4 | 19,8 | 43,9 (13,0) | 1,2 (1,2) | 1,6 (2,5) | - |
| Wilson et coll. (2007a et 2007b) | 228 | 1994-2007 | 5,2 (3,0) | 10,5 | 25,9 | 35,5 | 41,7 (11,4) | 2,8 (1,5) | 5,1 (2,3) | - |
| Total | 7 491 | 1957-2007 | 8,3 (5,2) | 12,0 | 22,4 | 35,9 | 39,8 (12,2) | 1,5 (1,3) | 2,7 (2,6) | 4,3 (2,7) |

Remarques : ^a Nombre d'années de suivi; ^b Récidive avec violence, y compris récidive sexuelle.

Validité ajoutée

La validité ajoutée a été examinée au moyen d'une analyse de régression de Cox (Allison, 1984). L'analyse de régression de Cox sert à estimer les rapports de risque (taux de risque) associés à une ou plusieurs variables prédictives à partir des données de survie étant donné un nombre inégal d'années de suivi. Chaque échantillon a été utilisé comme une strate pour permettre d'associer des fonctions de risque de base différentes (c.-à-d. taux de récurrence) à chaque valeur de la variable stratifiée, ce qui élimine de l'analyse la variabilité des taux de récurrence de base entre les échantillons.

Résultats

Validité prédictive

La validité prédictive des trois échelles a été mesurée au moyen des ASC. L'annexe A présente l'ASC de l'ERRRS, de la Statique-99R et de la Statique-2002R pour chaque échantillon. Les tableaux 3 à 5 présentent l'ASC pondérée pour chaque instrument d'évaluation du risque et les résultats du test de corrélation de Hanley et McNeil. Les échelles Statique-99R et Statique-2002R ont prédit le risque de récurrence sexuelle, de récurrence avec violence et de récurrence en général de façon similaire, aucune ne se démarquant sur le plan de l'exactitude prédictive (tableau 3). Vu que le coefficient τ permet un test plus prudent que ne le ferait le coefficient de corrélation de Bravais-Pearson, les analyses ont également comporté des calculs au moyen du coefficient de corrélation de Bravais-Pearson. Les résultats étaient similaires, sauf dans un cas : la Statique-2002R a significativement mieux évalué le risque de récurrence en général que la Statique-99R (écart avec le modèle à effets fixes = 0,0133, IC à 95 % de 0,00275 à 0,0238; écart avec le modèle à effets aléatoires = 0,0138, IC à 95 % de 0,00115 à 0,0265) au moyen du coefficient de corrélation de Bravais-Pearson, mais cet écart n'a pas été constaté avec l'utilisation du coefficient τ .

Le tableau 4 présente les résultats des méta-analyses de l'ASC de l'ERRRS et de la Statique-99R. Selon le test de corrélation de Hanley et McNeil, la Statique-99R a significativement mieux évalué le risque de récurrence sexuelle, de récurrence avec violence et de récurrence en général que l'ERRRS, particulièrement pour la récurrence avec violence (y compris la récurrence sexuelle) et la récurrence en général. Il en va de même entre l'ERRRS et la Statique-2002R, soit que la Statique-2002R a significativement mieux évalué le risque de récurrence sexuelle, de récurrence avec violence et de récurrence en général que l'ERRRS (voir tableau 5).

Les différences entre l'exactitude prédictive des échelles d'évaluation étaient similaires tant dans les analyses à effets fixes que celles à effets aléatoires. En outre, les différences entre l'exactitude prédictive des échelles étaient remarquablement uniformes d'un échantillon à l'autre sur le plan de la récurrence sexuelle et de la récurrence avec violence, comme l'indique un Q non significatif. En ce qui concerne la récurrence en général, une variation importante entre l'ERRRS et la Statique-99R ainsi qu'entre l'ERRRS et la Statique-2002R a été constatée, ce qui révèle que les différences de degré d'exactitude prédictive observées lors de ces comparaisons n'étaient pas uniformes d'un échantillon à l'autre.

Tableau 3 – Méta-analyse des ASC des échelles Statique-99R et Statique-2002R.

| Variable étudiée (récidive) | Échelle d'évaluation | Effets fixes | | | | Effets aléatoires | | | | k | N | Q |
|--|----------------------|------------------------|-----------|--------|------------------------|-------------------|--------|----------|--------|---|-------|----------|
| | | ASC de la FER pondérée | IC à 95 % | | ASC de la FER pondérée | IC à 95 % | | | | | | |
| | | | LI | LS | | LI | LS | | | | | |
| Sexuelle | Statique-99R | 0,684 | 0,655 | 0,713 | 0,699 | 0,641 | 0,757 | 0,641 | 0,757 | 7 | 2 609 | 19,40** |
| | Statique-2002R | 0,686 | 0,657 | 0,714 | 0,696 | 0,644 | 0,749 | 0,644 | 0,749 | 7 | 2 609 | 14,53* |
| Avec violence | Statique-99R | 0,703 | 0,679 | 0,727 | 0,705 | 0,658 | 0,752 | 0,658 | 0,752 | 6 | 2 419 | 16,05** |
| | Statique-2002R | 0,708 | 0,684 | 0,731 | 0,708 | 0,659 | 0,756 | 0,659 | 0,756 | 6 | 2 419 | 18,02** |
| En général | Statique-99R | 0,718 | 0,697 | 0,739 | 0,712 | 0,657 | 0,768 | 0,657 | 0,768 | 6 | 2 419 | 32,50*** |
| | Statique-2002R | 0,732 | 0,711 | 0,753 | 0,727 | 0,674 | 0,780 | 0,674 | 0,780 | 6 | 2 419 | 31,01*** |
| Différence entre les échelles Statique-99R et Statique-2002R | | | | | | | | | | | | |
| Sexuelle | | 0,00183 | -0,0183 | 0,0220 | 0,000787 | -0,0226 | 0,0242 | -0,0226 | 0,0242 | 7 | 2 609 | 2,15 |
| Avec violence | | 0,00341 | -0,0126 | 0,0194 | 0,00325 | -0,0146 | 0,0211 | -0,0146 | 0,0211 | 6 | 2 419 | 1,29 |
| En général | | 0,0132 | -0,000933 | 0,0274 | 0,0135 | -0,00224 | 0,0292 | -0,00224 | 0,0292 | 6 | 2 419 | 1,21 |

Remarques : IC = intervalle de confiance; LI = limite inférieure; LS = limite supérieure.
 * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$, suivant la loi du khi carré, à $k - 1$ degrés de liberté.

Tableau 4 – Méta-analyse des ASC des échelles ERRRS et Statique-99R.

| Variable étudiée (récidive) | Échelle d'évaluation | Effets fixes | | | | Effets aléatoires | | | | |
|---|----------------------|------------------------|-----------|--------|------------------------|-------------------|--------|----|-------|----------|
| | | ASC de la FER pondérée | IC à 95 % | | ASC de la FER pondérée | IC à 95 % | | k | N | Q |
| | | | LI | LS | | LI | LS | | | |
| Sexuelle | ERRRS | 0,661 | 0,642 | 0,680 | 0,660 | 0,628 | 0,691 | 19 | 7 418 | 28,16 |
| | Statique-99R | 0,694 | 0,675 | 0,713 | 0,697 | 0,664 | 0,730 | 19 | 7 418 | 35,71** |
| Avec violence | ERRRS | 0,614 | 0,597 | 0,631 | 0,605 | 0,574 | 0,636 | 16 | 6 163 | 29,76** |
| | Statique-99R | 0,725 | 0,710 | 0,740 | 0,707 | 0,675 | 0,739 | 16 | 6 163 | 48,85*** |
| En général | ERRRS | 0,582 | 0,564 | 0,600 | 0,576 | 0,547 | 0,605 | 14 | 4 657 | 19,79 |
| | Statique-99R | 0,709 | 0,693 | 0,724 | 0,700 | 0,665 | 0,735 | 14 | 4 657 | 48,71*** |
| Différence entre les échelles ERRRS et Statique-99R | | | | | | | | | | |
| Sexuelle | | 0,0304 | 0,0120 | 0,0489 | 0,0349 | 0,00847 | 0,0613 | 19 | 7 418 | 16,37 |
| | Avec violence | 0,104 | 0,0877 | 0,120 | 0,101 | 0,0752 | 0,127 | 16 | 6 163 | 20,69 |
| En général | | 0,123 | 0,106 | 0,139 | 0,124 | 0,0939 | 0,154 | 14 | 4 657 | 27,19** |

Remarques : IC = intervalle de confiance; LI = limite inférieure; LS = limite supérieure.
 * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$, suivant la loi du khi carré, à $k - 1$ degrés de liberté.

