

L'efficacité d'outils d'évaluation du risque d'abus physique envers les enfants

Septembre 2015

Une production de l'Institut national d'excellence en santé et en services sociaux

Rapport rédigé par
Jean Damasse

Avec la collaboration de
Isabelle Beaumier (CJQ-IU)
Ève-Line Bussières (CJQ-IU)

Sous la direction de
Sylvie Bouchard

Le présent rapport a été présenté au Comité scientifique permanent de l'Institut national d'excellence en santé et en services sociaux (INESSS) lors de sa réunion du 27 mars 2015.

Le contenu de cette publication a été rédigé et édité par l'INESSS.

Ce document et ses annexes sont accessibles en ligne dans la section *Publications* de notre site *Web*.

Équipe de projet

Auteur

Jean Damasse, M.A.

Collaborateurs

Isabelle Beaumier, M.A.

Ève-Line Bussièrès, Ph. D.

Direction scientifique

Sylvie Bouchard, Ph. D., M. Sc., M.B.A.

Coordination scientifique

Micheline Lapalme, Ph. D.

Recherche d'information scientifique

Lysane St-Amour, MBSI

Mathieu Plamondon, MSI

Soutien documentaire

Micheline Paquin, *tech. doc.*

Édition

Responsable

Renée Latulippe

Coordination

Véronique Baril

Traduction

Jocelyne Lauzière

Mise en page

Service des communications et du
transfert de connaissance

Vérification bibliographique

Denis Santerre

L'Institut remercie les membres de son personnel qui ont contribué à l'élaboration du présent document.

Dépôt légal

Bibliothèque et Archives nationales du Québec, 2015

Bibliothèque et Archives Canada, 2015

ISSN 1915-3104 INESSS (PDF)

ISBN 978-2-550-73988-3 (PDF)

© Gouvernement du Québec, 2015

La reproduction totale ou partielle de ce document est autorisée à condition que la source soit mentionnée.

Pour citer ce document : Institut national d'excellence en santé et en services sociaux (INESSS). L'efficacité d'outils d'évaluation du risque d'abus physique envers les enfants. Rapport rédigé par Jean Damasse avec la collaboration d'Isabelle Beaumier et d'Ève-Line Bussièrès. ETMIS 2015; 11(3): 1-92.

L'Institut remercie les membres de son personnel qui ont contribué à l'élaboration du présent document.

Comité de suivi

Isabelle Beaumier, Centre jeunesse de Québec – Institut universitaire (CJQ-IU)

Catherine Bessette, Alliance des comités des usagers des centres jeunesse du Québec (ACUCJQ)

Louise Bourassa, Centre jeunesse de la Mauricie et du Centre-du-Québec (CJMCQ)

Gilles Chabot, Centre hospitalier universitaire (CHU) Sainte-Justine

Geneviève Cloutier, Ordre des travailleurs sociaux et des thérapeutes conjugaux et familiaux du Québec (OTSTCFQ)

Nicole D’Astous, Centre jeunesse de Montréal – Institut universitaire (CJM-IU)

Maryse Davreux, Centre jeunesse de la Montérégie (CJM)

Anne Duret, Association des centres jeunesse du Québec (ACJQ)

Nathalie Girouard, Ordre des psychologues du Québec (OPQ)

Jean Hénault, Ordre des psychoéducateurs et psychoéducatrices du Québec (OPPQ)

Brigitte Moreault, Association des centres jeunesse du Québec (ACJQ)

Chantal Pilote, Centres jeunesse du Québec -IU

Marc Plamondon, ministère de la Santé et des Services sociaux (MSSS)

François Tremblay, Conseil d’administration transitoire pour l’intégration des criminologues au système professionnel québécois

Marie-Andrée Ulysse, Association québécoise d’établissements de santé et de services sociaux (AQESSS)

Experts indépendants

Tonino Esposito, Ph. D., professeur adjoint, École de service social – FAS, Université de Montréal

Marie-Ève Clément, Ph. D., chercheuse à l’Université du Québec en Outaouais

Sonia Hélie, Ph. D., chercheuse au CJM-IU

Lecteurs externes

La lecture externe est un des mécanismes utilisés par l’INESSS pour assurer la qualité de ses travaux. Les lecteurs externes valident les aspects méthodologiques de l’évaluation, de même que l’exactitude du contenu, en fonction de leurs domaines d’expertise respectifs.

Les lecteurs externes du présent rapport sont :

Delphine Collin-Vézina, Ph. D., directrice du Centre de recherche sur l’enfance et la famille, titulaire de la Chaire de recherche en protection de l’enfance de l’Université McGill

Carl Lacharité, Ph. D., professeur titulaire au Département de psychologie, directeur du Centre d’études interdisciplinaires sur le développement de l’enfant et la famille, Université du Québec à Trois-Rivières (UQTR)

Chantal Lavergne, Ph. D., chercheuse, CJM-IU, professeure associée à l’École de service social, Université de Montréal

Autres contributions

Outre les lecteurs externes, l'INESSS tient aussi à remercier les personnes suivantes, qui ont contribué à la rédaction de ce rapport : les intervenants, les agents de liaison et les gestionnaires de CSSS et de CJ qui ont fait part de leurs points de vue dans des groupes de discussion focalisée. L'INESSS remercie également les personnes qui ont assuré l'organisation de ces rencontres.

Déclaration d'intérêts

Carl Lacharité est membre du conseil d'administration de l'INESSS.

Responsabilité

L'INESSS assume l'entière responsabilité de la forme et du contenu définitifs de ce document. Les conclusions et les recommandations ne reflètent pas forcément les opinions des lecteurs externes ou des autres personnes consultées aux fins de son élaboration et de sa rédaction.

COMITÉ SCIENTIFIQUE PERMANENT EN SANTÉ ET EN SERVICES SOCIAUX

Membres

M. Pierre Dostie

- Directeur des clientèles en Dépendances (CSSS et Centre de réadaptation en dépendances), Santé mentale, Enfance, jeunesse et famille, et Santé publique au CSSS de Jonquière
- Chargé de cours en travail social, Université du Québec à Chicoutimi

M. Hubert Doucet

- Consultant en bioéthique
- Professeur associé, Faculté de théologie et de sciences des religions, Université de Montréal

M. Serge Dumont

- Directeur scientifique, CSSS de la Vieille-Capitale
- Directeur du Réseau de collaboration sur les pratiques interprofessionnelles en santé
- Professeur titulaire, École de service social, Université Laval
- Chercheur, Centre de recherche en cancérologie, Hôtel-Dieu de Québec

M^{me} Isabelle Ganache

- Consultante en éthique, Commissaire à la santé et au bien-être
- Professeure adjointe de clinique, Programmes de bioéthique, Département de médecine sociale et préventive, Faculté de médecine, Université de Montréal

M. Jude Goulet

- Pharmacien, chef du Département de pharmacie, Hôpital Maisonneuve-Rosemont

M^{me} Marjolaine Landry

- Professeure, Département des sciences infirmières, Centre universitaire de Drummondville, Université du Québec à Trois-Rivières
- Chercheuse, Centre affilié universitaire, Centre de santé et de services sociaux - Institut universitaire de gériatrie de Sherbrooke

M^{me} Claudine Laurier

- Pharmacienne
- Professeure titulaire, Faculté de pharmacie, Université de Montréal

M^{me} Esther Leclerc (présidente)

- Infirmière
- Ex-directrice générale adjointe, Centre hospitalier de l'Université de Montréal

D^r Michaël Malus

- Médecin de famille,
- Chef du département de médecine familiale et responsable du Centre Réseau Intégré Universitaire (CRIU) Herzl de l'Hôpital général juif – Sir Mortimer B. Davis
- Professeur associé, Département de médecine familiale, Université McGill

D^r Maurice St-Laurent

- Gériatre
- Professeur agrégé de clinique, Faculté de médecine, Université Laval

M. Jean Toupin (vice-président)

- Professeur associé, Département de psychoéducation, Université de Sherbrooke
- Chercheur, Institut universitaire en santé mentale de Montréal

Membres citoyens

M. Marc Bélanger

- Psychoéducateur à la retraite
- Professionnel expert de l'intervention des autorités publiques et de l'organisation des services en interdisciplinarité et intersectorialité (santé et services sociaux, justice, éducation) auprès des personnes éprouvant des difficultés d'adaptation psychosociale

M^{me} Jeannine Tellier-Cormier

- Professeure en soins infirmiers à la retraite, Cégep de Trois-Rivières

Membres experts invités

M. Éric A. Latimer

- Économiste et professeur titulaire, Département de psychiatrie, Faculté de médecine, Université McGill
- Chercheur, Institut universitaire en santé mentale Douglas

M. Aimé Robert LeBlanc

- Ingénieur, professeur émérite, Institut de génie biomédical, Département de physiologie, Faculté de médecine, Université de Montréal
- Directeur adjoint à la recherche et au développement, Centre de recherche de l'Hôpital du Sacré-Cœur de Montréal

D^r Réginald Nadeau

- Cardiologue et chercheur, Centre de recherche de l'Hôpital du Sacré-Cœur de Montréal
- Professeur émérite, Faculté de médecine, Université de Montréal

D^r Raghu Rajan

- Hématologue et oncologue médical, Centre universitaire de santé McGill
- Professeur associé, Université McGill
- Membre du Comité de l'évolution des pratiques en oncologie (CEPO) et du programme de gestion thérapeutique des médicaments

Membre observateur MSSS

M^{me} Natalie Rosebush

- Directrice générale adjointe aux services aux aînés, Direction générale des services sociaux, ministère de la Santé et des Services sociaux

TABLE DES MATIÈRES

AVANT-PROPOS	i
EN BREF	ii
RÉSUMÉ	iv
SUMMARY	viii
SIGLES ET ABRÉVIATIONS	xi
GLOSSAIRE	xii
INTRODUCTION	1
1 MÉTHODOLOGIE	3
1.1 Modèle logique et questions clés d'évaluation.....	3
1.2 Question 1 – Efficacité théorique (revue systématique)	4
1.3 Questions 2 et 3 – Efficacité pratique (revue narrative et consultation).....	6
1.4 Question 4 – Adaptation au Québec	9
2 RÉSULTATS	10
2.1 Question 1 – Efficacité théorique des outils (fiabilité, validité)	10
2.2 Validité des outils	13
2.3 Questions 2 et 3 – Efficacité pratique des outils (utilité, applicabilité)	34
2.3.1 Revue de la littérature.....	34
2.3.2 Utilité des outils.....	35
2.3.3 Applicabilité des outils.....	36
2.4 Consultation des milieux d'intervention	38
2.4.1 Groupes de discussion focalisée – Synthèse des résultats.....	38
2.5 Question 4 – Consultations des parties prenantes	42
3 DISCUSSION	44
3.1 Efficacité théorique (fiabilité, validité) des outils identifiés.....	44
3.1.1 Fiabilité des outils.....	44
3.1.2 Validité des outils	44
3.2 Efficacité pratique (utilité, applicabilité) des outils.....	47
3.3 Forces et limites de l'avis	49
CONCLUSIONS ET RECOMMANDATIONS	51
RÉFÉRENCES	55

LISTE DES TABLEAUX

Tableau 1	Taux de nouveaux signalements d'abus (tous types confondus) selon les niveaux de risque.....	17
Tableau 2	Taux de réévaluation des cas d'abus (tous types confondus) selon les niveaux de risque .	21
Tableau 3	Taux de nouvelles allégations fondées d'abus (tous types confondus) selon les niveaux de risque.....	25
Tableau 4	Taux de placement subséquent d'un enfant en dehors du foyer familial en raison de maltraitance (abus et négligence), selon les niveaux de risque	28
Tableau 5	Validité prédictive d'outils d'évaluation de la maltraitance (abus et négligence).....	30

LISTE DES FIGURES

Figure 1	Modèle logique	3
Figure 2	Validité prédictive et prise de décision	45

AVANT-PROPOS

Le présent avis est issu d'une requête de l'Association des centres jeunesse du Québec (ACJQ) qui représentait les 16 centres jeunesse du Québec au début des travaux de l'INESSS. L'ACJQ a été dissoute le 1^{er} avril 2015 en vertu de la *Loi modifiant l'organisation et la gouvernance du réseau de la santé et des services sociaux, notamment par l'abolition des agences régionales*. Les centres jeunesse ont été intégrés dans de nouvelles structures de gouvernance : les centres intégrés de santé et de services sociaux (CISSS) ou les centres intégrés de santé et de services sociaux instituts universitaires (CIUSSS).

Dans le présent document, le terme « centres jeunesse du Québec » réfère à la mission de ces établissements avant la sanction du projet de Loi 10. Cette mission n'a pas été modifiée. Depuis le 1^{er} avril 2015, les centres jeunesse du Québec et les équipes dédiées au programme services « Jeunes en difficulté », en première ligne dans les CSSS, sont regroupés sous la direction « Programme jeunesse » à l'intérieur des CISSS et des CIUSSS.

La réorganisation du réseau est susceptible d'entraîner des changements de pratique entre les partenaires du réseau « Jeunes en difficulté », de manière à assurer une offre de services harmonisée pour une clientèle commune en besoin de services ou de protection. La question de l'évaluation du risque d'abus physique abordée dans cet avis doit être considérée dans l'optique de cette nouvelle offre de services intégrés pour les jeunes en difficulté, où le partage des responsabilités reste à préciser. Bien qu'elles s'inscrivent dans une structure et une organisation de services modifiées, les recommandations formulées par l'INESSS demeurent applicables de même que les fondements théoriques et pratiques sur lesquels elles s'appliquent. Bien que ces recommandations s'adressaient aux centres jeunesse lors de la rédaction du présent avis, le lecteur comprendra que celles-ci s'adressent désormais aux directions « Programme jeunesse » à l'intérieur des CISSS et des CIUSSS.