Tableau 5 – Méta-analyse des ASC des échelles ERRRS et Statique-2002R.

| Variable étudiée (récidive) | Échelle d'évaluation | Effets fixes | | | | Effets aléatoires | | | | |
|---|----------------------|------------------------|-----------|--------|---------------|------------------------|-----------|----|-------|----------|
| | | ASC de la FER pondérée | IC à 95 % | LI | LS | ASC de la FER pondérée | IC à 95 % | LI | LS | |
| | | | | | | | | | | k |
| Sexuelle | ERRRS | 0,650 | 0,621 | 0,680 | 0,655 | 0,609 | 0,702 | 7 | 2 609 | 8,76 |
| | Statique-2002R | 0,686 | 0,657 | 0,714 | 0,696 | 0,644 | 0,749 | 7 | 2 609 | 14,53* |
| Avec violence | ERRRS | 0,603 | 0,577 | 0,630 | 0,604 | 0,566 | 0,642 | 6 | 2 419 | 5,37 |
| | Statique-2002R | 0,708 | 0,684 | 0,731 | 0,708 | 0,659 | 0,756 | 6 | 2 419 | 18,02** |
| En général | ERRRS | 0,586 | 0,562 | 0,610 | 0,585 | 0,553 | 0,618 | 6 | 2 419 | 4,52 |
| | Statique-2002R | 0,732 | 0,711 | 0,753 | 0,727 | 0,674 | 0,780 | 6 | 2 419 | 31,01*** |
| Différence entre les échelles ERRRS et Statique-2002R | | | | | | | | | | |
| Sexuelle | | 0,0349 | 0,00650 | 0,0633 | 0,0370 | 0,00311 | 0,0710 | 7 | 2 609 | 2,40 |
| Avec violence | | 0,0985 | 0,0739 | 0,123 | 0,102 | 0,0613 | 0,142 | 6 | 2 419 | 9,41 |
| En général | | 0,139 | 0,117 | 0,161 | 0,139 | 0,0952 | 0,184 | 6 | 2 419 | 17,34** |

Remarques : IC = intervalle de confiance; LI = limite inférieure; LS = limite supérieure.
 * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$, suivant la loi du khi carré, à $k - 1$ degrés de liberté.

Validité ajoutée

Les tableaux 6 à 8 présentent les analyses de régression de Cox utilisées pour examiner la validité ajoutée des instruments d'évaluation pour chacun des types de récurrence. Dans le cas de la récurrence sexuelle, il s'est révélé que les instruments d'évaluation du risque ajoutaient l'un l'autre à leur validité prédictive en dépit d'importantes corrélations entre eux, de 0,70 à 0,92 (tableau 6). L'ERRRS et la Statique-99R ont ajouté l'une l'autre à leur validité prédictive, la Statique-99R et la Statique-2002R ont ajouté réciproquement à leur validité prédictive et, enfin, si la Statique-2002R a ajouté à la validité prédictive de l'ERRRS, cette dernière n'a toutefois pas ajouté à la validité prédictive de la Statique-2002R. De plus, un modèle comportant les trois instruments a révélé que les échelles Statique-99R et Statique-2002R ajoutaient à la validité prédictive du modèle, mais pas l'ERRRS, à savoir que l'ajout de l'ERRRS après la prise en compte des échelles Statique-99R et Statique-2002R n'a pas ajouté de façon significative à la validité prédictive du modèle (variation du $\chi^2 = 0,48$; $dl = 1$; $p = 0,49$).

Tableau 6 – Validité ajoutée de l'instrument d'évaluation quant au risque de récurrence sexuelle.

| | Récurrence sexuelle | | | | | Wald |
|----------------------|---------------------|-------|--------|-----------|------|-----------|
| | N | r | Exp(B) | IC à 95 % | | |
| | | | | LI | LS | |
| Comparaison 1 | | | | | | |
| ERRRS | 7 410 | 0,702 | 1,11 | 1,04 | 1,19 | 9,75** |
| Statique-99R | | | 1,26 | 1,21 | 1,31 | 143,15*** |
| Comparaison 2 | | | | | | |
| ERRRS | 2 606 | 0,703 | 1,06 | 0,96 | 1,17 | 1,27 |
| Statique-2002R | | | 1,23 | 1,17 | 1,30 | 55,17*** |
| Comparaison 3 | | | | | | |
| Statique-99R | 2 606 | 0,925 | 1,14 | 1,04 | 1,25 | 8,22** |
| Statique-2002R | | | 1,12 | 1,03 | 1,23 | 6,62* |
| Comparaison 4 | | | | | | |
| ERRRS | 2 606 | - | 1,04 | 0,94 | 1,15 | 0,48 |
| Statique-99R | | | 1,14 | 1,04 | 1,25 | 7,41** |
| Statique-2002R | | | 1,11 | 1,02 | 1,22 | 5,14* |

Remarques : Analyses effectuées séparément pour chacune des comparaisons, chaque échantillon étant utilisé comme une strate et les deux instruments étant introduits à l'étape 1. La taille des échantillons fluctue à cause du nombre de cas censurés avant la prochaine occurrence. r = corrélation entre les instruments d'évaluation; IC = intervalle de confiance; LI = limite inférieure; LS = limite supérieure.

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$.

Dans le cas de la récurrence avec violence (y compris la récurrence sexuelle), les trois instruments d'évaluation du risque ont fourni de l'information additionnelle dans toutes les analyses. Il est à noter toutefois que l'ERRRS avait un effet inverse sur le plan de la validité ajoutée, c'est-à-dire que les scores faibles à cet instrument étaient associés à un risque élevé de récurrence avec violence une fois neutralisées

les autres échelles (voir tableau 7). De plus, un modèle comportant les trois instruments a révélé une validité ajoutée significative pour chacun des instruments (des scores faibles à l'ERRRS prédisant des taux élevés de récidive avec violence).

Tableau 7 – Validité ajoutée de l'instrument d'évaluation quant au risque de récidive avec violence.

| | Récidive avec violence (y compris la récidive sexuelle) | | | | | Wald |
|----------------|---|-------|--------|-----------|------|-----------|
| | N | r | Exp(B) | IC à 95 % | | |
| | | | | LI | LS | |
| Comparaison 1 | | | | | | |
| ERRRS | 6 161 | 0,691 | 0,83 | 0,79 | 0,88 | 40,30*** |
| Statique-99R | | | 1,42 | 1,37 | 1,46 | 499,54*** |
| Comparaison 2 | | | | | | |
| ERRRS | 2 417 | 0,708 | 0,83 | 0,76 | 0,91 | 17,09*** |
| Statique-2002R | | | 1,34 | 1,28 | 1,40 | 165,81*** |
| Comparaison 3 | | | | | | |
| Statique-99R | 2 417 | 0,927 | 1,16 | 1,08 | 1,26 | 15,38*** |
| Statique-2002R | | | 1,10 | 1,02 | 1,18 | 6,33* |
| Comparaison 4 | | | | | | |
| ERRRS | 2 417 | - | 0,80 | 0,74 | 0,88 | 23,53*** |
| Statique-99R | | | 1,20 | 1,11 | 1,30 | 22,04*** |
| Statique-2002R | | | 1,16 | 1,07 | 1,25 | 13,88*** |

Remarques : Analyses effectuées séparément pour chacune des comparaisons, chaque échantillon étant utilisé comme une strate et les deux instruments étant introduits à l'étape 1. La taille des échantillons fluctue à cause du nombre de cas censurés avant la prochaine occurrence. *r* = corrélation entre les instruments d'évaluation; IC = intervalle de confiance; *LI* = limite inférieure; *LS* = limite supérieure.

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$.

Dans le cas de la récidive en général, toutes les comparaisons ont révélé que les instruments d'évaluation du risque ajoutaient l'un l'autre à leur validité prédictive (voir tableau 8). En particulier, l'ERRRS et la Statique-99R ont ajouté l'une l'autre à leur validité prédictive, la Statique-99R et la Statique-2002R ont ajouté réciproquement à leur validité prédictive et, enfin, l'ERRRS a ajouté à la validité prédictive de la Statique-2002R et inversement. De manière analogue à l'analyse sur la récidive avec violence, les scores élevés à l'ERRRS étaient associés à un faible risque de récidive en général. Enfin, un modèle comportant les trois instruments a révélé une validité ajoutée significative pour chacun des instruments (des scores faibles à l'ERRRS prédisant des taux élevés de récidive en général).

Tableau 8 – Validité ajoutée de l'instrument d'évaluation quant au risque de récidive en général.

| | Récidive en général | | | | | Wald |
|----------------|---------------------|-------|--------|-----------|------|-----------|
| | N | r | Exp(B) | IC à 95 % | | |
| | | | | LI | LS | |
| Comparaison 1 | | | | | | |
| ERRRS | 4 655 | 0,697 | 0,77 | 0,73 | 0,81 | 98,04*** |
| Statique-99R | | | 1,40 | 1,36 | 1,44 | 538,38*** |
| Comparaison 2 | | | | | | |
| ERRRS | 2 418 | 0,708 | 0,74 | 0,68 | 0,79 | 71,66*** |
| Statique-2002R | | | 1,40 | 1,35 | 1,46 | 337,57*** |
| Comparaison 3 | | | | | | |
| Statique-99R | 2 418 | 0,927 | 1,10 | 1,03 | 1,17 | 9,09** |
| Statique-2002R | | | 1,15 | 1,09 | 1,22 | 21,76*** |
| Comparaison 4 | | | | | | |
| ERRRS | 2 418 | - | 0,72 | 0,67 | 0,77 | 81,16*** |
| Statique-99R | | | 1,15 | 1,08 | 1,23 | 19,32*** |
| Statique-2002R | | | 1,25 | 1,17 | 1,33 | 48,36*** |

Remarques : Analyses effectuées séparément pour chacune des comparaisons, chaque échantillon étant utilisé comme une strate et les deux instruments étant introduits à l'étape 1. La taille des échantillons fluctue à cause du nombre de cas censurés avant la prochaine occurrence. *r* = corrélation entre les instruments d'évaluation; IC = intervalle de confiance; *LI* = limite inférieure; *LS* = limite supérieure.

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$.

Pour examiner l'importance pratique des constatations sur la validité ajoutée, les participants ont également été classés en catégories de risque de récidive (faible, moyen et élevé) selon une définition proposée par Babchishin et Hanson (2009) des catégories de risque nominales indépendante de l'échelle pour la Statique-99R et la Statique-2002R. Particulièrement, les délinquants dont le score lié au taux de récidive sexuelle est moins de la moitié de celui d'un délinquant type (taux de risque $< 0,50$) sont à « faible risque ». Les délinquants dont le score lié au taux de récidive sexuelle est plus de la moitié de celui d'un délinquant type, mais moins du double de celui d'un délinquant type (taux de risque 0,50-1,99) sont à « risque moyen ». Enfin, les délinquants dont le score lié au taux de récidive sexuelle est au moins le double de celui d'un délinquant type (taux de risque $> 2,00$) sont à « risque élevé » (voir Hanson, Lloyd, Helmus et Thornton, 2010, pour plus de détails sur le rapport de risque). Les rapports de risque, centrés sur les scores médians dans les échantillons courants (non choisis) de délinquants sexuels, ont été calculés dans une étude antérieure (Hanson et coll., 2010) au moyen des coefficients de régression (Cox), ce qui a donné les catégories suivantes pour la Statique-99R (faible : de -3 à -1; moyen : de 0 à 4; élevé : 5 et plus) et la Statique-2002R (faible : de -2 à 1; moyen : de 2 à 6; élevé : 7 et plus; voir l'annexe B).

Le tableau 9 est un tableau croisé simple des taux de récidive sexuelle évalués en fonction des catégories de risque des échelles Statique-99R et Statique-2002R. Il permet de voir les taux de récidive des délinquants auxquels les échelles d'évaluation accordent des résultats discordants (lorsque les deux échelles classent les délinquants dans des catégories de risque différentes). Les taux de récidive des

délinquants des groupes aux résultats discordants se situent entre les taux de récidive des délinquants des deux catégories de risque voisines. Par exemple, lorsque les deux échelles ont évalué que les délinquants représentaient un risque moyen, le taux de récidive observé était de 10,7 % (146/1 360) et lorsque les deux échelles ont évalué que les délinquants représentaient un risque élevé, le taux de récidive observé était de 34,4 % (174/506). Lorsqu'une échelle a évalué que les délinquants représentaient un risque moyen et l'autre, que les délinquants représentaient un risque élevé, le taux de récidive observé était de 21,9 % (73/334).

Tableau 9 – Répartition des catégories de risque des échelles Statique-99R et Statique-2002R et des taux de récidive sexuelle observés.

| | Statique-2002R | | | |
|---------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|
| | Faible | Moyen | Élevé | Total |
| | % ($n_{\text{récidiviste}}/n$) |
| Statique-99R | | | | |
| Faible | 2,9 % (7/244) | 0,0 % (0/5) | - | 2,8 % (7/249) |
| Moyen | 6,2 % (10/160) | 10,7 % (146/1 360) | 20,4 % (10/49) | 10,6 % (166/1 569) |
| Élevé | - | 22,1 % (63/285) | 34,4 % (174/506) | 30,0 % (237/791) |
| Total | 4,2 % (17/404) | 12,7 % (209/1 650) | 33,2 % (184/555) | $N = 2 609$ |

Remarques : Taux de récidive sexuelle établis à partir de tous les cas, sans neutraliser le nombre d'années de suivi. Nombre moyen d'années de suivi = 8,0 ($ÉT = 4,9$).

Analyse

La présente étude visait à examiner la validité prédictive et la validité ajoutée de trois échelles utilisées pour évaluer le risque de récidive des délinquants sexuels. Elle a révélé que la Statique-99R et la Statique-2002R évaluent mieux le risque de récidive sexuelle, de récidive avec violence et de récidive en général que l'ERRRS. Aucune différence sur le plan de l'exactitude prédictive n'a été observée entre la Statique-99R et la Statique-2002R. En dépit d'importantes corrélations entre les échelles d'évaluation, elles ont toutes trois ajouté l'une l'autre à leur validité prédictive, sauf dans un cas : l'ERRRS n'a pas ajouté à la validité prédictive de la Statique-2002R en ce qui concerne la récidive sexuelle. Fait intéressant, l'ERRRS avait un effet inverse sur le plan de la validité ajoutée en ce qui concerne la récidive avec violence et la récidive en général, c'est-à-dire que les scores élevés à cet instrument étaient associés à un risque de récidive diminué.

La validité ajoutée constatée dans le cadre de la présente étude est vraiment remarquable vu le chevauchement considérable des items de ces échelles d'évaluation. Par ailleurs, cette constatation est en nette contradiction avec celles de Seto (2005), selon lesquelles des échelles semblables n'ajoutaient pas l'une l'autre à leur validité prédictive (l'échantillon étant toutefois de bien plus petite taille). Il serait facile de supposer que les corrélations importantes entre les échelles d'évaluation du risque excluraient la possibilité d'une validité ajoutée. Vu le chevauchement considérable du contenu de ces échelles, Vrieze et Grove (2010) ont supposé que les résultats discordants entre les échelles seraient [TRADUCTION] « de prime abord une raison pour ne croire » à ni l'une ni l'autre des échelles et « réduiraient leur force en tant qu'outils d'acquisition de connaissances » (Vrieze et Grove, 2010, p. 388). Les constatations de la présente étude semblent indiquer que Vrieze et Grove (2010) n'ont que partiellement raison. Des instruments d'évaluation aussi valides l'un que l'autre peut donner des résultats divergents. Même lorsque les items se ressemblent, ils peuvent être liés au risque de récidive selon divers mécanismes causaux, un point abordé plus loin.

Une méta-analyse antérieure recouvrant sept des séries de données utilisées dans la présente étude a révélé que la Statique-2002 évaluait mieux le risque de récurrence sexuelle, de récurrence avec violence et de récurrence en général que la Statique-99 (Hanson et coll., 2010). Stalans et coll. (2010) ont aussi constaté que la Statique-2002 évaluait mieux le risque de récurrence sexuelle que la Statique-99. La présente étude n'a pas relevé de différence entre les versions révisées des échelles Statique-99 et Statique-2002 sur le plan de l'exactitude prédictive (malgré l'utilisation des mêmes échantillons que ceux employés par Hanson et coll., 2010), ce qui peut vraisemblablement être attribuable aux nouvelles pondérations du facteur âge dans les échelles révisées. Les pondérations révisées du facteur âge ont sensiblement accru l'exactitude prédictive de la Statique-99R, mais peu celle de la Statique-2002R (Helmus et coll., 2010). Ainsi l'exactitude prédictive de la Statique-99R et celle de la Statique-2002R se ressemblent davantage que celle des deux échelles d'évaluation d'origine. Les analyses statistiques de Hanson et coll. (2010) et celles de la présente étude comportaient également des différences. En particulier, Hanson et coll. (2010) ont recouru au coefficient de corrélation de Bravais-Pearson pour effectuer le test de corrélation de Hanley et McNeil (1983) (un test moins prudent que celui effectué avec le coefficient de corrélation de Kendall $[\tau]$), alors que dans la présente étude, c'est τ qui a été utilisé. Une nouvelle analyse des données de Hanson et coll. (2010) avec le coefficient de corrélation de Kendall n'a toutefois pas influé sur les constatations (c.-à-d. que la Statique-2002 a significativement mieux évalué le risque de récurrence que la Statique-99). Ainsi l'exactitude prédictive similaire des versions révisées de la Statique-99 et de la Statique-2002 est probablement attribuable aux pondérations révisées du facteur âge dans ces échelles plutôt qu'à la méthode utilisée pour examiner la différence entre leur exactitude prédictive.

Pondération des items

La validité ajoutée constatée dans la présente étude montre que la pondération initiale des items de l'ERRRS, de la Statique-99R et de la Statique-2002R n'était pas optimale. Fait étonnant, l'ERRRS s'est révélée ajouter à la validité prédictive de la Statique-99R quant au risque de récurrence sexuelle en dépit du fait que tous les items de l'ERRRS se retrouvent dans la Statique-99R. (En fait, nous avons utilisé les items de la Statique-99R pour calculer l'ERRRS.) La validité ajoutée constatée ne peut donc être attribuable à de nouveaux concepts représentés par l'ERRRS, mais à la pondération différente des items.

Selon nos constatations, il est clair que la pondération des items des échelles actuarielles ne sera probablement jamais optimale. En présence d'échantillons d'une assez bonne taille, l'hypothèse nulle (aucune validité ajoutée constatée) peut presque toujours être rejetée (Cohen, 1994). La mise au point des pondérations est toutefois une tâche interminable qui nécessite des échantillons de plus grande taille pour obtenir un degré de précision de plus en plus grand. Ceux qui conçoivent les tests doivent également faire attention à ne pas surajuster les données, car il est rare que de petits ajustements soient applicables à d'autres séries de données (Cureton, 1950). Par ailleurs, les pondérations complexes compliquent la notation et augmentent le risque d'erreur; l'emploi des nombres entiers est assez simple.