EN BREF

La demande adressée à l'Institut national d'excellence en santé et en services sociaux (INESSS) par l'Association des centres jeunesse du Québec (ACJQ) visait à repérer des outils d'évaluation du risque d'abus physique envers les enfants, à évaluer leur efficacité théorique (fiabilité, validité) et leur efficacité pratique (utilité, applicabilité dans les milieux d'intervention). Les résultats d'une revue systématique de la littérature, les points de vue d'acteurs sur le terrain recueillis lors de groupes de discussion focalisée, ainsi que les réflexions des parties prenantes à ce projet ont contribué à la formulation de conclusions et de diverses recommandations.

Les conclusions sur l'efficacité théorique et sur l'efficacité pratique de ce type d'outil sont les suivantes :

- La majorité des études recensées sur ce type d'outil ne porte pas uniquement sur le risque d'abus physique et sa récurrence, puisque la prédiction englobe l'abus en général et, parfois, la négligence;
- Aucun des outils évalués, et habituellement utilisés à l'étape de l'évaluation de la situation de l'enfant, ne satisfait à la fois aux critères d'une bonne fiabilité et d'une bonne validité. Les outils CAPI (Child Abuse Potential Inventory) et FRAAN (Family Risk of Abuse and Neglect) offrent une bonne fiabilité. La validité de l'outil CFRA (California Family Risk Assessment) est jugée adéquate. Ce dernier a été adapté en Ontario (Canada) sous le nom OFRA (Ontario Family Risk Assessment), où il est actuellement en usage;
- Lorsqu'il fait partie d'un système d'évaluation intégré, tel que le SDM (Structured Decision MakingMD) mis au point par le Children's Research Center (CRC) du National Council on Crime and Delinquency (NCCD) aux États-Unis, l'outil actuariel fournit une meilleure évaluation du risque d'abus physique que lorsqu'il est utilisé seul;
- Les acteurs sur le terrain qui ont été interrogés se disent intéressés par un outil d'évaluation du risque d'abus physique pour standardiser le processus d'évaluation des cas, favoriser un langage commun, partager les mêmes critères d'évaluation, et harmoniser les pratiques d'intervention. L'applicabilité de cet outil est conditionnelle à une formation appropriée du personnel, au caractère distinct de l'outil par rapport aux autres types d'outils qui sont déjà en usage au Québec, et à sa capacité de soutenir la prise de décision sur la priorité et l'intensité des services à offrir à la clientèle concernée.

En conséquence, les trois recommandations suivantes sont adressées aux centres jeunesse du Québec (regroupés sous la direction « Programme jeunesse » à l'intérieur des CISSS et des CIUSSS depuis le 1er avril 2015).

- 1) Intégrer un outil de type actuariel pour soutenir l'évaluation du niveau de risque d'abus physique que présentent les familles. Ce type d'évaluation ne doit pas s'appuyer sur un seul outil, mais sur un ensemble d'outils d'évaluation, tel que le SDM, afin de tenir compte de la situation, des besoins et du contexte des familles évaluées;
- 2) Avant de procéder à l'intégration proprement dite d'un outil actuariel, considérer et documenter la nature des facteurs de risque au Québec, la pertinence et la faisabilité d'adapter l'outil CFRA (dont la version adaptée au contexte ontarien est l'OFRA), le degré

de correspondance entre le Système de soutien à la pratique (SSP), en usage dans les centres jeunesse du Québec, et le SDM;

- 3) Dans l'éventualité où le déploiement à grande échelle de l'outil serait proposé, s'assurer d'un engagement organisationnel afin d'appuyer les changements apportés aux pratiques d'intervention, de définir clairement les procédures et les usages de l'outil au moyen de lignes directrices et de pouvoir répondre à la demande de services.

RÉSUMÉ

Mise en contexte et mandat

Au Québec, en 2007, l'introduction de la notion de *risque sérieux* en matière de négligence, d'abus sexuel et d'abus physique dans la Loi sur la protection de la jeunesse rend compte de l'ampleur croissante de la récurrence de ces types de maltraitance. Dans les services de protection de la jeunesse (SPJ), cela se manifeste, entre autres, par de nouveaux signalements, une réévaluation de cas, de nouvelles allégations fondées et le placement successif d'enfants en dehors du foyer familial. Le parcours cyclique entre les services de première ligne - centres de santé et de services sociaux (CSSS) - et les services de deuxième ligne - centres jeunesse (CJ) - destinés aux enfants ayant vécu ou à risque de revivre une ou plusieurs situations d'abus physique rend compte de ce phénomène qui préoccupe les autorités responsables de ces services. Dans l'optique d'améliorer la pratique d'évaluation, de favoriser une meilleure gestion du risque et de mieux cibler la clientèle afin de lui offrir des services adaptés selon les ressources disponibles, l'Association des centres jeunesse du Québec (ACJQ) a présenté une demande à l'Institut national d'excellence en santé et en services sociaux (INESSS) afin de repérer un ou des outils d'évaluation du risque d'abus physique envers les enfants.

Par définition, un outil d'évaluation du risque d'abus physique est composé d'une série de questions qui sert à évaluer, avec plus de précision et de cohérence que le seul jugement professionnel, les enfants à risque de subir des sévices corporels pouvant compromettre leur santé et leur sécurité. Un tel outil devrait répondre à la fois aux critères d'efficacité théorique (fiabilité, validité), sur le plan scientifique, et aux critères d'efficacité pratique (utilité et applicabilité dans les milieux d'intervention). L'INESSS a donc produit, en collaboration avec le Centre jeunesse de Québec – Institut universitaire (CJQ-IU), le présent avis, qui vise à répondre aux quatre questions suivantes :

- 1) Quelle est l'efficacité reconnue, sur le plan scientifique, des outils d'évaluation du risque d'abus physique envers les enfants?
- 2) Ces outils dont l'efficacité est reconnue sont-ils utiles à la prise de décision au moment d'évaluer des enfants qui font l'objet d'un signalement ou d'un nouveau signalement pour un risque d'abus physique?
- 3) Comment ces outils dont l'efficacité est reconnue s'intègrent-ils au contexte d'intervention?
- 4) Ces outils dont l'efficacité est reconnue sont-ils adaptables au contexte légal québécois?

Méthodologie

Une revue systématique de la littérature scientifique couvrant les années 2003 à 2013 a été réalisée afin de répondre à la première question. Les réponses à la deuxième et à la troisième question s'appuient sur une revue de la littérature grise, et la tenue de groupes de discussion focalisée auxquels ont participé 21 personnes, soit des intervenants, des agents de liaison et des gestionnaires de CSSS et de CJ du Québec. La quatrième question a été débattue avec les membres du comité de suivi du projet et commentée par des experts indépendants.

Résultats

La revue systématique a permis de recenser 21 études qui présentent des résultats sur l'efficacité théorique de 7 outils d'évaluation du risque d'abus physique envers les enfants. Utilisés à l'étape d'évaluation de la situation de l'enfant, ces outils sont les suivants : AAPI-2 (Adult Adolescent Parenting Inventory-2), C-CAPS (Cleveland Child Abuse Potential Scale), CAPI (Child Abuse Potential Inventory), CFRA (California Family Risk Assessment), CFRAT (Colorado Family Risk Assessment Tool), FRAAN (Family Risk Assessment of Abuse and Neglect) et OFRA (Ontario Family Risk Assessment). Il est à noter que la majorité des études recensées ne porte pas uniquement sur le risque d'abus physique, et sur sa récurrence, puisque la prédiction englobe l'abus en général et, parfois, la négligence.

Ces outils sont de deux types : 1) des outils d'évaluation structurée du risque de type contextuel, qui s'appuient sur le consensus, issu de la littérature et des conseils d'experts, au sujet des facteurs et des conditions jugés le plus fortement associés à la récurrence de la maltraitance; 2) des outils statistiques de type actuariel, qui font appel à des méthodes statistiques permettant de déterminer et de pondérer les principaux facteurs associés à l'occurrence et à la récurrence de sévices corporels infligés aux enfants (résultats d'intérêt).

Question 1 - Efficacité théorique des outils (fiabilité, validité)

En réponse à la question sur l'efficacité théorique des sept outils évalués, les résultats obtenus indiquent qu'aucun de ces outils ne présente à la fois une bonne fiabilité et une bonne validité. Certains d'entre eux (CAPI, FRAAN) se démarquent sur le plan de la fiabilité. Aussi, les résultats indiquent que les outils de type contextuel (CAPI, AAPI-2) ont une faible validité pour prédire un abus futur (tous types confondus). Les outils de type actuariel sont plus efficaces pour évaluer le niveau de risque d'abus ou sa récurrence. Parmi ces outils, le CFRA et, dans une moindre mesure, l'OFRA – qui est sa version adaptée au contexte ontarien – présentent une validité adéquate (efficacité moyenne) en vertu des critères scientifiques reconnus.

Questions 2 et 3 - Efficacité pratique des outils (utilité, applicabilité)

La revue de la littérature grise a permis de repérer neuf publications dont les auteurs soulèvent des questions sur l'utilité et l'applicabilité des outils d'évaluation du risque d'abus physique dans les services de protection de la jeunesse (SPJ).

Les atouts associés à ce type d'outil sont les suivants : l'information additionnelle obtenue à l'aide de l'outil, la fiabilité de l'évaluation, le soutien apporté par l'outil à la prise de décision, la possibilité de déterminer un ordre de priorité des services à offrir selon le niveau de risque que présentent les familles évaluées. À l'inverse, les limites associées à ce type d'outil concernent son usage obligatoire, le surclassement des familles à risque, la perception selon laquelle l'outil porte atteinte au jugement professionnel de l'intervenant, l'absence de considération des conditions de vie et du contexte socioéconomique des familles évaluées, l'absence de prise en compte des ressources et des forces dont disposent ces familles, et les facteurs de protection de celles-ci contre le risque d'abus physique. En conséquence, il en résulterait parfois un détournement de l'usage, une sous-utilisation ou encore un usage différencié de l'outil, selon ses utilisateurs et les établissements.

Quant à l'applicabilité (conditions d'implantation) de ce type d'outil, les principales préoccupations soulevées dans ces publications sont les suivantes : une formation adéquate du personnel, le soutien des gestionnaires, une définition claire de la finalité de l'outil et de

ses usages, l'accès à des services qui répondent aux besoins des familles évaluées à l'aide de l'outil.

Afin de contextualiser les résultats issus de la littérature, 4 groupes de discussion focalisée ont été formés; 21 personnes (intervenants, agents de liaison, gestionnaires) en provenance de 2 CSSS et de 2 CJ du Québec y ont participé. De façon générale, les propos de ces personnes au sujet des atouts et des limites de l'utilité de ce type d'outil ainsi que sur ses conditions d'implantation vont dans le même sens que ceux rapportés dans la littérature. Cependant, en raison du contexte québécois, la coordination des évaluations de la clientèle et l'objectif d'assurer une continuité entre l'offre de services de première ligne et l'offre de services de deuxième ligne ont marqué les discussions. À cet égard, les liens à établir (partenariat, entente) entre ces différents niveaux de services et avec leurs partenaires pour assurer des conditions favorables à l'intégration d'un tel outil ont été soulevés. L'importance d'éviter un chevauchement avec les outils déjà utilisés, ou une répétition inutile de ceux-ci a été mentionnée.

Question 4 – Adaptabilité au contexte légal québécois

L'exercice de délibération engagé entre les membres du comité de suivi leur a permis de se prononcer sur la pertinence et la possibilité d'adapter un outil actuariel, le CFRA (dont la version adaptée au contexte ontarien et en usage en Ontario est l'OFRA), au contexte légal québécois, moyennant certaines conditions énumérées dans les recommandations.

Conclusions

Aucun des outils évalués ne satisfait à la fois aux critères d'une bonne fiabilité et d'une bonne validité (efficacité théorique). Quand il fait partie d'un système d'évaluation intégré ou SDM (*structured decision making*), l'outil actuariel fournit une meilleure évaluation du risque d'abus physique que lorsqu'il est utilisé seul. Aussi, les acteurs sur le terrain consultés ont manifesté un intérêt à l'égard d'un outil d'évaluation du risque d'abus physique. Selon eux, l'utilité d'un tel outil réside dans l'harmonisation des pratiques d'évaluation qu'il introduit (standardisation du processus d'intervention, langage et critères d'évaluation communs, meilleure évaluation de la priorité et de l'intensité des services à offrir à la clientèle). Ils voient également l'applicabilité de ce type d'outil soumis à certaines conditions : l'étape du processus d'intervention à laquelle il doit être utilisé, son caractère distinct par rapport aux outils qui sont déjà utilisés, sa capacité à prendre en compte les caractéristiques du client et l'évolution de sa situation, une formation commune à tout le personnel, des lignes directrices claires sur ses procédures et ses usages de même qu'une entente inter établissements sur le choix, les orientations et les résultats attendus de cet outil.

Recommandations

À la lumière de ces résultats, l'INESSS adresse les trois recommandations suivantes aux centres jeunesse du Québec (regroupés sous la direction « Programme jeunesse » à l'intérieur des CISSS et des CIUSSS depuis le 1^{er} avril 2015) :

- 1) Intégrer un outil actuariel dans le processus d'évaluation du risque d'abus physique.
- 2) Considérer et documenter certains éléments avant de songer à l'implantation d'un outil de ce type. Par exemple, les facteurs de risque associés aux caractéristiques de la population québécoise, la pertinence et la faisabilité d'adapter l'outil CFRA au contexte

québécois, le degré de correspondance entre le Système de soutien à la pratique (SSP) et le système d'évaluation intégré, soit le SDM.