Bien que l'amélioration de la pondération des items puisse faire progresser l'évaluation du risque de récurrence, nous ne croyons pas qu'il s'agit là de la solution aux problèmes les plus pressants liés à l'évaluation du risque de récurrence dans des contextes appliqués. Nous sommes plutôt d'avis que pour faire avancer les choses, il faudrait porter davantage attention à la validité conceptuelle des instruments d'évaluation.

Validité conceptuelle et combinaison des résultats de plusieurs échelles d'évaluation du risque

La plupart des tests psychologiques sont conçus pour évaluer les traits psychologiques latents, comme la santé mentale et l'intelligence. Ainsi on s'attend à ce que les résultats de divers instruments d'évaluation dont la conception est la même concordent (p. ex. tests d'intelligence différents); par ailleurs, les évaluateurs font régulièrement la moyenne des résultats de plusieurs instruments d'évaluation (Weiner, 2003). Cette méthode consistant à effectuer la moyenne est fondée sur l'hypothèse de la théorie classique des tests selon laquelle un nombre d'items accru devrait réduire l'erreur d'échantillonnage et produire des

résultats plus fiables (Nunally et Bernstein, 1994). Les évaluateurs qui constatent que les résultats d'instruments d'évaluation différents concordent ont davantage confiance dans ces résultats.

Cependant, le choix des scores utilisés dans l'évaluation du risque de récidive avec violence a souvent été fondé uniquement sur des observations empiriques, sans vraiment tenir compte de la validité conceptuelle. Si les évaluateurs ne savent pas ce qui est évalué, ils peuvent difficilement savoir comment combiner les résultats d'instruments d'évaluation différents. La méthode privilégiée de combinaison variera selon que les échelles évaluent des concepts semblables ou différents.

Lorsque les échelles d'évaluation comportent des items des mêmes domaines et ont une relation similaire avec la variable étudiée (c.-à-d. la récidive), il est alors possible de baser des conclusions sur la moyenne des résultats de ces échelles. Par exemple, dans la présente étude, la Statique-99R et la Statique-2002R contribuent de manière semblable à l'évaluation du risque de récidive sexuelle et on peut supposer que ces échelles portent sur les mêmes traits psychologiques latents et leur accordent la même pondération. En dépit d'un assez faible effet de validité ajoutée entre la Statique-99R et la Statique-2002R, la différence entre les taux de récidive des groupes de délinquants aux résultats discordants était notable, à savoir que lorsque les résultats de ces deux échelles étaient discordants, une différence d'environ 10 % était observée dans les taux de récidive, ceux des groupes de délinquants aux résultats discordants se situant entre ceux des délinquants des deux catégories de risque voisines. Un écart de 10 % se compare aux effets constatés de la plupart des facteurs de risque reconnus (p. ex. au moins une victime de sexe masculin, célibataire, au moins une victime sans lien de parenté; Hanson et Bussière, 1998).

Lorsque les échelles d'évaluation comportent des items de domaines différents, la façon de combiner leurs résultats en un seul jugement cohérent n'est pas aussi évidente. Si les échelles évaluent des concepts différents, il n'est pas surprenant qu'elles classent les délinquants différemment. Il n'est peut-être pas conseillé de faire la moyenne des résultats des deux échelles distinctes, car cela pourrait se traduire par une exactitude prédictive moindre comparativement à d'autres méthodes de combinaison des résultats. Par exemple, l'ERRRS accorde plus de poids à la déviance sexuelle que la Statique-99 (Doren, 2004; Roberts, Doren et Thornton, 2002), qui inclut des items portant sur la déviance sexuelle ainsi que sur le comportement antisocial. La méthode de combinaison des résultats d'échelles évaluant des concepts différents doit donc aussi prendre en considération 1) le ou les domaines évalués par les échelles et 2) le lien entre chaque domaine et la variable étudiée (soit la récidive). Dans la présente étude, l'ERRRS a ajouté à la validité de la Statique-99R, mais dans des sens opposés selon le type de récidive (c.-à-d. validité ajoutée dans le sens habituel dans le cas de la récidive sexuelle, mais validité ajoutée en sens inverse dans le cas de la récidive avec violence et de la récidive en général). La relation inverse de l'ERRRS dans le cas de la récidive avec violence et de la récidive en général semble indiquer que de soustraire les résultats de l'ERRRS de ceux de la Statique-99R serait une meilleure méthode de combinaison que de faire la moyenne. Cependant, dans le cas de la récidive sexuelle, où les deux échelles ajoutent réciproquement à leur validité prédictive, il est possible qu'une approche selon laquelle les résultats des échelles sont additionnés ou font l'objet d'une moyenne serait plus précise.

En somme, pour déterminer la méthode à utiliser pour combiner les résultats d'échelles d'évaluation du risque portant sur différents domaines, il faut préciser ce que les échelles évaluent vraiment. Cette tâche ne va toutefois pas de soi. Bien que tous les items de l'ERRRS se retrouvent dans la Statique-99R, les deux échelles d'évaluation avaient des relations opposées avec la récidive avec violence, une fois neutralisés les autres instruments. Par conséquent, on peut supposer (bien qu'à posteriori) que les deux échelles d'évaluation portent sur différents domaines. Cerner les concepts évalués nécessite des données tant théoriques qu'empiriques, sinon, on peut s'attendre à ce que la fiabilité des résultats sur les traits psychologiques latents soit faible entre les différents évaluateurs.

Implications pour les chercheurs

Nous sommes d'avis que selon les résultats de la présente étude, il serait justifié de prendre davantage en compte la validité conceptuelle lors de l'élaboration des instruments empiriques d'évaluation du risque. S'il est possible de résoudre le problème que représente l'utilisation combinée de plusieurs instruments

d'évaluation sans comprendre ce qu'ils évaluent, recourir à une approche pure de prédiction pour y parvenir comporte des limites considérables. Vrieze et Grove (2010), par exemple, ont proposé de créer une super-échelle, dont les items seraient les échelles existantes. Bien qu'une telle approche soit logiquement cohérente, elle est inefficace et peu pratique. En particulier, cette super-échelle nécessiterait de passer par toutes les mêmes étapes que lors de la création de toute nouvelle échelle d'évaluation, comme produire un manuel de notation et effectuer une contrevalidation. Étant donné que les items des échelles d'évaluation sont identiques ou presque, les évaluateurs se lasseraient vite de la répétition et essaieraient rapidement de trouver des façons de combiner les items plutôt que les scores totaux des diverses échelles.

Nous croyons que les futurs travaux de recherche sur l'évaluation du risque devraient préciser et évaluer les caractéristiques psychologiques significatives associées à la récidive (Mann, Hanson et Thornton, 2010). Par exemple, une seule dimension ou propension (p. ex. comportement antisocial) serait composée de plusieurs marqueurs et influencée par ces derniers (p. ex. chômage, toxicomanie, antécédents criminels, attitudes favorables au crime). Une fois évaluées les mesures valides des principaux concepts, les chercheurs peuvent examiner la contribution distincte de chacune de ces dimensions. Selon la théorie sur les dimensions (Loftus, Oberg et Dillon, 2004), les facteurs de risque pourraient être pondérés au niveau conceptuel (p. ex. comportement antisocial, déviance sexuelle) et le poids accordé à chaque concept peut dépendre du type de récidive évalué (p. ex. récidive avec violence comparativement à récidive sexuelle).

Une telle échelle actuarielle conceptuelle aurait l'avantage de comporter des sous-éléments définis et, par conséquent, de permettre aux évaluateurs de préciser les raisons expliquant le score d'un délinquant. En comprenant ce que les items mesurent, les évaluateurs pourraient expliquer les incohérences entre les instruments en ce qui concerne l'évaluation du risque, ce qui servirait de base à la méthode consistant à utiliser une combinaison de plusieurs échelles d'évaluation du risque. Ce ne serait toutefois pas une mince tâche, car elle exigerait non seulement de comprendre les concepts sous-jacents, mais également de savoir comment les items particuliers mesurent ces concepts. Néanmoins, cette tâche est essentielle, étant donné que la validité ajoutée des échelles ne se limite vraisemblablement pas aux trois échelles actuarielles examinées dans la présente étude. Si les échelles d'évaluation du risque sont créées à partir d'une approche essentiellement prédictive, les évaluateurs continueront de se retrouver dans la situation où ils sauront que d'autres variables (et d'autres échelles) ajoutent à la validité sans pouvoir expliquer pourquoi. Pour faire avancer la question de l'évaluation du risque, il faut conjuguer prédiction empirique et validité conceptuelle traditionnelle.

Implications pour la pratique

Selon la présente étude, ni la Statique-99R ni la Statique-2002R ne s'est révélée mieux évaluer le risque de récidive sexuelle, de récidive avec violence et de récidive en général (mais les deux ont surpassé l'ERRRS à cet égard). Par conséquent, les évaluateurs devant choisir l'une ou l'autre devraient prendre d'autres critères en considération. Par exemple, s'ils souhaitent évaluer le taux de récidive absolu, ils pourraient préférer la Statique-99R à la Statique-2002R en raison de la taille assez importante des échantillons normatifs disponibles (Helmus, 2009). Pour d'autres types d'évaluations, il serait préférable de recourir à la Statique-2002R plutôt qu'à la Statique-99R, car les items de la première sont regroupés en sous-échelles (c.-à-d. âge, déviance sexuelle, criminalité en général) qui indiquent la source du risque. Dans le cas d'évaluations du risque où l'enjeu est considérable, les évaluateurs pourraient vouloir utiliser les deux échelles : elles ajoutent réciproquement à leur validité prédictive, les taux de récidive des groupes de délinquants aux résultats discordants se situant entre ceux indiqués par chacune des échelles.

Pour évaluer le risque de récidive avec violence, les échelles Statique-99R et Statique-2002R peuvent toutes deux être utilisées. Les évaluateurs devraient toutefois être conscients que la pondération des items de ces échelles n'est pas optimale pour prédire la récidive avec violence (c.-à-d. poids trop important attribué aux items évaluant la déviance sexuelle). En conséquence, si l'évaluation porte principalement

sur la récidive avec violence, nous recommandons de recourir à des instruments conçus à cet effet (p. ex. VRAG, SORAG [Quinsey et coll., 2006]; Risk Matrix 2000v et Risk Matrix 2000c [Thornton et coll., 2003]). Ces instruments attribuent un poids plus important à la criminalité en général que l'ERRRS, la Statique-99R et la Statique-2002R.

Dans cette étude, nous avons présenté les poids de prédiction (coefficients de régression normalisés des analyses de régression de Cox) de l'ERRRS, de la Statique-99R et de la Statique-2002R à des fins d'illustration seulement. Nous ne recommandons pas l'utilisation de ces coefficients dans un contexte appliqué parce qu'ils seraient probablement sujet à un surajustement (Cureton, 1950). Sans de nouvelles études sur la question (avec des échantillons de taille importante), il n'est pas possible de savoir la mesure dans laquelle les coefficients calculés sont exacts et si on peut généraliser leur utilisation.

En résumé, pour les évaluateurs qui choisissent des échelles évaluant des domaines similaires de facteurs de risque (p. ex. Statique-99R et Statique-2002R), il semble que faire la moyenne des résultats serait l'approche optimale de combiner les résultats des échelles. Une telle approche s'inscrit dans la théorie classique des tests en ce sens qu'un nombre accru d'items évaluant un même ensemble de concepts (et ayant une exactitude prédictive semblable) réduit l'erreur de mesure et augmente l'exactitude prédictive. Par conséquent, si les résultats des différentes échelles concordent, les évaluateurs auront davantage confiance dans l'exactitude de l'évaluation du risque. En revanche, si les échelles ne portent pas sur les mêmes traits psychologiques latents, il faudrait aux évaluateurs un modèle défendable en ce qui concerne 1) les traits psychologiques latents évalués par les échelles, 2) la façon dont les domaines de facteurs de risque sont liés à la variable étudiée et 3) les données empiriques décrivant la manière dont les concepts devraient être pondérés et combinés. En l'absence d'un tel modèle fondé sur des données empiriques, les évaluateurs devraient faire preuve de prudence et privilégier l'échelle dans laquelle ils ont le plus confiance.

Bibliographie

Les notices bibliographiques précédées d'un astérisque ont fait l'objet de la méta-analyse.

- *ALLAN, M., R. C. GRACE, B. RUTHERFORD et S. M. HUDSON. 2007. « Psychometric assessment of dynamic risk factors for child molesters », *Sexual Abuse: A Journal of Research and Treatment*, vol. 19, p. 347-367. DOI : 10.1007/s11194-007-9052-5.
- ALLISON, P. D. 1984. *Event history analysis: Regression for longitudinal event data*, Beverly Hills (Calif.), Sage.
- ANDERSON, D., et R. K. HANSON. 2010. « Static-99: An actuarial tool to assess risk of sexual and violent recidivism among sexual offenders », dans R. K. Otto et K. S. Douglas, *Handbook of Violence Risk Assessment*, New York, Taylor & Francis Group, p. 251-267.
- ANDREWS, D. A., et J. BONTA. 1995. *The Level of Service Inventory – Revised*, Toronto (Ont.), Multi-Health Systems.
- ARCHER, R. P., J. K. BUFFINGTON-VOLLUM, R. V. STREDNY et R. W. HANDEL. 2006. « A survey of psychological test use patterns among forensic psychologists », *Journal of Personality Assessment*, vol. 87, p. 84-94. DOI : 10.1207/s15327752jpa8701_07.
- AVIS, J. M., J. D. KUDISCH et V. J. FORTUNATO. 2002. « Examining the incremental validity and adverse impact of cognitive ability and conscientiousness on job performance », *Journal of Business and Psychology*, vol. 17, p. 87-105. DOI : 0889-3268/02/0900-0087/0.
- BABCHISHIN, K. M., et R. K. HANSON. 2009. « Improving our talk: Moving beyond ‘low’, ‘moderate’, and ‘high’ in risk communication », *Crime Scene*, vol. 16, n° 1, p. 11-14. Adresse : <http://www.cjsw.ac.uk/cjsw/files/Hanson%202009.pdf>.
- BARBAREE, H. E., C. M. LANGTON et E. J. PEACOCK. 2006. « Different actuarial risk measures produce different risk rankings for sexual offenders », *Sexual Abuse: A Journal of Research and Treatment*, vol. 18, p. 423-440. DOI : 10.1007/s11194-006-9029-9.
- BARBAREE, H. E., M. C. SETO, C. M. LANGTON et E. J. PEACOCK. 2001. « Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders », *Criminal Justice and Behavior*, vol. 28, p. 490-521. DOI : 10.1177/009385480102800406.
- *BENGTSON, S. 2008. « Is newer better? A cross-validation of the Static-2002 and the Risk Matrix 2000 in a Danish sample of sexual offenders », *Psychology, Crime & Law*, vol. 14, p. 85-106. DOI : 10.1080/10683160701483104.
- *BIGRAS, J. 2007. « La prédiction de la récidive chez les délinquants sexuels [Prediction of recidivism among sex offenders] », *Dissertation Abstracts International : Section B*, vol. 68, n° 9. N° du UMI : NR30941.
- *BOER, A. 2003. *Evaluating the Static-99 and Static-2002 risk scales using Canadian sexual offenders*, University of Leicester, United Kingdom. Thèse de maîtrise inédite.
- BOER, D. P., S. D. HART, P. R. KROPP et C. D. WEBSTER. 1997. *Manual for the Sexual Violence Risk - 20 : Professional guidelines for assessing risk of sexual violence*, Vancouver (B.C.), British Columbia Institute on Family Violence and Mental Health, Law and Policy Institute, Simon Fraser University.
- *BONTA, J., et A. K. YESSINE. 2005. *Système national de repérage : identification des délinquants violents à risque élevé et intervention*, Rapport pour spécialistes n° 2005-04, Ottawa (Ont.), Sécurité publique et Protection civile Canada. Données brutes inédites sur la récidive des 124 délinquants sexuels parmi les délinquants identifiés dans le rapport.

- BORUM, R., P. BARTEL et A. FORTH. 2002. *Manual for the Structured Assessment of Violence Risk in Youth (SAVRY)*, Tampa (Fl.), University of South Florida.
- BROOKS, B. L., E. STRAUSS, E. M. S. SHERMAN, G. L. IVERSON et D. J. SLICK. 2009. « Developments in neuropsychological assessments: Refining psychometric and clinical interpretive methods », *Canadian Psychology = Psychologie canadienne*, vol. 50, p. 196-209. DOI : 10.1037/a0016066.
- *BROUILLETTE-ALARIE, S., et J. PROULX. 2008. *Predictive and convergent validity of phallometric assessment in relation to sexual recidivism risk*. Affiche présentée à la conférence annuelle de l'Association for the Treatment of Sexual Abusers, Atlanta (Ga.).
- CELLA, D. F., A. E. BONOMI, S. R. LLOYD, D. S. TULSKY, E. KAPLAN et P. BONOMI. 1995. « Reliability and validity of the Functional of Cancer Therapy – Lung (FACT-L) quality of life instrument », *Lung Cancer*, vol. 12, p. 199-220.
- COHEN, J. 1994. « The earth is round ($p < .05$) », *The American Psychologist*, vol. 49, p. 997-1003.
- *CORTONI, F., et K. L. NUNES. 2007. *Évaluation de l'efficacité du Programme national pour délinquants sexuels*, Ottawa (Ont.), Service correctionnel du Canada. Adresse : <http://www.csc-scc.gc.ca/text/rsrch/reports/r183/r183-fra.shtml>. Rapport de recherche n° R-183.
- CRONBACH, L. J., et P. E. MEEHL. 1955. « Construct validity in psychological tests », *Psychological Bulletin*, vol. 52, p. 281-302.
- CURETON, E. E. 1950. « Validity, reliability, and baloney », *Educational and Psychological Measurement*, vol. 10, p. 94-96.
- Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579. 1993.
- DAWES, R. M., D. FAUST et P. E. MEEHL. 1989. « Clinical versus actuarial judgment », *Science*, vol. 243, p. 1668-1674. DOI : 10.1126/science.2648573.
- DOREN, D. 2004. « Toward a multidimensional model for sexual recidivism risk », *Journal of Interpersonal Violence*, vol. 19, p. 835-856. DOI : 10.1177/0886260504266882.
- DOUGLAS, K., et P. R. KROPP. 2002. « A prevention-based paradigm for violence risk assessment: Clinical and research applications », *Criminal Justice and Behavior*, vol. 29, p. 617-658. DOI : 10.1177/009385402236735.
- *EHER, R., M. RETTENBERGER, F. SCHILLING et F. PFAFFLIN. 2009. Données brutes inédites sur les délinquants sexuels mis en liberté en Autriche.
- *EPPERSON, D. L. 2003. *Validation of the MnSOST-R, Static-99, and RRASOR with North Dakota prison and probation samples*, North Dakota Division of Parole and Probation. Rapport d'assistance technique inédit.
- EPPERSON, D. L., J. D. KAUL, S. J. HUOT, D. HESSELTON, R. GOLDMAN et W. ALEXANDER. 1998. *Minnesota Sex Offender Screening Tool-Revised (MnSOST-R)*, St. Paul (Minn.), Minnesota Department of Corrections.
- FORTH, A. E., D. S. KOSSON et R. D. HARE. 2003. *The Hare Psychopathy Checklist: Youth Version*, North Tonawanda (N.Y.), Multi-Health Systems.
- GARB, H. N. 2003. « Clinical judgment and mechanical prediction », dans J. R. Graham, J. A. Naglieri et I. B. Weiner (éd.), *Handbook of psychology: Volume 10, Assessment psychology*, Hoboken (N.J.), John Wiley & Sons, Inc., p. 27-42.
- GENDREAU, P., C. GOGGIN et M. LAW. 1997. « Predicting prison misconduct », *Criminal Justice and Behavior*, vol. 24, p. 414-431. DOI : 10.1177/0093854897024004002.