- 3) Dans l'éventualité du déploiement de l'outil suggéré, s'assurer que certaines conditions d'implantation avantageuses soient en place dans les milieux d'intervention : définition claire des procédures et des usages de ce type d'outil au moyen de lignes directrices, offre de formation appropriée, adaptation des résultats obtenus à l'aide de l'outil aux politiques et aux ressources disponibles des centres jeunesse afin de pouvoir répondre à la demande de services.

SUMMARY

Effectiveness of child maltreatment risk assessment tools

Background and mandate

In 2007, the introduction in Québec of the concept of *serious risk* concerning neglect, sexual abuse and physical abuse in the Youth Protection Act reflected the growing recurrence of these types of maltreatment. This is evidenced in child protective services (CPS), for example, by new child abuse reports, case reassessments, new substantiated allegations, and repeated child placements in foster homes. The cyclical path between primary care services (health and social service centres, CSSS) and secondary care services (youth centres) destined for children who have experienced or who are likely to experience one or more situations of physical abuse reflects this phenomenon that concerns the authorities responsible for these services. With a view to improving assessment practices, promoting better risk management and improving the identification of clients in order to offer appropriate services based on available resources, the Association des centres jeunesse du Québec (ACJQ) asked the Institut national d'excellence en santé et en services sociaux (INESSS) to identify one or more child maltreatment risk assessment tools.

By definition, a child maltreatment risk assessment tool is composed of a series of questions serving to assess, with greater accuracy and consistency than professional judgment alone, children at risk for bodily injury potentially endangering their health and safety. Such a tool should meet scientific criteria both for efficacy (reliability, validity) and for effectiveness (utility and applicability in intervention settings). INESSS, in collaboration with the Centre jeunesse de Québec – Institut universitaire (CJQ–IU), therefore produced this evidence brief, which is designed to answer the following four questions:

- 1) What is the recognized scientific efficacy of child maltreatment risk assessment tools?
- 2) Are the tools with recognized efficacy useful in decision making during the assessment of children who are the subject of initial or subsequent child abuse reports?
- 3) How can these tools with recognized efficacy be incorporated in the context of intervention?
- 4) Are the tools with recognized efficacy adaptable to the Québec legal context?

Methodology

A systematic review of the literature spanning from 2003 to 2013 was performed to respond to the first question. The second and third questions were addressed through review of the grey literature and a focus group of 21 participants, including practitioners, liaison agents and managers employed in CSSS and youth centres in Québec. The fourth question was discussed by the members of the project steering committee and commented on by independent experts.

Results

The systematic review identified 21 studies with results on the efficacy of 7 child maltreatment risk assessment tools. Used during evaluation of the child's situation, these tools are the following: AAPI-2 (Adult Adolescent Parenting Inventory-2), C-CAPS (Cleveland

Child Abuse Potential Scale), CAPI (Child Abuse Potential Inventory), CFRA (California Family Risk Assessment), CFRAT (Colorado Family Risk Assessment Tool), FRAAN (Family Risk Assessment of Abuse and Neglect) and OFRA (Ontario Family Risk Assessment). It should be noted that most of the identified studies did not address only the risk of physical abuse and its recurrence since prediction encompasses abuse in general and sometimes neglect.

Two types of assessment tools are used: (1) consensus-based structured risk assessment tools derived from the literature and expert opinion on the factors and conditions considered to be most highly associated with the recurrence of maltreatment; (2) statistical actuarial tools used to determine and weigh the main factors associated with the occurrence and recurrence of bodily injury inflicted upon children (outcomes of interest).

Question 1: Efficacy of tools (reliability, validity)

Regarding efficacy of the seven tools assessed, results indicated that none of these tools have both good reliability and good validity. Some of them (CAPI, FRAAN) stand out in terms of reliability. The results also indicated that consensus-based tools (CAPI, AAPI-2) have low validity for predicting future abuse (all types combined). Actuarial tools are more effective for assessing the level of risk of abuse or its recurrence. Of these tools, the CFRA and, to a lesser extent, the OFRA (i.e., the version adapted to the Ontario context) have adequate validity (average efficacy) according to recognized scientific criteria.

Questions 2 and 3: Effectiveness of the tools (utility, applicability)

The review of the grey literature identified nine publications in which the authors raised questions about the utility and applicability of child maltreatment risk assessment tools in child protective services (CPS).

The advantages associated with this type of tool are the following: the additional information obtained, the reliability of the assessment, the decision support provided, the possibility of determining the order of priority of the services to be offered according to the level of risk that the families present. Conversely, the limitations of this type of tool relate to its compulsory use, the over-classification of at-risk families, the perception that the tool undermines practitioners' professional judgment, the lack of consideration given to the living conditions and socio-economic status of the assessed families, the lack of consideration given to family resources, strengths, and protective factors against the risk of physical abuse. This may consequently lead to misuse, underuse or inconsistent use of the tool, depending on the practitioners and institutions.

Regarding the applicability (implementation conditions) of this type of tool, the main concerns raised in the publications were the following: appropriate staff training, support provided by management, clear definition of the purpose and uses of the tool, and access to services that meet the needs of the families assessed with this tool.

In order to contextualize the results found in the literature, 4 focus groups were formed, including 21 persons (practitioners, liaison officers, managers) employed by 2 CSSS and 2 youth centres in Québec. In general, the comments on the advantages and limitations of this type of tool and on its implementation conditions were consistent with those reported in the literature. However, topics that permeated the discussions, owing to the Québec context, were the coordination of client assessments and the need to ensure continuity between primary care and secondary care services. In this regard, participants raised the issue of the

links to be established (partnerships, agreements) between different service levels and with partners to facilitate use of such a tool. The importance of preventing overlap with the tools already in use, or unnecessary duplication, was mentioned.

Question 4: Adaptability to Québec's legal context

The deliberative exercise allowed the members of the steering committee to comment on the appropriateness and possibility of adapting an actuarial tool, the CFRA (of which the OFRA is the Ontario version), to Québec's legal context, subject to certain conditions listed in the recommendations.

Conclusions

None of the tools assessed meet the criteria of both good reliability and good validity (efficacy). When actuarial tools are part of an integrated assessment system, or structured decision making (SDM), they provide better assessment of the risk of maltreatment than when used alone. In addition, the consulted practitioners expressed interest in a maltreatment risk assessment tool. According to them, the utility of this type of tool lies in the fact that it harmonizes the assessment practices that it introduces (standardization of the intervention process, common terminology and assessment criteria, better assessment of the priority and intensity of services). They also see the applicability of this type of tool as being subject to certain conditions: the stage of the intervention process at which it must be used, its distinctiveness in relation to the tools already in use, its capacity to take into account client characteristics and the progress in their situations, common training offered to all staff, clear guidelines on its procedures and uses, and an inter-institutional agreement on the choice, directions and outcomes expected from this tool.

Recommendations

In light of these results, INESSS makes the following three recommendations to youth centres in Québec (grouped under the youth program division ["Programme jeunesse"] within the various CISSS [integrated health and social service centres] and CIUSSS [integrated university health and social service centres] as of April 1, 2015):

- 1) Incorporate an actuarial tool into the process for assessing the risk of physical abuse.
- 2) Consider and document certain elements before contemplating use of this type of tool, such as the risk factors associated with characteristics of the Québec population, the appropriateness and feasibility of adapting the CFRA tool to Québec, the level of correspondence between the "Système de soutien à la pratique" (SSP) and structured decision making (SDM).
- 3) In the event that the suggested tool is used, ensure that certain positive implementation conditions are in place in the intervention settings: clear definition of the procedures and uses of this type of tool through practice guidelines; appropriate training; adaptation of the results obtained with the tool to youth centre policies and to available resources in order to meet service demand.

SIGLES ET ABRÉVIATIONS

AAPI-2	Adult Adolescent Parenting Inventory-2
ACJQ	Association des centres jeunesse du Québec
AETMIS	Agence d'évaluation des technologies et des modes d'intervention en santé
AMSTAR	Assessment of Multiple Systematic Reviews
AQESSS	Association québécoise d'établissements de santé et de services sociaux
C-CAPS	Cleveland Child Abuse Potential Scale
CAP	Child Abuse Potential Inventory
CFRA	California Family Risk Assessment
CFRAT	Colorado Family Risk Assessment Tool
CJ	Centres jeunesse
CJQC-IU	Centre jeunesse de Québec – Institut universitaire
CRC	Children's Research Center
CSSS	Centre de santé et de services sociaux
DIFR	Dispersion Index for Risk
FRAAN	Family Risk Assessment of Abuse and Neglect
INESSS	Institut national d'excellence en santé et en services sociaux
LPJ	Loi sur la protection de la jeunesse
MSSS	Ministère de la Santé et des Services sociaux
NHMRC	National Health and Medical Research Council
OFRA	Ontario Family Risk Assessment
PSEP	Prognostic Separation Index
SDM	Structured Decision Making
SIPPE	Services intégrés en périnatalité et pour la petite enfance
SPJ	Services de protection de la jeunesse
SSP	Système de soutien à la pratique

GLOSSAIRE

Abus physique

L'abus physique chez l'enfant correspond à un acte de violence physique commis à son endroit par un proche de celui-ci, souvent un adulte (père, mère ou conjoint) dans le contexte d'une discipline abusive (punition ou châtiment corporel) ou encore par l'irruption d'une brutalité irréfléchie. Par exemple, secouer, pousser, attraper, projeter, pousser ou tirer un enfant, secouer un bébé, frapper avec la main, donner un coup de poing, un coup de pied ou mordre, frapper avec un objet ou lancer un objet sur l'enfant, etc. (Trocmé et coll., 2010 : 69-70).

Au Québec, la Loi sur la protection de la jeunesse définit l'abus physique comme suit [LPJ, chapitre 34.1, article 38e, juillet 2007] :

« Lorsque l'enfant subit des sévices corporels ou est soumis à des méthodes éducatives déraisonnables de la part de ses parents ou de la part d'une autre personne et que ses parents ne prennent pas les moyens nécessaires pour mettre fin à cette situation. »

Aire (surface) sous la courbe ROC

L'aire sous la courbe ROC (Receiver Operating Characteristic) représente la sensibilité d'un test en fonction de 1-spécificité du test lorsque l'on fait varier le seuil de positivité. Si le seuil est haut, il y a peu de faux positifs, donc une spécificité élevée. S'il y a beaucoup de faux négatifs : la sensibilité est faible. Si le seuil est bas, il y a beaucoup de faux positifs, donc une spécificité faible. S'il y a peu de faux négatifs, la sensibilité est élevée. Dans le cas d'un test parfait, la sensibilité = 100 % et la spécificité = 100 %.

Efficacité pratique

Bénéfice que procure l'usage d'une technologie, d'un programme ou d'une intervention pour traiter un problème particulier dans des conditions générales ou courantes plutôt que dans des conditions contrôlées, notamment dans le cadre de l'utilisation d'une technologie par un médecin dans un hôpital ou par un patient à domicile¹.

Efficacité théorique

Bénéfice que procure l'usage d'une technologie, d'un programme ou d'une intervention pour traiter un problème particulier dans des conditions idéales, notamment dans le cadre d'une recherche en laboratoire ou d'un protocole d'essai clinique randomisé rigoureux².

¹ Institut national d'excellence en santé et en services sociaux (INESSS). Glossaire en ETS [site Web]. Disponible à : <http://htaglossary.net/accueil>.

² *Ibid.*

Évaluation sur le terrain

Étude conçue expressément pour recueillir des données originales sur un mode d'intervention prometteur afin de vérifier son efficacité pratique lorsqu'on ne dispose pas de données probantes suffisantes pour prendre une décision éclairée [INESSS, 2015 : vii].

Exactitude

Pour déterminer la validité d'un test, on le compare avec un test de référence dont on est sûr qu'il mesure l'effet visé. On dira d'un test hautement sensible et spécifique qu'il est très exact. Il sera considéré aussi comme très performant [AETMIS, 2008 : 10].

Fiabilité

La fiabilité d'un instrument de mesure est sa capacité à reproduire un résultat de façon constante dans le temps et dans l'espace lorsqu'il est utilisé correctement par des observateurs différents.

Trois approches permettent d'évaluer la fiabilité d'un instrument :

- « la comparaison des résultats obtenus par l'utilisation d'un même instrument à différents moments pour évaluer sa stabilité (fiabilité test-retest);
- l'appréciation de l'équivalence des résultats obtenus quand un même phénomène est mesuré par plusieurs observateurs à un même moment (fiabilité interjuges);
- l'appréciation de la fiabilité d'un instrument composé de plusieurs éléments ou indicateurs, en mesurant l'homogénéité de ses composantes [validité interne] » [Contandriopoulos *et al.*, 1990 : 75].

Groupe de discussion focalisée

« Une méthode d'enquête qualitative [...] qui consiste à rassembler un petit groupe d'individus homogène appartenant à une population précise, afin de provoquer des échanges et des discussions autour de thèmes prédéfinis » [Desrosiers et Larivière, 2014 : 258].

Littérature grise

Documents publiés pour un public restreint, en dehors des grands circuits de distribution, et difficilement repérables dans les bases de données courantes, tels que des présentations à des congrès, des évaluations de technologies de la santé réalisées par des hôpitaux, certains documents gouvernementaux, etc. Note : La définition de la littérature grise ne fait pas l'unanimité, car la frontière qui la délimite est mouvante et évolue avec les nouveaux moyens techniques³.

³ Institut national d'excellence en santé et en services sociaux (INESSS). Glossaire en ETS [site Web]. Disponible à : <http://htaglossary.net/accueil>.

Outil d'évaluation structurée du risque d'abus physique

Instrument composé d'une série de questions servant à déterminer, avec plus de précision et de cohérence que le seul jugement professionnel, les enfants à risque de subir des sévices corporels pouvant compromettre leur santé et leur sécurité [Knoke et Trocmé, 2004].

Parties prenantes

Personnes ou groupes qui sont potentiellement ou effectivement concernés par l'élaboration d'un avis ou d'un guide de pratique [INESSS, 2015 : vii].

Récurrence de l'abus physique

Probabilité de récurrence de ce phénomène selon une période de temps plus ou moins étendue.

Revue narrative

Synthèse d'études originales qui n'ont pas été répertoriées ou analysées de façon systématique, c'est-à-dire standardisées et objectives⁴.

Revue systématique

Forme de recension structurée des publications portant sur une question formulée de façon à ce qu'on puisse y répondre en analysant les articles qui s'y rapportent⁵.

Risque d'abus physique

Selon la Loi sur la protection de la jeunesse du Québec (LPJ)⁶, il y a risque d'abus physique « lorsque l'enfant encourt un risque sérieux de subir des sévices corporels ou d'être soumis à des méthodes éducatives déraisonnables de la part de ses parents ou de la part d'une autre personne et que ses parents ne prennent pas les moyens nécessaires pour mettre fin à cette situation ».

La notion de risque sérieux d'abus physique « fait référence à la forte probabilité que des abus physiques se produisent. Les situations de compromission basées sur le risque sérieux doivent être démontrées selon le même fardeau de preuve que les autres situations de compromission, c'est-à-dire par prépondérance de preuve. L'existence actuelle ou immédiate du danger n'est pas en cause, mais plutôt son caractère futur et probable » [MSSS, 2010 : 389].

Seuil de signification statistique

Valeur basale déterminée (arbitrairement) pour le résultat d'un test. Si le résultat du test est égal ou plus élevé que la valeur seuil, le test est alors considéré comme positif. [Van Driel, 2004].

⁴ *Ibid.*

⁵ *Ibid.*

⁶ Loi sur la protection de la jeunesse, L.R.Q., c. P-34.1, art. 38e.

Spécificité (Sp), Sensibilité (Se) (voir Validité prédictive, Aire sous la courbe ROC)

Validité

« Caractéristique du résultat d'un test correspondant à sa capacité d'apprécier correctement un phénomène ou sa valeur réelle [Fletcher et al., 1988b] [...] la validité indique le degré auquel un examen ou un instrument de mesure fournit une évaluation correcte de ce qu'il est censé mesurer [Van Driel, 2004] » [AETMIS, 2008 : 10].

Validité de construit (ou conceptuelle)

La validité de construit tente de valider le corps théorique sous-jacent à la mesure et de vérifier des hypothèses d'associations. Le point central de la validité de construit repose sur le concept abstrait qui est mesuré et sa relation avec d'autres concepts. C'est aussi l'analyse de la signification des scores d'un test par rapport au concept mesuré.

Deux principales approches permettent d'évaluer la validité de construit, soit :

- 1) Validité de convergence : différentes échelles de mesure du même construit sont appliquées à des sujets. Les analyses de corrélation effectuées sur ces mesures doivent produire des résultats similaires;
- 2) Validité de différenciation : des échelles de mesure reliées au concept d'intérêt sont appliquées à des sujets; on évalue la capacité de l'instrument de mesure à différencier le construit mesuré des autres qui lui sont similaires [Fortin, 1994 : 61].

Validité de critère (ou critérielle)

« Elle est évaluée en examinant la corrélation, ou autre relation statistique, entre la performance au test et un critère externe de performance. En d'autres mots, les individus qui présentent un score élevé au test tendent à être plus performants sur ce critère externe que les individus qui ont un score faible au test. Si le critère est obtenu en même temps que la passation du test, on parle de validité concourante ou concomitante. La validité concourante (ou concomitante) d'un test se manifeste à travers ses corrélations avec d'autres tests semblables. Si le critère est obtenu plus tard, on parle de validité prédictive »⁷.

Validité externe

Capacité d'un plan de recherche à fournir des résultats généralisables à d'autres populations, à d'autres contextes et à d'autres périodes⁸.

Validité interne

Capacité d'un plan de recherche à représenter la véritable relation de cause à effet entre une intervention et un effet dans les circonstances particulières de cette recherche⁹.

⁷ Central Test. Validation psychométrique. Paris, France : Central Test; 2011. Disponible à : http://www.centraltest.com/ct_fr/upload/documents/ValidationPsychometrique.pdf.

⁸ Institut national d'excellence en santé et en services sociaux (INESSS). Glossaire en ETS [site Web]. Disponible à : <http://htaglossary.net/accueil>.

⁹ *Ibid.*

Validité prédictive

Ce terme désigne l'aptitude d'un dispositif d'évaluation ou de mesure (conçu par exemple dans une perspective d'évaluation pronostique) à prédire de manière aussi précise que possible le niveau de réussite, d'adaptation, d'intégration, etc., d'un ensemble d'individus par rapport à des situations qu'ils rencontreront (ou pourraient rencontrer) ultérieurement¹⁰.

¹⁰ Institut de recherche et de documentation pédagogique (IRDp). Valeur prédictive (d'un dispositif d'évaluation ou de mesure) [site Web]. Disponible à : http://www.irdp.ch/edumetrie/lexique/validite_predictive.htm.

INTRODUCTION

Au Québec, la notion de risque sérieux en matière de négligence, d'abus sexuel et d'abus physique fait partie intégrante de la Loi sur la protection de la jeunesse (LPJ), dans sa version révisée de 2007. Depuis ce temps, les données tirées des bilans des directeurs de la protection de la jeunesse indiquent une croissance faible, mais continue du nombre de signalements retenus pour cause d'abus physique ou de risques sérieux d'abus physique. Par exemple, le bilan 2014 des directeurs de la protection de la jeunesse fait état de 7 520 signalements retenus d'abus physique et de 2 308 signalements retenus de risque sérieux d'abus physique envers des enfants de 0 à 17 ans; cela totalise 29,6 % des signalements retenus en 2013-2014, tous types confondus [ACJQ, 2014]. Par ailleurs, la récurrence de la maltraitance (abus de tous types et négligence) envers des enfants de ce groupe d'âge atteint 36 % en ce qui concerne les allégations fondées de cas qui ont été signalés à nouveau au cours des 5 années suivant la fin de l'intervention par les services de protection de la jeunesse (SPJ) [Hélie et Laurier, 2012].

Dans le domaine de la protection de la jeunesse, le personnel qui intervient auprès des familles et de leurs enfants d'âge mineur doit considérer le risque de maltraitance future de l'enfant au moment de prendre une décision sur le type et l'intensité des services à offrir ou non à la clientèle visée par la LPJ. Or, c'est souvent dans un contexte marqué par l'urgence de la situation vécue par l'enfant que cette décision est prise. En l'absence d'un outil validé sur le plan scientifique, cette décision dépend, pour l'essentiel, du jugement de l'intervenant et suivant la collecte d'une quantité appréciable de renseignements de sources variées (grilles d'évaluation, conseils de pairs, avis d'experts, etc.), qu'il obtient sur la famille concernée. Dans ce contexte, toute erreur résultant de cette prise de décision s'avère hasardeuse en raison de la gravité des conséquences qui s'ensuivent pour les enfants (récidive de l'abus physique, atteinte à la santé et au bien-être de l'enfant maltraité, stigmatisation de la famille, etc.).

Plusieurs auteurs ont effectivement relevé des erreurs dans la prise de décision des intervenants concernés [Barlow *et al.*, 2012; Baumann *et al.*, 2011; Price-Robertson et Bromfield, 2011; Guay, 2006; White et Walsh, 2006; D'Andrade *et al.*, 2005]. L'approche utilisée par ces derniers pour évaluer l'abus physique et sa récurrence est ici pointée du doigt. Dans la littérature consultée, cette approche est classée dans les trois grandes catégories suivantes [Rycus et Hughes, 2003] :

- 1) Le **jugement professionnel** s'appuie sur l'intuition, la formation et l'expérience de l'intervenant. Plusieurs études font état de la faible valeur prédictive du jugement professionnel, parfois entremêlé de diverses approches heuristiques (par exemple, une confiance exagérée dans l'intuition ou le jugement personnel, une association induite entre un cas présent et un cas antérieur, se faire rapidement une idée du client et qu'on ne cherche qu'à confirmer par la suite) sur lesquelles s'appuierait l'intervenant pour prendre une décision sur la protection de l'enfant;
- 2) Les outils d'évaluation structurée du risque de :
 - **type contextuel** s'appuient sur le consensus (issu de la littérature et des conseils d'experts) sur les facteurs et les conditions jugés le plus fortement associés à la récurrence de la maltraitance. L'interaction entre ces facteurs, leur pondération et les résultats obtenus ne font pas l'objet d'études empiriques. Le recours au jugement

professionnel est de mise. Le grand nombre de faux positifs (familles à faible risque de récidive dont le signalement est indûment retenu) et de faux négatifs (familles à risque élevé de récidive dont le signalement n'est pas retenu) obtenus à l'aide de ces outils a contribué à l'essor d'un troisième type d'outils : les outils de type actuariel;

- **type actuariel** sont élaborés à partir de méthodes statistiques permettant de déterminer et de pondérer les principaux facteurs associés à l'occurrence et à la récurrence de mauvais traitements physiques infligés aux enfants (résultats d'intérêt).

Les centres jeunesse du Québec disposent du Système de soutien à la pratique (SSP) dont l'usage est devenu obligatoire au cours des années 2000 [Frenette *et al.*, 2002]. Ce système vise à normaliser l'application pratique du concept de « protection » découlant de la LPJ concernant les jeunes d'âge mineur. En cela, ses principaux modules sont liés aux différentes étapes d'application de cette loi : réception et traitement du signalement, protection immédiate, évaluation de la situation, orientation (choix des mesures de protection). Ce système expert d'aide à la décision constitue un moyen de systématiser la démarche clinique des intervenants. Élaboré selon un modèle d'intelligence artificielle, il se présente sous la forme d'une arborescence exploitant ou non diverses avenues (ou branches) du questionnaire informatisé, selon les réponses fournies par l'intervenant (approche déductive). Ainsi qu'il est précisé dans un document produit par la Sogique (Direction générale des technologies de l'information du MSSS), « les éléments étudiés ne sont pas présentés comme des indicateurs ayant une valeur prédictive, mais comme des données organisées avec d'autres dans une perspective d'établir un argumentaire »¹¹. Le SSP ne contient pas d'outil actuariel d'évaluation du risque d'abus physique. De plus, il n'a jamais fait l'objet d'une validation scientifique.

Dans ce contexte, disposer d'un outil d'évaluation du risque d'abus physique qui soit à la fois standardisé, fiable et valide permettrait d'appuyer la prise de décision et de limiter les risques d'occurrence et de récidive d'abus physique envers les enfants. La demande présentée à l'INESSS par l'ACJQ vise justement à recenser ces outils et à évaluer leur efficacité. Une entente de coproduction a été établie entre l'INESSS et le CJQ-IU à cette intention. La question décisionnelle à laquelle cet avis doit répondre est la suivante : Existe-t-il des outils dont l'efficacité est reconnue, sur le plan scientifique, pour évaluer le risque d'abus physique envers des enfants, et sa récurrence? Le cas échéant, peut-on adapter ces outils au contexte légal québécois? L'utilité et l'applicabilité de ces outils que soulève leur intégration possible dans les milieux d'intervention ont également été considérées.

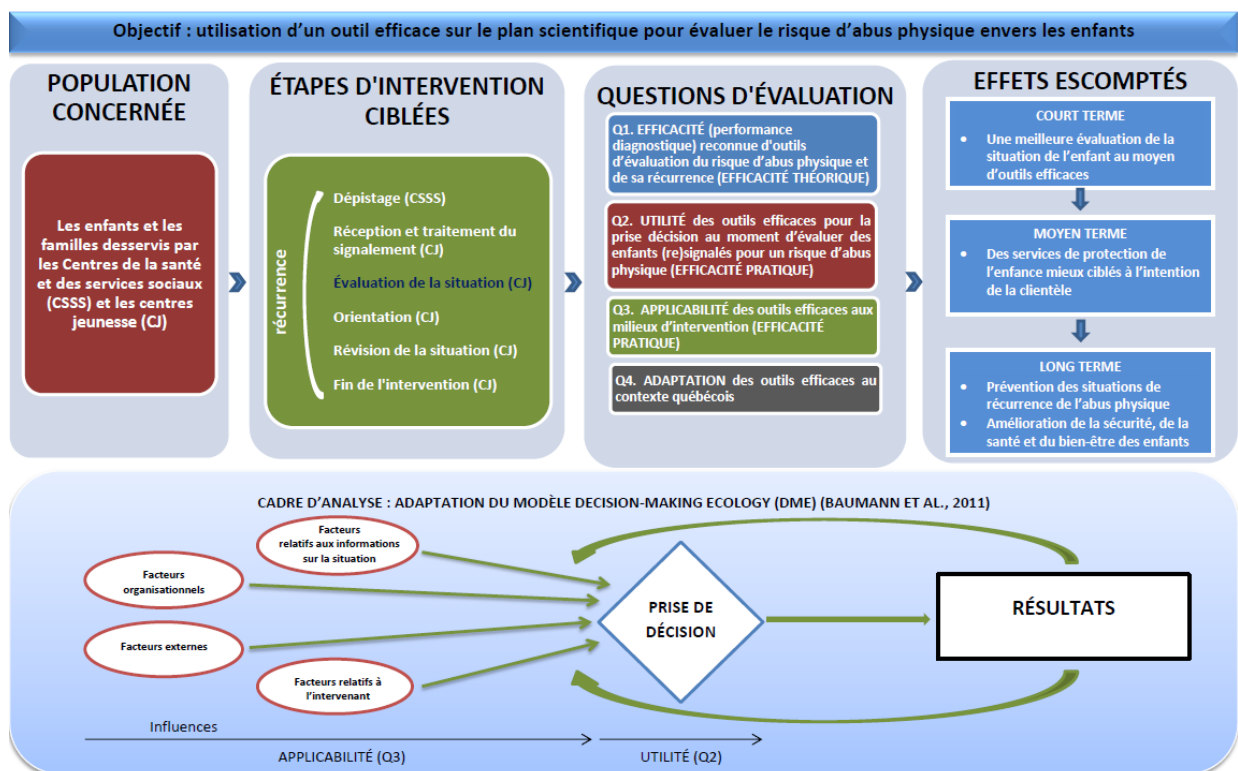
¹¹ Dufault D. Le système de soutien à la pratique (SSP). Outils cliniques et SSP. 2006 [page 5, présentation PowerPoint (document interne)].