- GROVE, W. M., D. H. ZALD, B. S. LEBOW, B. E. SNITZ et C. NELSON. 2000. « Clinical versus mechanical prediction: A meta-analysis », *Psychological Assessment*, vol. 12, p. 19-30. DOI : 10.1037/1040-3590.12.1.19.
- *HAAG, A. M. 2005. « Do psychological interventions impact on actuarial measures: An analysis of the predictive validity of the Static-99 and Static-2002 on a re-conviction measure of sexual recidivism », *Dissertation Abstracts International : Section B*, vol. 66, n° 8, p. 4531. Données brutes inédites sur la récidive de 198 délinquants incarcérés jusqu'à la date d'expiration du mandat.
- HANLEY, J. A., et B. J. MCNEIL. 1983. « A method of comparing the Areas under ROC curves derived from same cases », *Radiology*, vol. 148, p. 839-843.
- HANSON, R. K. 1997. *Établissement d'une échelle actuarielle sommaire du risque de récidive sexuelle*, Rapport pour spécialistes n° 1997-04, Ottawa (Ont.), Ministère du Solliciteur général du Canada. Adresse : <http://www.publicsafety.gc.ca/res/cor/rep/cprminindex-eng.aspx#a97>.
- HANSON, R. K. 2005. « Twenty years of progress in violence risk assessment », *Journal of Interpersonal Violence*, vol. 20, p. 212-217. DOI : 10.1177/0886260504267740.
- HANSON, R. K. 2008. « What statistics should we use to report predictive accuracy », *Crime Scene*, vol. 15, n° 1, p. 15-17. Adresse : <http://www.cpa.ca/cpsite/userfiles/Documents/Criminal%20Justice/Crime%20Scene%202008-04.pdf>.
- HANSON, R. K. 2009. « The psychological assessment of risk for crime and violence », *Canadian Psychology = Psychologie canadienne*, vol. 20, p. 172-182. DOI : 10.1037/a0015726.
- HANSON, R. K., et M. T. BUSSIÈRE. 1998. « Predicting relapse: A meta-analysis of sexual offender recidivism studies », *Journal of Consulting and Clinical Psychology*, vol. 66, p. 348-362.
- *HANSON, R. K., A. J. R. HARRIS, T.-L. SCOTT et L. HELMUS. 2007. *Évaluation du risque chez les délinquants sexuels soumis à une surveillance dans la collectivité : le Projet de surveillance dynamique*, Rapport pour spécialistes n° 2007-05, Ottawa (Ont.), Sécurité publique Canada. Adresse : http://www.publicsafety.gc.ca/res/cor/rep/_fl/crp2007-05-fr.pdf.
- HANSON, R. K., L. HELMUS et D. THORNTON. 2010. « Predicting recidivism among sexual offenders: A multi-site study of Static-2002 », *Law and Human Behavior*, vol. 34, p. 198-211. DOI : 10.1007/s10979-009-9180-1.
- HANSON, R. K., C. D. LLOYD, L. HELMUS et D. THORNTON. 2010. *Using multiple samples to estimate relative risk for actuarial risk tools: A Canadian example using Static-99 and Static-2002*. Manuscrit inédit.
- HANSON, R. K., et K. E. MORTON-BOURGON. 2009. « The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies », *Psychological Assessment*, vol. 21, p. 1-21. DOI : 10.1037/a0014421.
- HANSON, R. K., et D. THORNTON. 2000. « Improving risk assessments for sex offenders: A comparison of three actuarial scales », *Law and Human Behavior*, vol. 24, p. 119-136. DOI : 10.1023/A:1005482921333.
- HANSON, R. K., et D. THORNTON. 2003. *Notes sur l'élaboration de la Statique-2002*, Rapport pour spécialistes n° 2003-01, Ottawa (Ont.), Ministère du Solliciteur général du Canada. Adresse : <http://www.securitepublique.gc.ca/res/cor/rep/2003-01-not-sttc-fra.aspx>.
- *HARKINS, L., et A.R. BEECH. 2007. *Examining the effectiveness of sexual offender treatment using risk band analysis*. Manuscrit inédit.

- HARRIS, A. J. R., et R. K. HANSON. 2010. « Clinical, Actuarial, and Dynamic risk assessment of sexual offenders: Why do things keep changing? », *Journal of Sexual Aggression*, vol. 16, p. 296-310.
- HARRIS, A. J. R., A. PHENIX, R. K. HANSON et D. THORNTON. 2003. *Statique-99 : règles de codage révisées – 2003*, Ottawa (Ont.), Ministère du Solliciteur général du Canada. Adresse : http://www.publicsafety.gc.ca/res/cor/rep/_fl/2003-03-stc-cde-fra.pdf.
- HARRIS, G. T., M. E. RICE, V. L. QUINSEY, M. L. LALUMIÈRE, D. BOER et C. LANG. 2003. « A multi-site comparison of actuarial risk instruments for sex offenders », *Psychological Assessment*, vol. 15, p. 413-425. DOI : 10.1037/1040-3590.15.3.413.
- HART, S. D., D. N. COX et R. D. HARE. 1995. *Manual for the Hare Psychopathy Checklist -Revised: Screening Version (PCL:SV)*, Toronto (Ont.), Multi-Health Systems.
- HEDGES, L. V. 1994. « Fixed effect models », dans H. Cooper et L. V. Hedges (éd.), *The handbook of research synthesis*, New York, Russell Sage, p. 285-299.
- HEDGES, L. V., et I. OLKIN. 1985. *Statistical methods for meta-analysis*, New York, Academic Press.
- HEDGES, L. V., et J. L. VEVEA. 1998. « Fixed- and random-effects models in meta-analysis », *Psychological Methods*, vol. 3, p. 486-504. DOI : 10.1037/1082-989X.3.4.486.
- HELMUS, L. 2009. *Re-norming Static-99 recidivism estimates: Exploring base rate variability across sex offender samples*. Thèse de maîtrise. N° du UMI : MR58443 (base de données *Dissertations and Theses* de ProQuest). Disponible également à l'adresse www.static99.org.
- HELMUS, L., D. THORNTON, R. K. HANSON et K. M. BABCHISHIN. 2011. *Évaluation du risque chez les délinquants sexuels âgés au moyen des échelles Statique-99R et Statique-2002R*, Rapport pour spécialistes n° 2011-01, Ottawa (Ont.), Sécurité publique Canada.
- *HILL, A., N. HABERMANN, D. KLUSMANN, W. BERNER et P. BRIKEN. 2008. « Criminal recidivism in sexual homicide perpetrators », *International Journal of Offender Therapy and Comparative Criminology*, vol. 52, p. 5-20. DOI : 10.1177/0306624X07307450.
- HOGUE, R. D., et D. A. ANDREWS. 2002. *The Youth Level of Service/Case Management Inventory manual and scoring key*, Toronto (Ont.), Multi-Health Systems.
- HUMPHREYS, L. G., et J. A. SWETS. 1991. « Comparison of predictive validities measured with biserial correlations and ROCs of signal detection theory », *Journal of Applied Psychology*, vol. 76, p. 316-321. DOI : 10.1037/0021-9010.76.2.316.
- INTERSTATE COMMISSION FOR ADULT OFFENDER SUPERVISION. 2007. *Sex offender assessment information survey*, Lexington (Ky.), chez l'auteur. Document de l'ICAOS n° 4-2007.
- JACKSON, R. L., et D. T. HESS. 2007. « Evaluation for civil commitment of sex offenders: A survey of experts », *Sexual Abuse: A Journal of Research and Treatment*, vol. 19, p. 409-448. DOI : 10.1007/s11194-007-9062-3.
- *JOHANSEN, S. H. 2007. « Accuracy of predictions of sexual offense recidivism: A comparison of actuarial and clinical methods », *Dissertation Abstracts International : Section B*, vol. 68, n° 3. N° du UMI : 3255527.
- *KNIGHT, R. A., et T D. HORNTON. 2007. *Evaluating and improving risk assessment schemes for sexual recidivism: A long-term follow-up of convicted sexual offenders*. Document n° 217618. Présenté au département de la Justice des États-Unis.
- KRAEMER, H. C., A. E. KAZDIN, D. R. OFFORD, R. C. KESSLER, P. S. JENSEN et D. J. KUPLER, 1997. « Coming to terms with the terms of risk », *Archives of General Psychiatry*, vol. 54, p. 337-343.