1 MÉTHODOLOGIE

1.1 Modèle logique et questions clés d'évaluation

Le modèle logique présenté dans la figure 1 illustre le contexte général dans lequel s'effectue la prise de décision suivant l'évaluation du risque d'abus physique envers des enfants québécois ainsi que les questions clés d'évaluation.

Figure 1 Modèle logique



Ce modèle logique illustre la démarche d'ensemble proposée. Dans le contexte québécois, la population concernée est composée des enfants, et de leurs familles, desservis par les établissements de première ligne (CSSS) et de deuxième ligne (CJ) en vertu de la Loi sur la santé et les services sociaux (LSSS) et de la Loi sur la protection de la jeunesse (LPJ). Les différentes interventions offertes par ces établissements sont mentionnées ainsi que la récurrence possible (parcours cyclique) des enfants à risque d'être à nouveau victimes d'un acte d'abus physique. L'étape qui correspond à l'« Évaluation de la situation » est celle à laquelle ce type d'outil a été conçu initialement, d'après la littérature consultée.

Ce modèle logique intègre également le cadre d'analyse proposé par Baumann et ses collaborateurs [2011], qui présente le contexte général de la prise de décision. Celle-ci s'effectue dans un contexte d'incertitude, caractérisé par un ensemble de facteurs qui exercent une influence sur elle et, par conséquent, sur les mesures adoptées concernant l'enfant et sa famille. Ces facteurs sont les suivants :

- facteurs relatifs aux informations sur la situation : caractéristiques de l'enfant et des parents, dynamique familiale, sécurité de l'enfant, type et antécédents de mauvais traitements, etc.
- facteurs externes : lois, politiques gouvernementales, cas médiatisés, etc.
- facteurs associés aux caractéristiques de l'intervenant qui évalue les cas : compétences, expérience, formation, valeurs personnelles;
- facteurs organisationnels : ressources disponibles, politiques internes des établissements, contraintes de temps, conditions de travail du personnel concerné, etc.

Dans un contexte d'incertitude, la prise de décision est sujette au risque d'erreur. Cela peut se traduire, par exemple, par une offre de services à des familles à faible risque de récurrence d'un abus physique envers leurs enfants (faux positif). À l'inverse, cela peut se traduire par l'absence de services offerts à des familles à risque élevé de récurrence (faux négatif). L'efficacité recherchée par ces outils consiste à réduire le plus possible le risque d'erreur dans la prise de décision. Il s'agit d'arriver à prédire, avec la plus grande précision possible, les cas probables d'occurrence ou de récurrence d'abus physique envers les enfants.

Les effets escomptés par l'utilisation d'outils fiables et valides pour évaluer le risque d'abus physique envers les enfants sont les suivants :

- à court terme : une meilleure évaluation de la situation de l'enfant au moyen d'outils efficaces.
- à moyen terme : des services de protection de la jeunesse (SPJ) mieux ciblés à l'intention de la clientèle.
- à long terme : la prévention des situations de récurrence de l'abus physique à l'égard des enfants et l'amélioration de leur sécurité, de leur santé et de leur bien-être.

Les questions clés d'évaluation et la démarche entreprise pour y répondre se présentent comme suit.

1.2 Question 1 – Efficacité théorique (revue systématique)

Quelle est l'efficacité reconnue, sur le plan scientifique, des outils d'évaluation du risque d'abus physique envers un enfant?

1.2.1 Revue systématique

Une revue systématique a été réalisée afin de répondre à la question 1. La méthodologie utilisée pour mener cette revue systématique respecte les normes de production des revues systématiques de l'INESSS [2013].

1.2.2 Stratégie de recherche de l'information

Une revue systématique des publications scientifiques originales et des synthèses (méta-analyses, revues systématiques, rapports d'évaluation) qui évaluent l'efficacité théorique (fiabilité et validité) des outils d'évaluation du risque d'abus physique à l'égard des enfants de 0 à 17 ans ont été réalisées.

La stratégie de recherche d'information a été élaborée en collaboration avec deux spécialistes en information scientifique. Cette stratégie est présentée en détail dans l'annexe A. Les bases de données MEDLINE (PubMed), CINAHL, PsycINFO, Embase, Éric, Érudit et Repère ont été interrogées. Les principaux mots clés utilisés pour décrire le mode d'intervention et la population sont les suivants : *child abuse, child maltreatment, child protection, risk, assessment, screening, tool, actuarial, consensus-based, efficacy, effectiveness, predictive, reliability, validity, sensibility, specificity, test, measure, scale, re-reported, re-substantiated, recurrence*. La recherche documentaire couvrait les publications parues de janvier 2003 à décembre 2013, de langue anglaise ou française (avec une indexation des mots clés en anglais). Une mise à jour de la recherche documentaire a été réalisée en octobre 2014. Aucune limite relative au type de publication n'a été définie en vue de ces recherches.

1.2.2.1 Critères de sélection des études scientifiques

Les publications répondant aux critères suivants ont été retenues aux fins de la revue systématique :

- population cible : Les jeunes âgés de 0 à 17 ans et ayant fait l'objet d'au moins 1 signalement (fondé ou non) par les SPJ.
- mode d'intervention : L'application d'un outil d'évaluation du risque d'abus physique envers un enfant.
- comparateur : En l'absence d'un modèle de référence, l'efficacité de l'outil est comparée à celle d'outils ayant la même finalité et dont traitent les publications retenues.
- résultats d'intérêt : L'efficacité théorique des outils est considérée à partir des mesures de fiabilité¹² et de validité en rapport avec les résultats obtenus sur le risque d'abus physique futur (réurrence).
- durée de l'intervention et durée du suivi postintervention : Elle est ajustée selon ce qui est mesuré dans les études retenues.
- milieux d'intervention et territoires inclus : Les SPJ en Amérique du Nord, en Europe, en Australie et en Nouvelle-Zélande. Les services de garde éducatifs à l'enfance, les établissements d'enseignement, les services des urgences des centres hospitaliers ainsi que les postes de police sont exclus des milieux d'intervention ciblés.
- étendue de la période des publications retenues : janvier 2003 à décembre 2013, afin de considérer les études de (re)validation des outils recensés.
- langue : française ou anglaise.

¹² Les critères d'une bonne fiabilité concernant la validité interne de l'outil sont établis à partir des mesures suivantes : un alpha (α) de Cronbach $\geq 0,80$; et concernant l'entente interjuges, un coefficient Kappa de Cohen $\geq 0,75$ [Noonan *et al.*, 2013]. La mesure Kuder-Richardson (KR-20) est utilisée pour estimer la validité interne des échelles dichotomiques, c'est-à-dire celles qui ont deux réponses seulement (oui ou non). Le KR-20 varie de 0,00 à 1,0, la valeur la plus élevée indique une meilleure validité interne.

1.2.2.2 Évaluation de la qualité des études

L'évaluation de la qualité méthodologique des études a été réalisée par trois professionnels scientifiques (dont les initiales sont J.D., I.B. et C.P.), qui ont travaillé de façon indépendante. Ces personnes étaient en accord sur 86 % des résultats de cette démarche. En fait, 20 % des études évaluées par un professionnel (J.D.) ont été validées par une autre (I.B.), à l'aide d'une grille d'extraction préétablie. Les divergences d'opinions ont été réglées par consensus. L'extraction des données a ensuite été réalisée par l'un des professionnels scientifiques (J.D.)

L'évaluation de la qualité méthodologique des études quantitatives a été réalisée selon les critères de la grille de Kmet et ses collaborateurs [2004], selon les critères de la grille de Letts et ses collaborateurs [2007] concernant les études qualitatives, et à l'aide des critères de la grille AMSTAR (Assessment of Multiple Systematic Reviews) [Shea *et al.*, 2009; 2007] concernant les revues systématiques (voir l'annexe C). Une catégorisation a été établie pour attribuer un niveau de qualité à chacune des études, comme suit :

- Faible : lorsque les conclusions de l'étude sont remises en question en raison de faiblesses et de limites méthodologiques majeures.
- Moyenne : lorsque les conclusions sont susceptibles d'être influencées par les faiblesses et les limites méthodologiques, mais que celles-ci ne sont pas majeures et ne remettent pas en cause l'ensemble des conclusions.
- Bonne : lorsqu'il est peu probable que les conclusions de l'étude soient influencées par les faiblesses et les limites méthodologiques [INESSS, 2014].

Les études de faible qualité ont été exclues de la sélection finale, la validité et l'utilité des résultats obtenus étant mises en doute en raison d'une information incomplète ou de limites méthodologiques majeures. La qualité des études a été prise en considération lors de l'interprétation des résultats et de la formulation des recommandations.

La preuve scientifique extraite des publications évaluées a été résumée sous la forme d'une synthèse narrative qui s'appuie sur une analyse descriptive des principaux résultats. Les résultats d'intérêt qui portent sur la validité d'outils actuariels établie à partir du lien entre le classement de familles selon différents niveaux de risque et des indicateurs d'abus futur sont présentés sous la forme de tableaux.

1.3 Questions 2 et 3 – Efficacité pratique (revue narrative et consultation)

Question 2 – Ces outils dont l'efficacité est reconnue sont-ils utiles à la prise de décision au moment d'évaluer des enfants ayant fait l'objet d'un signalement ou d'un nouveau signalement concernant un risque d'abus physique?

Question 3 – Comment ces outils dont l'efficacité est reconnue s'intègrent-ils au contexte d'intervention?

Ces deux questions ont fait l'objet :

- d'une revue narrative à partir de la recension de la littérature grise de manière à inclure les études descriptives (rapports d'évaluation, documentation gouvernementale, sites d'organismes de recherche, etc.) qui abordent l'efficacité pratique (utilité, atouts et limites, conditions d'implantation) des outils d'évaluation du risque d'abus physique et;

- de groupes de discussion focalisée auxquels ont participé des intervenants, des agents de liaison et des gestionnaires en provenance de deux milieux d'intervention différents (CSSS, CJ) et qui sont concernés par l'utilisation de ce type d'outils.

1.3.1 Revue narrative

1.3.1.1 Recension de la littérature grise

La revue narrative visant à répondre à ces questions prend appui sur une recherche documentaire fondée sur les mêmes critères que ceux relatifs la question 1, à l'exception du type de publication, pour lequel aucune limite n'a été définie. La recherche d'information relative à cette littérature grise a été faite au moyen des moteurs de recherche Google, Google Scholar, Exalead, Dogpile, Scirus et Social Care Online. Les mots clés utilisés – et associés aux fins de cette recherche, aux mots *risk assessment child physical abuse* – sont : *utility, feasibility, implement*, relevant*, context, condition, decision making, practitioner, caseworker, protection worker, perception, tool, instrument, acceptability, compliance*. Une recherche à l'aide des mêmes mots clés a également été effectuée sur les sites Web des agences, des organismes, des associations et des organismes gouvernementaux, en particulier ceux qui consacrent leurs activités à la protection de la jeunesse ou à la recherche en matière de maltraitance à l'égard des enfants, notamment : Child Welfare Information Gateway, Children and Family Research Center, California Evidence-Based Child Clearinghouse for Child Welfare, Factor-Inwentash Faculty of Social Work Research Commons, McGill Centre for Research on Children and Families, National Council on Crime and Delinquency's Children's Research Center, National Resource Center for Child Protective Services, Portail canadien de la recherche en protection de l'enfance.

Dans la documentation ainsi recensée, la qualité des études descriptives, réalisées selon une approche qualitative ou mixte (quantitative et qualitative), a été évaluée. La qualité des autres types de publications, dont le contenu prend plutôt la forme d'une synthèse narrative, n'a pas été évaluée.

1.3.2 Groupes de discussion focalisée

Le choix du groupe de discussion focalisée s'inscrit dans une démarche d'évaluation sur le terrain qui s'appuie ici sur une collecte de données primaires sur les pratiques réelles, donc de tirer profit de l'expérience d'acteurs sur le terrain. Cette démarche vise à mettre en contexte des résultats de la revue narrative de la littérature sur l'efficacité pratique de ces outils, en les situant par rapport au processus d'évaluation des cas réalisé par les établissements de première ligne et de deuxième ligne concernés par la protection de la jeunesse au Québec. Trois groupes de discussion focalisée, chacun formé de huit à dix personnes en provenance de deux CSSS et de deux CJ, étaient prévus. Les participants ne se rencontraient qu'une seule fois. Le choix des lieux de provenance de ces derniers s'explique par le type de services abordés (services de première ou de deuxième ligne) et par la rapidité et la facilité d'accès à ces milieux d'intervention par l'entremise des membres de l'équipe du CJQ-IU qui coproduit le présent avis. La participation des personnes recrutées par l'entremise des directeurs des services professionnels et de chefs de service des établissements sélectionnés est volontaire. Ces participants sont des intervenants, des agents de liaison et des gestionnaires concernés par l'utilisation d'un outil d'évaluation du risque d'abus physique envers les enfants. Ces

personnes sont recrutées, selon le service concerné, dans l'un ou l'autre de ces deux types d'établissements, et selon la fonction que ces personnes y occupent.

Le contenu des échanges entre les participants a été enregistré à l'aide d'un magnétophone numérique puis retranscrit, sous forme de résumé contenant des extraits mot pour mot, au besoin. Un compte-rendu de ces rencontres a ensuite été envoyé aux participants à des fins de validation; par la suite, une analyse de contenu a été réalisée afin de dégager leurs points de vue, de les comparer et de les synthétiser. Enfin, une correspondance a été établie entre ces points de vue et les préoccupations soulevées dans la littérature sur l'efficacité pratique de ces outils, soit : la nécessité d'évaluer le risque d'abus physique envers les enfants et sa récurrence; les méthodes utilisées à cette fin dans les établissements; la valeur ajoutée ou non de ce type d'outil à la pratique d'évaluation en vigueur; les acteurs qui pourraient bénéficier de ce type d'outil; le moment opportun de son application, les conséquences, les atouts et les limites relatives à l'utilisation de ce type d'outil; les conditions propices à l'intégration de ce type d'outil (voir l'annexe I).

1.3.3 Critères de sélection des études

Concernant la revue de la littérature effectuée en vue de répondre aux questions 2 et 3, les critères de sélection des études sont similaires à ceux relatifs à la question 1 (PICOTS), sauf en ce qui a trait aux résultats d'intérêt. En effet, seuls les résultats portant sur l'utilité et l'applicabilité des outils ont été pris en compte. Le contenu du guide d'entretien pour animer les groupes de discussion focalisée a donc été élaboré à partir des préoccupations soulevées dans la littérature sur l'efficacité pratique de ces outils.

1.3.4 Méthodes d'extraction et d'analyse des données

Les données relatives aux questions 2 (utilité) et 3 (applicabilité) apparaissent sur la même grille d'extraction des données que celle de la question 1. Dans la littérature grise, l'utilité (question 2) de ce type d'outil est considérée comme une valeur ajoutée ou, au contraire, comme une entrave à la pratique d'évaluation menant à la prise de décision. En ce qui concerne l'applicabilité (question 3), elle y est considérée sous l'angle des conditions d'usage et d'application de ce type d'outil dans les milieux d'intervention. Les données ont également été analysées en vue de faire ressortir les atouts et les limites de l'efficacité pratique de ce type d'outil.

1.3.5 Méthodes d'analyse des résultats (procédure)

L'analyse de contenu de l'information recueillie par l'entremise des groupes de discussion focalisée formés d'intervenants, d'agents de liaison et de gestionnaires de CSSS et de CJ a suivi la lecture approfondie de leurs propos. Une catégorisation thématique de ces données a été dégagée. Ensuite, une analyse des données axée sur une comparaison systématique des propos des participants à ces groupes de discussion focalisée a permis de dégager les convergences et les divergences sur les thèmes abordés dans le guide d'entretien. Une analyse intragroupe (entre les intervenants, entre les agents de liaison et entre les gestionnaires) et une analyse intergroupe (les intervenants des CSSS et des CJ, les agents de liaison et les gestionnaires des CJ) ont été effectuées afin d'obtenir une meilleure représentation de l'expérience à la fois propre à chacun de ces groupes et commune à ceux-ci.

1.4 Question 4 – Adaptation au Québec

Afin de répondre à la quatrième question, « Les outils dont l'efficacité est reconnue sont-ils adaptables au contexte légal québécois? », les parties prenantes du projet (membres du comité de suivi, experts indépendants) ont été invitées à se prononcer sur les recommandations qui accompagnent la présentation des résultats et des conclusions du présent avis.

1.4.1 Consultations des parties prenantes

Un comité de suivi, composé de 14 représentants d'organismes gouvernementaux et professionnels, avait pour mandat de soutenir les travaux de l'INESSS sur ce projet. À cette fin, le comité s'est rencontré à quatre reprises. Il a notamment contribué à la définition initiale du projet et a validé les orientations prises en cours de réalisation. Il a également contribué à la formulation des recommandations finales et s'est prononcé sur les produits de transfert de connaissances qui accompagnent la diffusion de l'avis. Cet avis final témoigne de la contribution des membres du comité au projet, mais n'engage nullement leur responsabilité.

De plus, trois experts indépendants ont aussi été consultés en raison de leur expertise scientifique et des connaissances qu'ils possèdent sur la maltraitance des enfants et sur la protection de la jeunesse. Ils ont émis des commentaires sur le plan de réalisation, le rapport d'étape et le rapport préliminaire de ce projet d'évaluation. Ces personnes se sont assurées de la rigueur de la méthodologie utilisée, des analyses effectuées ainsi que de la pertinence de la démarche. La consultation de ces experts a été effectuée au moyen d'échanges de courriels et d'appels téléphoniques.

Enfin, une remarque s'impose au sujet de la catégorisation (échelle de niveaux) des éléments de preuve et de la recommandation sur l'efficacité théorique de ces outils. Bien qu'elle soit largement utilisée dans le domaine biomédical, l'échelle d'évaluation GRADE (Grading of Recommendations, Assessment, Development and Evaluation) appliquée à des domaines d'intervention pour lesquels il existe des devis d'études d'une autre nature que celui conçu pour les études cliniques aléatoires s'avère moins évidente [European Centre for Disease Prevention and Control, 2011]. C'est pour pallier cette limite que le conseil australien de la santé et de la recherche médicale, le National Health and Medical Research Council (NHMRC) [2009], a élaboré un système de catégorisation qui s'applique à diverses questions qui ne portent pas sur les interventions cliniques. Ce système a donc été retenu pour évaluer la qualité de la preuve scientifique sur l'efficacité théorique (fiabilité, validité) des outils d'évaluation du risque d'abus physique investigué (voir l'annexe J).

La formulation de recommandations qui a suivi cette étape résulte d'un processus de délibération avec les membres du comité de suivi sur les résultats obtenus et de commentaires exprimés par des experts indépendants sur la démarche scientifique menant à la réalisation de ce projet.

2 RÉSULTATS

Une recherche de l'information dans les bases de données scientifiques a permis de recenser 1 932 documents. La recherche de la littérature grise a aussi permis de repérer 386 documents. La figure B-1 de l'annexe B présente le processus de sélection des études sous la forme d'un diagramme. Une fois retirés les 522 documents repérés en double, 1 708 documents ont été exclus en raison de leurs titres ou de leurs résumés. Des 88 documents qui ont été retenus, 54 ont été exclus parce qu'ils ne répondaient pas aux critères d'inclusion sur l'intervention.

Parmi les 34 documents qui restaient, et dont la qualité a été évaluée, 5 documents ont été exclus en raison de leur faible qualité [Johnson *et al.*, 2012; Bolton et Lennings, 2010; Fuller et Nieto, 2010; Fitch, 2006; Agar, 2003]. La mise à jour de la revue de la littérature a donné lieu à l'ajout de 1 publication qui a également été évaluée [Dankert et Johnson, 2014]. Finalement, 30 documents jugés de bonne ou de moyenne qualité ont été retenus. Ces documents se répartissent comme suit :

- source de l'information : 21 documents issus des bases de données scientifiques et 9 documents issus de la littérature grise;
- territoire : 1 document produit en Australie, 6 documents produits au Canada et 23 documents produits aux États-Unis;
- plan d'étude : 1 étude à séquence temporelle avec groupe témoin non équivalent, 1 étude post-test à groupes multiples, 1 revue systématique de la littérature, 4 revues non systématiques de la littérature, 5 études comprenant un devis qualitatif, 17 études de cohortes (rétrospectives ou prospectives) et 1 opinion d'expert.

Les études exclues et les raisons de leur exclusion ainsi que la liste des articles retenus sont présentées dans l'annexe D. Les caractéristiques des études incluses sont présentées dans l'annexe E.

Au terme de cette démarche, 7 outils d'évaluation du risque d'abus physique envers des enfants ont été repérés dans la littérature. La revue systématique porte sur les résultats de 21 publications scientifiques dans lesquelles des mesures de fiabilité ou de validité de ces outils sont rapportées. Les principales caractéristiques de ces outils sont présentées dans l'annexe F. Les 9 autres documents retenus abordent les préoccupations relatives à l'utilité ou à l'applicabilité de ce type d'outils dans les SPJ.

2.1 Question 1 – Efficacité théorique des outils (fiabilité, validité)

Afin de répondre à la question qui porte sur l'efficacité théorique des outils d'évaluation du risque d'abus physique, 21 documents ont été retenus. Il s'agit de 17 études de cohortes (rétrospectives ou prospectives), de 1 revue systématique de la littérature scientifique, de 2 revues non systématiques de la littérature et de 1 étude post-test à groupes multiples. L'évaluation de la qualité scientifique de ces documents a été effectuée à l'aide de la grille de Kmet et ses collaborateurs [2004], à l'exception de la revue systématique de la littérature, dont l'évaluation de la qualité a été réalisée à l'aide de la grille AMSTAR (voir l'annexe C).

Une synthèse narrative décrit les résultats des mesures de fiabilité et de validité de chacun des outils ainsi que les particularités méthodologiques des études qui s'y rapportent. Les résultats d'intérêt sur la validité des outils actuariels sont présentés sous la forme de tableaux synthèses.

Les sept outils évalués sont les suivants :

- Adult Adolescent Parenting Inventory-2 (AAPI-2)
- Cleveland-Child Abuse Potential Scale (C-CAPS)
- Child Abuse Potential Inventory (CAPI) (la version française se nomme Inventaire du potentiel d'abus)
- California Family Risk Assessment (CFRA)
- Colorado Family Risk Assessment Tool (CFRAT)
- Family Risk Assessment of Abuse and Neglect (FRAAN)
- Ontario Family Risk Assessment (OFRA)

2.1.1 Fiabilité des outils

La fiabilité de 6 des 7 outils repérés a été mesurée dans 9 études; la fiabilité de l'outil CFRAT n'est pas abordée dans l'étude de Cooney et ses collaborateurs [2013].

2.1.2 Outils de type contextuel (CAPI, AAPI-2)

Les résultats des études sur la fiabilité des deux outils de type contextuel (les outils CAPI et AAPI-2) sont divergents.

Les résultats des études sur l'outil CAPI indiquent une bonne fiabilité. Au sujet de la stabilité de l'outil, c'est-à-dire de la constance des résultats obtenus par cet outil au gré de son application répétée, un prétest de 19 cas mené par Chaffin et Valle [2003], au cours d'une période de 2 semaines, fait état d'un coefficient de corrélation de Pearson égal à 0,74 concernant l'ensemble des éléments (160 éléments) qui composent l'outil, et de 0,91 concernant l'échelle de l'abus physique (77 éléments). Les évaluateurs s'accordent donc entre eux près de 3 fois sur 4 au moment d'évaluer les réponses fournies au moyen de l'outil, et de 9 fois sur 10 dans le cas de l'échelle de l'abus physique. Par ailleurs, l'étude menée par Wells et ses collaborateurs (2011) montre que cet outil offre une bonne validité interne : les facteurs de risque corrélaient fortement entre eux et avec le score total des sous-échelles qui composent l'outil. En effet, le coefficient de fiabilité Kuder-Richardson (KR20) est élevé relativement à un échantillon de mères adolescentes (KR20 = 0,90 à 0,96). Aussi, les caractéristiques sociodémographiques (sexe, âge, éducation, ethnie) de ces dernières influencent peu la validité interne de l'outil (KR20 = 0,85 à 0,96). Enfin, la fiabilité de l'outil se maintient, quels que soient sa procédure d'application et son milieu d'application (clientèle recevant des services à domicile, étudiants, etc.).

À l'opposé, les résultats de l'étude de Connors et ses collaborateurs [2006] sur l'outil AAPI-2 indiquent une faible validité interne entre les 5 échelles qu'il comprend : le coefficient alpha (α) de Cronbach varie de 0,59 à 0,79 relativement à ces échelles. La validité interne de l'outil est tout au plus acceptable ($\alpha \geq 0,60$) relativement à 3 de ces échelles (Expectations of

Children, Empathy Towards Children's Needs, Use of Corporal Punishment as a Mean of Discipline), qui concernent les attentes de l'enfant, l'empathie relative aux besoins de l'enfant et le recours aux châtements corporels disciplinaires. Elle est encore plus faible ($\alpha < 0,60$) relativement aux deux autres échelles (Parent-Child Role Responsibilities, Children's Power and Independence), qui concernent le rôle et les responsabilités parent-enfant ainsi que l'autonomie et le pouvoir de l'enfant. De l'avis même des auteurs de l'étude, ces résultats ne répondent pas aux standards de détection ($\alpha \geq 0,80$) ou de l'évaluation diagnostique ($\alpha \geq 0,90$). Ces auteurs concluent que le caractère unidimensionnel de l'outil fait défaut : ses différentes échelles sont peu reliées les unes aux autres.

2.1.2.1 Outils de type actuariel (C-CAPS, FRAAN, OFRA¹³, CFRA)

Les résultats des études sur la fiabilité de quatre outils de type actuariel varient passablement d'une étude à une autre.