- *LANGSTRÖM, N. 2004. « Accuracy of actuarial procedures for assessment of sexual offender recidivism risk may vary across ethnicity », *Sexual Abuse: A Journal of Research and Treatment*, vol. 16, p. 107-120. DOI : 10.1177/107906320401600202.
- LLOYD, M. D. 2008. « Incremental validity of commonly-used risk assessment measures in predicting serious sexual recidivism », *Dissertation Abstracts International : Section B: The Sciences and Engineering*, vol. 69, n° 9, p. 5784.
- LOFTUS, G. R., M. A. OBERG et A. M. DILLON. 2004. « Linear theory, dimensional theory, and the face-inversion effect », *Psychological Review*, vol. 111, p. 835-863. DOI : 10.1037/0033-295X.111.4.835.
- MALLOY, P. F., J. L. CUMMINGS, C. E. COFFEY, J. DUFFY, M. FINK, E. C. LAUTERBACH, M. LOVELL, D. ROYALL et S. SALLOWAY. 1997. « Cognitive screening instruments in neuropsychiatry: A report of the Committee on Research of the American Neuropsychiatric Association », *The Journal of Neuropsychiatry and Clinical Neurosciences*, vol. 9, p. 189-197.
- MANN, R. E., R. K. HANSON et D. THORNTON. 2010. « Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors », *Sexual Abuse: A Journal of Research and Treatment*, vol. 22, p. 191-217. DOI : 10.1177/1079063210366039.
- MCGRATH, R. J., G. F. CUMMING, B. L. BURCHARD, S. ZEOLI et L. ELLERBY. 2010. *Current practices and emerging trends in sexual abuser management: The Safer Society 2009 North American Survey*, Brandon (Vt.), Safer Society Press. ISBN : 978-1-884444-85-2. Résumé des conclusions canadiennes en français, *Pratiques courantes dans les programmes canadiens de traitement des agresseurs sexuels : sondage 2009 de la Safer Society Foundation*, à l'adresse http://www.publicsafety.gc.ca/res/cor/rep/_fl/2010-02-sss-fra.pdf.
- MEEHL, P. E. 1954. *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*, Minneapolis (Minn.), University of Minnesota Press.
- MEEHL, P. E. 1956. « Wanted—A good cook-book », *American Psychologist*, vol. 11, p. 263-272. DOI : 10.1037/h0044164.
- MILLS, J. F., et D. G. KRONER. 2006. « The effect of discordance among violence and general recidivism risk estimates on predictive accuracy », *Criminal Behaviour and Mental Health*, vol. 16, p. 155-166. DOI : 10.1002/cbm.623.
- MONAHAN, J., et L. WALKER. 2010. *Social science in law: Cases and materials*, New York, Foundation Press.
- *NICHOLAICHUK, T. 2001. *The comparison of two standardized risk assessment instruments in a sample of Canadian Aboriginal sexual offenders*. Présenté à la conférence annuelle sur la recherche et le traitement de l'Association for the Treatment of Sexual Abusers, San Antonio (Tex.).
- NUFFIELD, J. 1982. *La libération conditionnelle au Canada : recherches en vue d'une normalisation des décisions*, Ottawa (Ont.), Solliciteur général du Canada.
- NUNNALLY, J. C., et I. H. BERNSTEIN. 1994. *Psychometric theory*, 3^e éd., New York, McGraw-Hill.
- PHENIX, A., D. DOREN, L. HELMUS, R. K. HANSON et D. THORNTON. 2009. *Règles de codage pour l'échelle Statique-2002*, Ottawa (Ont.), Sécurité publique Canada. Adresse : <http://www.securitepublique.gc.ca/res/cor/rep/sttc-2002-fra.aspx>.
- QUINSEY, V. L., G. T. HARRIS, M. E. RICE et C. A. CORMIER. 2006. *Violent offenders: Appraising and managing risk*, 2^e éd., Washington (D.C.), American Psychological Association.

- RETTENBERGER, M., A. MATTHES, D. P. BOER et R. EHER. 2010. « Prospective actuarial risk assessment: A comparison of five risk assessment instruments in different sexual offender subtypes », *International Journal of Offender Therapy and Comparative Criminology*, vol. 54, p. 169-186. DOI : 10.1177/0306624X08328755.
- RICE, M.E., et G.T. HARRIS. 1995. « Violent recidivism: Assessing predictive validity », *Journal of Consulting and Clinical Psychology*, vol. 63, p. 737-748.
- ROBERTS, C. F., D. M. DOREN et D. THORNTON. 2002. « Dimensions associated with assessments of sex offender recidivism risk », *Criminal Justice and Behavior*, vol. 29, p. 569-589. DOI : 10.1177/009385402236733.
- RUST, J., et S. GOLOMBOK. 2009. *Modern psychometrics: The science of psychological assessment*, 3^e éd., London (U.K.), Taylor and Francis.
- SCHMIDT, F. L., I. OH et T. L. HAYES. 2009. « Fixed- versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results », *British Journal of Mathematical and Statistical Psychology*, vol. 62, p. 97-128. DOI : 10.1348/000711007X255327.
- SETO, M. C. 2005. « Is more better? Combining actuarial risk scales to predict recidivism among adult sex offenders », *Psychological Assessment*, vol. 17, p. 156-167. DOI : 10.1037/1040-3590.17.2.156.
- SLEDJESKI, E. M., L. C. DIERKER, D. COSTELLO, S. SHIFFMAN, E. DONNY et B. R. FLAY. 2007. « Predictive validity of four nicotine dependence measures in a college sample », *Drug and Alcohol Dependence*, vol. 87, p. 10-19. DOI : 10.1016/j.drugalcdep.2006.07.005.
- STALANS, L. J., R. HACKER et M. E. TALBOT. 2010. « Comparing nonviolent, other-violent, and domestic batterer sex offenders: Predictive accuracy of risk assessments on sexual recidivism », *Criminal Justice and Behavior*, vol. 37, p. 613-628. DOI : 10.1177/0093854810363794.
- *SWINBURNE ROMINE, R., S. M. DWYER, C. MATHIOWETZ et M. THOMAS. 2008. *Thirty years of sex offender specific treatment: A follow-up study*. Affiche présentée à la conférence de l'Association for the Treatment of Sexual Abusers, Atlanta (Ga.).
- *TERNOWSKI, D. R. 2004. « Sex offender treatment: An evaluation of the Stave Lake Correctional Centre Program », *Dissertation Abstracts International : Section B*, vol. 66, n° 6, p. 3428.
- THORNTON, D. 2002. « Constructing and testing a framework for dynamic risk assessment », *Sexual Abuse: A Journal of Research and Treatment*, vol. 14, p. 139-153. DOI : 10.1177/107906320201400205.
- THORNTON, D., R. MANN, S. WEBSTER, L. BLUD, R. TRAVERS, C. FRIENDSHIP et M. ERIKSON. 2003. « Distinguishing and combining risks for sexual and violent recidivism », *Annals of the New York Academy of Sciences*, vol. 989, p. 225-235.
- VRIEZE, S. I., et W. M. GROVE. 2009. « Survey on the use of clinical and mechanical prediction methods in clinical psychology », *Professional Psychology: Research and Practice*, vol. 40, p. 525-531. DOI : 10.1037/a0014693.
- VRIEZE, S. I., et W. M. GROVE. 2010. « Multidimensional assessment of criminal recidivism: Problems, pitfalls, and proposed solutions », *Psychological Assessment*, vol. 22, p. 382-395. DOI : 10.1037/a0019228.
- WEINER, I. B. 2003. « The assessment process », dans J. R. Graham, J. A. Naglieri et I. B. Weiner (éd.), *Handbook of psychology: Volume 10, Assessment psychology*, Hoboken (N.J.), John Wiley & Sons, Inc., p. 3-26.

- WELSH, J. L., F. SCHMIDT, L. MCKINNON, H. K. CHATTHA et J. R. MEYERS. 2008. « A comparative study of adolescent risk assessment instruments: Predictive and incremental validity », *Assessment*, vol. 15, p. 104-115. DOI : 10.1177/1073191107307966.
- *WILSON, R. J., F. CORTONI et M. VERMANI. 2007a. *Cercles de soutien et de responsabilité : reproduction à l'échelle nationale des résultats obtenus*, Rapport de recherche n° R-185, Ottawa (Ont.), Service correctionnel du Canada. Adresse : <http://www.csc-scc.gc.ca/text/rsrch/reports/r185/r185-fra.shtml>.
- *WILSON, R. J., J. E. PICHECA et M. PRINZO. 2007b. « Evaluating the effectiveness of professionally-facilitated volunteerism in the community-based management of high-risk sexual offenders: Part two – A comparison of recidivism rates », *The Howard Journal*, vol. 46, p. 327-337. DOI : 10.1111/j.1468-2311.2007.00480.x.