Les auteurs de l'étude qui porte sur l'outil C-CAPS [Ezzo et Young, 2012] mentionnent sa bonne validité interne (alpha de Cronbach = 0,86) moyennant la combinaison de trois échelles, soit les antécédents, les caractéristiques familiales et les facteurs environnementaux. Toutefois, ce résultat ne s'applique pas aux quatre autres échelles qui composent l'outil. De plus, la possibilité d'un biais à l'absence d'insu, lors de l'application de l'outil, met en cause ce résultat : les cinq psychologues connaissaient bien les cas évalués; ils disposaient d'une importante documentation sur ces cas avant même d'utiliser l'outil C-CAPS. Leur propre interprétation de ces cas peut donc avoir influencé les résultats obtenus sur la fiabilité de l'outil. Ces éléments permettent de douter de fiabilité accordée à cet outil par les auteurs de cette étude.

Parmi les deux études sur l'outil FRAAN, une seule en a mesuré la fiabilité. En effet, Wood [2011] souligne un bon degré d'entente interjuges (Rho de Spearman = 0,78, $p < 0,001$). Ce constat concerne 20 cas choisis au hasard, et à l'insu d'un lecteur externe (un des superviseurs de l'étude), par l'auteure de cette étude.

Une seule des six études sur l'outil OFRA aborde la fiabilité de cet outil. En effet, Regehr et ses collaborateurs [2010] arrivent au constat d'une faible fiabilité de cet outil à partir des résultats obtenus à la suite d'une simulation qui consiste en 2 scénarios de 15 minutes chacun. Ces scénarios portent sur la rencontre entre un intervenant qui doit évaluer deux parents typiques (des acteurs engagés pour l'occasion). Dans ce contexte précis, la fiabilité de l'outil diffère selon les intervenants : certains évaluent le risque d'abus futur comme étant élevé, d'autres non. S'appuyant sur une échelle maison, les auteurs de l'étude mentionnent que sur une possibilité de 0 à 18 points pour l'échelle du risque d'abus, le plus haut résultat accordé est de 15 et le plus faible est de 1. La moyenne est de 4,82 (écart-type = 2,14) et les résultats obtenus par 62,4 % des répondants sont de 4, 5 ou 6. Ni l'âge, ni la scolarité, ni l'expérience de travail de ces derniers n'expliquent les écarts à cet égard¹⁴. En conclusion, les auteurs de l'étude mentionnent que la confiance des répondants en leur évaluation des cas, à l'aide de l'outil

¹³ L'outil OFRA – une composante de l'ensemble ORAM (SDM) – n'a pas été intégré aux autres études qui portent sur l'outil CFRA, dont il est la version adaptée en usage dans la province de l'Ontario (Canada). L'implantation de cet outil dans une province voisine du Québec présente un intérêt en soi, ce qui explique le choix de le considérer à part.

¹⁴ Précisons que 48 % des 96 participants à cette étude qualitative ont à intervenir à l'étape qui correspond, dans le contexte des SPJ au Québec, à la rétention et au traitement du signalement.

OFRA, s'expliquerait par la relation de collaboration qu'ils développent avec le parent évalué et non par l'efficacité de l'outil d'évaluation.

Concernant l'outil CFRA, trois études font état de sa fiabilité. Loman et Siegel [2004] mettent en évidence la faible fiabilité de cet outil. D'une part, l'alpha (α) de Cronbach est de 0,65 pour l'échelle du risque d'abus. Ce résultat est donc légèrement inférieur à la norme de référence ($\alpha \geq 0,70$). Par ailleurs, les résultats du sondage mené auprès de 459 intervenants ayant évalué des cas exemplaires, à l'aide de capsules, indiquent que la fiabilité interjuges pour l'échelle du risque d'abus diminue avec l'augmentation du nombre d'évaluations réalisées par ces intervenants. En effet, les intervenants ayant effectué 10 évaluations et plus à l'aide de l'échelle du risque d'abus obtiennent un α de 0,61; après 20 évaluations et plus : $\alpha = 0,56$; après 30 évaluations et plus : $\alpha = 0,51$; après 40 évaluations et plus : $\alpha = 0,56$; et après 50 évaluations et plus : $\alpha = 0,46$. D'après Loman et Siegel [2004], ce problème serait particulièrement évident au moment de délimiter le classement final des familles à risque de récidive. Pourtant, dans leur revue systématique de la littérature, Barlow et ses collaborateurs [2012] jugent que la fiabilité du CFRA est bonne, citant notamment l'étude de Loman et Siegel [2004].

En résumé, 9 études ont mesuré, de diverses manières, la fiabilité de 6 des 7 outils repérés dans la revue systématique de la littérature. Parmi les outils de type contextuel, les résultats de 2 études indiquent une bonne fiabilité de l'outil CAPI sur le plan de sa stabilité et de sa validité interne. En ce qui a trait aux outils de type actuariel, seul l'outil FRAAN montre une bonne fiabilité, selon une étude réalisée au Royaume-Uni, c'est-à-dire un pays étranger à celui où l'outil a été conçu, soit les États-Unis.

2.2 Validité des outils

2.2.1 Prédiction et classification des familles à risque dans le contexte des services de la protection de la jeunesse

On observe une ambiguïté concernant la validité prédictive des outils de type actuariel. Baird [2002] associe cette ambiguïté à l'utilisation interchangeable des termes « prédiction » et « classification » par les chercheurs, ce qui aurait pour effet de brouiller les attentes à l'égard de la validité de ce type d'outils¹⁵. Dans une étude de validation de l'outil OFRA, Shlonsky et ses collaborateurs [2011 : 8] évitent justement cette confusion des termes « prédiction » et « classification » par une mise en garde, sur la finalité de cet outil, qui souligne l'importance de se rappeler qu'il sert à classer adéquatement les familles concernées selon différents niveaux de risque (faible, modéré, élevé, très élevé) plutôt que de prédire, sans réserve, que ces familles commettront ou non à nouveau un acte de maltraitance¹⁶.

¹⁵ « Prediction is more precise than classification, and refers to the expectation of a particular outcome on the basis of observation, experience, or scientific reason. Classification, on the other hand, refers to the systematic arrangement of cases into groups or categories according to established criteria. While accurate prediction would greatly benefit child protective services and society, it has not proven feasible. For the present research, the goals of risk assessment are more modest. For the validity research component, the goal is to assign cases to different categories based on observed rates of behavior » [Baird, 2002 : 8].

¹⁶ « [...] it is extremely important to remember that this tool is designed to adequately differentiate between levels of risk (i.e., low, moderate, high, very high) rather predict, outright, whether families will come back with some form of maltreatment recurrence. »

La synthèse des résultats sur la validité des 7 outils recensés dans la revue de la littérature tient compte de ces deux approches.

D'une part, une approche fondée sur des critères de validité prédictive (sensibilité, spécificité, aire sous la courbe, etc.) s'applique, le plus souvent, à un outil qui contribue à l'évaluation diagnostique d'un acte de maltraitance. Les résultats sur la validité se présentent alors sous une forme dichotomique, à l'aide d'un tableau croisé à quatre cellules (ou tableau 2 X 2). Les vrais positifs et les faux positifs sont associés aux familles ayant ou non déjà abusé de leur enfant, et selon que ces familles sont ou non à risque de récidiver¹⁷.

D'autre part, une approche évalue la précision de la classification de groupes (familles) à risque à partir de leurs caractéristiques (similaires aux familles ayant ou non maltraité leur enfant) et selon différents niveaux de risque (faible, modéré, élevé, très élevé). La précision de l'outil réside d'une part, dans sa capacité de discrimination qui consiste à assurer une proportion suffisante de familles classées dans les différents niveaux de risque et d'obtenir un écart marqué entre celles qui sont situées aux extrêmes de la classification (familles à faible risque et familles à risque très élevé). La précision de l'outil dépend de sa capacité à classer les familles qui se distinguent entre elles par rapport au niveau de risque d'un acte d'abus futur. Les résultats de cette classification sont alors associés à au moins un indicateur d'abus futur : taux de nouveaux signalements, taux de réévaluation des cas, taux de nouvelles allégations fondées, taux de placement subséquent de l'enfant. Ces taux augmentent en fonction de la progression du niveau de risque déterminé à l'aide de l'outil au moment de l'évaluation initiale des cas. En cela, un outil sera d'autant plus valide que la probabilité d'un acte d'abus sera moins forte dans les familles évaluées « à faible risque » de récidive que dans les familles évaluées « à risque modéré », etc., dans les semaines, les mois ou les années à venir.

2.2.2 Validité des outils de type contextuel

La validité prédictive de l'outil **CAPi** a été examinée dans deux des trois articles retenus.

En effet, dans une revue (non systématique), Walker et ses collaborateurs (2010) observent que la structure factorielle de l'outil et la pondération des facteurs sont moins efficaces dans certains pays (Croatie, Grèce, Chili) qu'aux États-Unis où l'outil a été conçu. Ce résultat met ainsi en doute l'adaptation de cet outil aux caractéristiques culturelles des pays où il est utilisé. Ces auteurs rapportent également les résultats d'autres études qui indiquent qu'une mesure postintervention (après une année) permettrait d'observer la capacité de cet outil à mesurer les changements dans les comportements et les attitudes des parents ayant suivi un programme d'intervention.

Ce dernier point avait déjà été mis en doute dans l'étude de cohorte (avant et après le programme d'intervention) menée par Chaffin et Valle [2003], dont les résultats indiquent que l'outil CAPi offre une bonne capacité d'évaluation diagnostique (préintervention) : il est capable de repérer les groupes de parents à risque élevé de maltraiter leurs enfants parmi les participants à un programme d'intervention. Cependant, les résultats obtenus à l'aide de cet outil ne permettent pas de refléter les changements observés dans les comportements violents des parents ayant suivi un programme d'intervention. Affichant une valeur seuil de

¹⁷ À ce sujet, voir la sous-section intitulée *Validité prédictive et prise de décision* dans la section « Discussion ».

166 lors de l'évaluation préintervention, 34 % des parents ont été jugés à risque de commettre un abus physique ou de faire preuve de négligence à l'égard de leurs enfants selon des facteurs statiques (ex. : leurs caractéristiques sociodémographiques). Après l'intervention, ce résultat perd de son acuité, car l'outil CAPI ne peut mesurer les changements survenus dans les facteurs dynamiques que contient l'outil. Les écarts statistiquement significatifs entre les sous-échelles « Détresse », « Rigidité » et « Problèmes » avant et après l'intervention en témoignent. Les auteurs de l'étude n'observent aucun écart entre les divers niveaux de risque chez les parents ayant suivi le programme d'intervention. Au sujet de la validité prédictive de l'outil, l'aire sous la courbe n'indique guère d'écart entre la mesure préintervention (AUC= 0,661; intervalle de confiance (IC) : 95 %) et la mesure postintervention (AUC= 0,646; IC : 95 %). De plus, toute information additionnelle obtenue sur ces familles après l'intervention diminue la capacité de prédiction d'un abus futur à l'aide de l'outil CAPI. Chaffin et Valle [2003] concluent que l'utilisation de cet outil, comme celle de tout outil de ce type, pour évaluer l'efficacité d'une intervention est inadéquate, bien que la capacité diagnostique de l'outil CAPI soit bonne.

Dans leur étude sur l'**AAPI-2**, Connors et ses collaborateurs [2006] concluent que la validité concomitante (et non, prédictive) de cet outil est faible. En vertu de ce type de validité, les analyses de corrélation effectuées sur différentes échelles de mesure du même construit, qui sont appliquées à des sujets, doivent produire des résultats similaires. Ces auteurs ont ainsi comparé les résultats des corrélations entre l'outil AAPI-2 et quatre autres outils. Les résultats obtenus confirment un lien entre l'ensemble des échelles de l'outil AAPI-2 et ces autres outils : toutes les corrélations vont dans le sens prévu. Toutefois, ce lien est de faible ampleur : le coefficient de régression varie de 0,18 à -0,45. Les items de l'échelle de l'AAPI-2 diffèrent donc de celles d'autres outils de mesure du même phénomène. Selon les auteurs, cette faible cohérence interne des items de l'outil révèle que le caractère unifactoriel de l'échelle de cet outil n'est pas probant.

Les outils contextuels CAPI et AAPI-2 permettent ainsi d'établir un diagnostic des parents susceptibles de commettre un abus physique par rapport à ceux qui ne sont pas à risque de commettre ce type de maltraitance. **Aucun de ces outils ne peut toutefois prédire avec précision la récurrence d'un abus physique commis par ces personnes.**

2.2.3 Validité des outils de type actuariel

La validité des outils actuariels tient compte de la précision de la classification de familles selon différents niveaux de risque (ce qui implique la présence ou l'absence de facteurs de risque), une fois ceux-ci mis en relation avec des indicateurs d'abus physiques futurs (résultats d'intérêt). Dans la majorité des cas, la validité prédictive de ces outils s'appuie sur des études de validation dont le plan de recherche (ou d'étude) inclut une cohorte prospective. Cela permet de comparer la précision de l'outil en cours d'usage dans un SPJ (version originale) à une version révisée de l'outil une fois certaines modifications apportées à sa structure interne (recalibrage) par les auteurs de l'étude. Dans les autres cas, les études de validation portent sur une cohorte rétrospective¹⁸.

¹⁸ À ce sujet, voir l'annexe G.