Annexe A

Tableau 1A – ASC de la FER des échelles ERRRS, Statique-99R et Statique-2002R, par échantillon.

| Échantillon | N | Récidive sexuelle | | Récidive avec violence | | Récidive en général | |
|--------------------------------------|-----|-------------------|-----------|------------------------|-----------|---------------------|-----------|
| | | ASC de la FER | IC à 95 % | ASC de la FER | IC à 95 % | ASC de la FER | IC à 95 % |
| Allan et coll. (2007) | 492 | 0,70 | 0,62 | 0,60 | 0,53 | 0,57 | 0,51 |
| | | 0,72 | 0,64 | 0,69 | 0,63 | 0,70 | 0,65 |
| Bengtson (2008) | 308 | 0,61 | 0,54 | 0,60 | 0,54 | 0,59 | 0,52 |
| | | 0,62 | 0,56 | 0,66 | 0,60 | 0,64 | 0,57 |
| | | 0,64 | 0,57 | 0,66 | 0,60 | 0,67 | 0,60 |
| Bigras (2007) | 457 | 0,60 | 0,48 | 0,54 | 0,47 | 0,55 | 0,48 |
| | | 0,71 | 0,60 | 0,69 | 0,62 | 0,69 | 0,64 |
| | | 0,70 | 0,59 | 0,69 | 0,63 | 0,71 | 0,65 |
| Boer (2003) | 296 | 0,69 | 0,57 | 0,62 | 0,55 | 0,60 | 0,53 |
| | | 0,75 | 0,65 | 0,75 | 0,69 | 0,79 | 0,74 |
| | | 0,73 | 0,63 | 0,74 | 0,67 | 0,81 | 0,76 |
| Bonta et Yessine (2005) | 133 | 0,50 | 0,36 | 0,47 | 0,36 | 0,48 | 0,38 |
| | | 0,64 | 0,52 | 0,66 | 0,57 | 0,65 | 0,56 |
| Brouillette-Alarie et Proulx (2008) | 228 | 0,67 | 0,59 | 0,60 | 0,53 | - | - |
| | | 0,68 | 0,60 | 0,69 | 0,62 | - | - |
| Cortoni et Nunes (2007) ^a | 73 | - | - | 0,73 | 0,56 | 0,64 | 0,49 |
| | | - | - | 0,71 | 0,53 | 0,70 | 0,54 |
| Eher et coll. (2008) | 706 | 0,75 | 0,65 | 0,68 | 0,63 | 0,63 | 0,58 |
| | | 0,71 | 0,61 | 0,76 | 0,71 | 0,71 | 0,67 |
| Epperson (2003) | 177 | 0,73 | 0,62 | - | - | - | - |
| | | 0,78 | 0,67 | - | - | - | - |
| Haag (2005) | 190 | 0,64 | 0,55 | - | - | - | - |
| | | 0,70 | 0,61 | - | - | - | - |
| | | 0,67 | 0,58 | - | - | - | - |

(Suite du tableau à la page suivante)

Tableau IA (suite).

| Échantillon | N | Récidive sexuelle | | Récidive avec violence | | Récidive générale | |
|----------------------------------|-------|-------------------|-----------|------------------------|-----------|-------------------|-----------|
| | | ASC de la FER | IC à 95 % | ASC de la FER | IC à 95 % | ASC de la FER | IC à 95 % |
| Hanson et coll. (2007) | 702 | | | | | | |
| ERRRS | | 0,70 | 0,64 | 0,77 | 0,57 | 0,61 | 0,56 |
| Statique-99R | | 0,76 | 0,69 | 0,82 | 0,70 | 0,76 | 0,72 |
| Statique-2002R | | 0,75 | 0,69 | 0,82 | 0,72 | 0,76 | 0,73 |
| Harkins et Beech (2007) | 190 | | | | | | |
| ERRRS | | 0,73 | 0,64 | 0,83 | 0,56 | 0,62 | 0,54 |
| Statique-99R | | 0,78 | 0,68 | 0,88 | 0,67 | 0,77 | 0,69 |
| Statique-2002R | | 0,79 | 0,68 | 0,89 | 0,69 | 0,77 | 0,70 |
| Hill et coll. (2008) | 86 | | | | | | |
| ERRRS | | 0,61 | 0,46 | 0,76 | 0,45 | 0,53 | 0,40 |
| Statique-99R | | 0,67 | 0,52 | 0,82 | 0,50 | 0,61 | 0,48 |
| Johansen (2007) | 273 | | | | | | |
| ERRRS | | 0,63 | 0,52 | 0,75 | 0,53 | 0,57 | 0,50 |
| Statique-99R | | 0,65 | 0,53 | 0,77 | 0,63 | 0,72 | 0,66 |
| Knight et Thornton (2007) | 466 | | | | | | |
| ERRRS | | 0,62 | 0,56 | 0,68 | 0,53 | 0,56 | 0,51 |
| Statique-99R | | 0,62 | 0,56 | 0,67 | 0,59 | 0,63 | 0,58 |
| Statique-2002R | | 0,63 | 0,58 | 0,69 | 0,59 | 0,64 | 0,59 |
| Långström (2004) | 1 278 | | | | | | |
| ERRRS | | 0,72 | 0,66 | 0,78 | 0,61 | - | - |
| Statique-99R | | 0,73 | 0,68 | 0,79 | 0,75 | - | - |
| Nicholaichuk (2001) | 281 | | | | | | |
| ERRRS | | 0,67 | 0,60 | 0,75 | - | - | - |
| Statique-99R | | 0,74 | 0,67 | 0,81 | - | - | - |
| Swinburne Romine et coll. (2008) | 680 | | | | | | |
| ERRRS | | 0,62 | 0,55 | 0,68 | - | - | - |
| Statique-99R | | 0,63 | 0,57 | 0,70 | - | - | - |
| Ternowski (2004) | 247 | | | | | | |
| ERRRS | | 0,61 | 0,48 | 0,74 | 0,50 | 0,61 | 0,52 |
| Statique-99R | | 0,75 | 0,66 | 0,85 | 0,69 | 0,77 | 0,70 |
| Wilson et coll. (2007a et b) | 228 | | | | | | |
| ERRRS | | 0,63 | 0,51 | 0,75 | 0,43 | 0,50 | 0,42 |
| Statique-99R | | 0,57 | 0,43 | 0,71 | 0,54 | 0,60 | 0,52 |

Remarques : Taux de récidive sexuelle établis à partir de tous les cas, sans neutraliser le nombre d'années de suivi utilisé pour calculer les valeurs de l'ASC.
^aCortoni et Nunes (2007) n'ont pu calculer aucune valeur de l'ASC pour la récidive sexuelle, car il n'y avait aucun récidiviste.

Annexe B

Tableau 1B – *Rapports de risque des échelles Statique-99R et Statique-2002R.*

| | | Statique-99R | | Statique-2002R | |
|---------------|------------------|--------------|------|----------------|------|
| | | Score | RR | Score | RR |
| Faible | < 0,50 | | | -2 | 0,11 |
| | | -3 | 0,26 | -1 | 0,17 |
| | | -2 | 0,34 | 0 | 0,26 |
| | | -1 | 0,45 | 1 | 0,38 |
| | | 0 | 0,59 | 2 | 0,54 |
| Moyen | 0,50-2,00 | 1 | 0,77 | 3 | 0,74 |
| | | 2 | 1,00 | 4 | 1,00 |
| | | 3 | 1,31 | 5 | 1,31 |
| | | 4 | 1,71 | 6 | 1,68 |
| | | 5 | 2,23 | 7 | 2,08 |
| Élevé | ≥ 2,00 | 6 | 2,91 | 8 | 2,52 |
| | | 7 | 3,80 | 9 | 2,97 |
| | | 8 | 4,96 | 10 | 3,40 |
| | | 9 | 6,48 | 11+ | 3,79 |
| | | 10+ | 8,47 | | |

Remarques : RR = risque relatif. Les rapports de risque pour la Statique-99R ont été calculés dans une étude antérieure (Hanson et coll., 2010) au moyen de coefficients de régression (Cox) établis à partir des scores de la Statique-99R introduits ($B = 0,267$; $ET = 0,013$; $Wald = 413,11$; $p < 0,001$), avec l'échantillon comme strate ($k = 22$, $n = 8\ 047$). Les rapports de risque pour la Statique-2002R ont été calculés dans une étude antérieure (Hanson et coll., 2010) au moyen de coefficients de régression (Cox) établis à partir des scores initiaux de la Statique-2002R ($B = 0,285$; $ET = 0,033$; $Wald = 74,24$; $p < 0,001$) et des scores de la Statique-2002R au carré ($B = -0,013$; $ET = 0,006$; $Wald = 4,66$; $p = 0,031$), avec l'échantillon comme strate ($k = 7$, $n = 2\ 610$).