Les résultats d'intérêt sur la validité de ces outils sont présentés dans les tableaux 1 à 5 (ci-après). Ces résultats font état du lien entre la classification des familles selon différents niveaux de risque (de récurrence) et les taux d'indicateurs d'abus futurs (tous types confondus) : le taux de nouveaux signalements, le taux de réévaluation des cas, le taux de nouvelles allégations fondées et le taux de placement subséquent d'un enfant en dehors du foyer familial pour cause de maltraitance (abus et négligence). Étant donné que le calibrage, le recalibrage et la capacité de discrimination d'un outil constituent des critères importants pour évaluer sa validité¹⁹, les résultats des mesures PSEP²⁰ et DIFR²¹ qui s'y rapportent sont également indiqués, lorsque ceux-ci sont disponibles.

2.2.3.1 Taux de nouveaux signalements d'abus (tous types confondus) selon la classification des familles à risque

Le tableau 1 regroupe les résultats obtenus sur les taux de nouveaux signalements d'abus (tous types confondus) au cours d'une période variant de 6 à 24 mois. Ces résultats proviennent d'études effectuées à la suite de l'évaluation des familles classées selon différents niveaux de risque.

¹⁹ Le calibrage (ou ajustement) d'un instrument de mesure est un « ensemble d'opérations réalisées sur un système de mesure pour qu'il fournisse des indications prescrites correspondant à des valeurs données des grandeurs à mesurer » (ISO/CEI GUIDE 99:2007). Dans le cas présent, le calibrage vise à améliorer le degré de correspondance entre le taux réel d'abus (résultant de l'évaluation initiale des cas) et le taux prédit (la récurrence) concernant les groupes (familles) à risque. Quant à la discrimination de l'outil, elle fait référence à sa capacité à distinguer les groupes (familles) plus ou moins susceptibles de vivre l'événement en cause, selon la classification obtenue (faible, modéré, élevé, très élevé) au moyen des scores globaux obtenus à l'aide des échelles de l'abus et de négligence qui composent l'outil.

²⁰ **PSEP (prognostic separation index)** : Indice de distinction pronostique qui mesure l'écart entre les taux de récurrence d'actes d'abus dans les familles à risque selon les niveaux de classification situés aux extrêmes (ex. : les familles à risque très élevé et les familles à faible risque). Bien qu'aucune norme ne soit établie, plus l'écart est grand entre ces classes extrêmes, meilleure est la capacité d'évaluer où diriger les services en priorité. Précisons que le calcul du PSEP présenté dans les tableaux 1 à 5 a été réalisé, dans la plupart des cas, par l'auteur du présent avis.

²¹ **DIFR (dispersion index for risk)** : Indice de dispersion relatif au risque qui permet d'évaluer comment les familles (cohortes) sont réparties en différents groupes à risque (trois groupes et plus) et la variabilité des résultats d'intérêt de chacun de ces groupes par rapport à l'ensemble des familles à risque (taux de référence). Le DIFR mesuré ici concerne les abus (tous types confondus) et la négligence.

Tableau 1 Taux de nouveaux signalements d'abus (tous types confondus) selon les niveaux de risque

Outil Auteurs (année) Durée de suivi Territoire		Échantillon (taille)	Niveaux de risque (%)					PSEP	DIFR
			faible	modéré	élevé	très élevé	Échantillon total (taux de référence)		
1	CFRAT Coohey <i>et al.</i> , 2013 24 mois Iowa (É-U)	6 832	11,8* 12,4**	16,6* 22,8**	25,9* 37,6**	s. o.	17,0	14,1* 25,2 **	0,40* 0,46**
2	FRAAN Johnson et Bogie, 2009 18 mois Caroline du Nord (É-U)	1 294	3,9* 2,9**	5,9* 5,6**	9,7* 15,1**	s. o.	4,7	5,8* 12,2**	-- -
3	OFRA Shlonsky <i>et al.</i> , 2011 24 mois Ontario (Canada)	7 782	8,52* 8,78**	10,35* 11,08**	12,07* 11,08**	7,27* s. o.	10,24	- 0,01* ,02**	0,12**
4	OFRA Lee <i>et al.</i> , 2014 12 mois Ontario (Canada)	8 210	13,1	28,2	42,2	26,2	24,8	13,1	-- -
5	CFRA Johnson <i>et al.</i> , 2006 18 mois Minnesota (É-U)	11 159	5,8* 5,5**	11,5* 11,9**	15,9* 21,0**	15,6* s. o.	9,4	9,8* 15,5**	-- -
6	CFRA Johnson et O'Connor, 2008 6 mois Californie (É-U)	5 706	4,0	7,4	13,2	15,6	7,1	11,6	

* Version originale de l'outil utilisée dans le SPJ

** Version révisée de l'outil (version améliorée proposée par les auteurs)

s. o. : sans objet (ne s'applique pas en raison de la combinaison des niveaux élevé et très élevé.)

PSEP (prognostic separation index) : Indice de distinction pronostique

DIFR (dispersion index for risk) : Indice de dispersion relatif au risque

-- - : non spécifié

Dans le tableau 1, on constate que la validité de l'outil **CFRAT²²** est adéquate en ce qui concerne le lien entre le classement des familles et les résultats d'intérêt concernant trois niveaux de risque (faible, modéré, élevé). Ce constat s'applique aux deux versions de l'outil : la version originale adoptée par le IADSH (association des SPJ en Iowa, aux États-Unis) et la version révisée (par l'ajout de facteurs de risque et l'ajustement de leur pondération) qui est proposée par Coohey et ses collaborateurs [2013]. En effet, 24 mois après l'évaluation initiale des familles, on observe une augmentation constante des taux de nouveaux signalements

²² Conçu à partir de l'outil CFRA (Californie, 1999), l'outil CFRAT a été élaboré au Colorado (États-Unis), puis adapté et appliqué partout en Iowa depuis 2005. Le « T » pour *tool* a été utilisé, dans le présent avis, afin de le distinguer du CFRA.

d'abus. Celle-ci correspond aux catégories de risque, à savoir les familles « à faible risque », les familles « à risque modéré » et les familles « à risque élevé » de faire l'objet de nouveaux signalements. On y trouve des proportions de 11,8 %, 16,6 % et 25,9 % respectivement, obtenues à l'aide la version originale de l'outil, et de 12,4 %, de 22,8 % et 37,6 % respectivement, obtenues à l'aide de la version révisée de l'outil.

Par ailleurs, la mesure PSEP obtenue à l'aide de la version révisée de l'outil indique un écart presque 2 fois plus élevé que celui obtenu à l'aide de la version originale de l'outil, soit 25,2 %, comparativement à 14,1 %. Ces résultats traduisent l'écart (la dispersion) entre les catégories situées aux extrêmes de la classification, c'est-à-dire entre les familles « à faible risque » et les familles « à risque élevé » de récidive. Par ailleurs, la mesure DIFR indique une capacité de discrimination adéquate de l'outil. Cette mesure DIFR obtenue à l'aide de la version révisée de l'outil est légèrement plus élevée (DIFR = 0,46) que celle obtenue à l'aide de la version originale de l'outil (DIFR = 0,40). Cela signifie que la répartition des familles entre les différents niveaux de risque de récidive obtenue à l'aide de la version révisée de l'outil est un peu mieux équilibrée que celle obtenue à l'aide de la version originale de l'outil. Précisons que l'outil CFRAT se distingue d'autres outils ayant la même finalité parce qu'il permet une classification à trois niveaux de risque. Cette particularité limite les comparaisons possibles avec les résultats d'outils qui comptent quatre niveaux de risque. En effet, une classification comportant un moins grand nombre de niveaux de risque tend à favoriser un plus grand écart entre ceux-ci lorsque les résultats d'intérêt sont pris en compte [Johnson *et al.*, 2007; Johnson et Wagner, 2003].

Les résultats de l'étude de validation interne de l'outil **FRAAN** par Johnson et Bogie [2009] indiquent, eux aussi, une meilleure validité de la version révisée de l'outil par rapport à la version originale, après une période de suivi de 18 mois. En effet, le taux obtenu à l'aide de la version originale de l'outil est 2,5 fois plus élevé concernant les familles à risque élevé de récidive (9,7 %) que celui obtenu concernant les familles à faible risque de récidive (3,9 %). Le taux obtenu à l'aide de la version révisée de l'outil est 5 fois plus élevé concernant les familles à risque élevé (15,1 %) que celui obtenu concernant les familles à faible risque de récidive (2,9 %). Quant au résultat du PSEP de la version révisée de l'outil (12,2 %), il est de plus du double du résultat du PSEP obtenu à l'aide de la version originale de l'outil (5,8 %). Ces résultats indiquent une validité adéquate selon les critères de calibrage et de discrimination entre les deux versions de l'outil (originale, révisée). Cependant, la version révisée de l'outil proposé par Johnson et Bogie [2009] permet d'obtenir une distinction encore plus fine du classement des familles selon le niveau de risque, notamment en ce qui concerne les familles à risque élevé, comparativement à celle obtenue à l'aide de la version originale de l'outil, utilisée par les SPJ en Caroline du Nord.

En ce qui a trait à la validité de l'outil **OFRA**, elle est mesurée dans trois des six études évaluées. Cependant, les résultats de deux de ces trois études ne sont plus d'actualité, parce qu'ils portent sur l'outil d'évaluation du risque RAT (*Risk Assessment Tool*), lequel a été remplacé en 2007 par l'outil OFRA. Il ne reste donc que l'étude de Shlonsky et ses

collaborateurs [2011], et sa mise à jour par Lee et ses collaborateurs [2014]²³, pour évaluer la validité de l'outil OFRA.

Les études sur l'outil OFRA menées par Shlonsky et ses collaborateurs, et par Lee et ses collaborateurs, portent sur des sous-échantillons de population qui proviennent de la même base de données (OCANDS) et dont la taille est presque similaire. La période consacrée à l'intervention (l'application de l'outil) dans l'étude menée par Lee et ses collaborateurs recouvre celle de Shlonsky et ses collaborateurs : la première est de 4 ans, soit de janvier 2008 à décembre 2011; la deuxième est de 6 mois, soit de janvier à juin 2008. Cependant, la période de suivi des cohortes de familles observées varie entre ces deux études, soit 12 mois et 24 mois respectivement.

Dans ces deux études sur l'outil OFRA, la catégorie « à risque très élevé » pose problème dans le cas des nouveaux signalements. Les familles qui y sont classées sont peu nombreuses (n = 110 ou 7,27 % dans l'étude de Shlonsky et ses collaborateurs; n = 99 ou 26,2 % dans l'étude de Lee et ses collaborateurs). La dispersion est plus marquée entre les catégories, notamment celles situées aux extrêmes de la classification (PSEP = 13,1 %) dans la deuxième étude en comparaison avec les résultats de la première, où l'écart est très faible entre les catégories « à risque élevé » et « à risque très élevé ». Aussi, l'indice PSEP se situe autour de zéro. Afin d'améliorer ces résultats, les auteurs de ces deux études proposent de combiner les catégories « à risque élevé » et « à risque très élevé », de manière à établir une classification à 3 niveaux plutôt qu'à 4 niveaux de risque.

Shlonsky et ses collaborateurs [2011] ont procédé à cette expérimentation. Dans ce scénario, la catégorie « à risque élevé et très élevé » est associée à 11,08 %, soit le même résultat que celui associé à la catégorie « à risque modéré ». Le niveau de dispersion demeure faible (DIFR : 0,12, PSEP : 0,02). Bref, en utilisant trois niveaux de risque au lieu de quatre, la validité interne de l'outil s'est améliorée, mais demeure faible concernant l'abus physique, d'où la recommandation faite par ces auteurs de continuer à utiliser l'outil dans sa version originale. Enfin, les auteurs respectifs des deux études concluent que la validité de l'outil OFRA est adéquate étant donné le lien établi entre la classification des familles selon les niveaux de risque et les indicateurs d'abus physique futurs, dont le taux de nouveaux signalements.

En ce qui a trait à la validité de l'outil **CFRA**, 8 des 10 études évaluées l'ont mesurée. Dans l'ensemble, les auteurs de ces études concluent, eux aussi, que la validité de cet outil est adéquate. Parmi ces huit études, deux d'entre elles abordent la validité en référence aux résultats sur le taux de nouveaux signalements d'abus. Dans l'une d'elles [Johnson *et al.*, 2006], les auteurs proposent une version améliorée de l'outil dans les SPJ du Minnesota à partir d'un échantillon de 11 159 familles. Au cours d'une période de 18 mois après l'évaluation initiale, la version révisée de l'outil permet d'obtenir un écart plus marqué entre les niveaux « à faible risque » et « à risque élevé », une fois ceux-ci mis en relation avec les taux de nouveaux signalements d'abus physique. En effet, dans sa version originale, l'outil permet d'obtenir un PSEP égal à 9,8 % alors que dans la version révisée, l'outil permet d'obtenir un PSEP égal à 15,5 %. Aussi, les résultats obtenus à l'aide de la version révisée de l'outil relativement aux autres niveaux de risque (faible, modéré) sont légèrement meilleurs. Il

²³ Il s'agit d'un bulletin sur la mise à jour de la validation de l'outil OFRA. Cette démarche poursuit celle réalisée par Shlonsky et ses collaborateurs [2011], et ce, à partir de la même base de données (OCANDS).

faut souligner que ce dernier ne possède plus que trois niveaux de risque étant donné les résultats presque similaires entre les niveaux « à risque modéré » (15,9 %) et « à risque élevé » (15,6 %) obtenus à l'aide de la version originale de l'outil. Cela explique pourquoi les auteurs proposent le regroupement de ces deux niveaux dans la version révisée. Il faut préciser que la représentation des familles « à risque très élevé » de la version originale est beaucoup plus faible (3,4 %) que celle de la version révisée (22,1 %), si l'on considère l'ensemble des types d'abus (tous types confondus) et la négligence. Cette situation rend difficiles l'interprétation et la comparaison des données de cette catégorie selon une classification à trois ou à quatre niveaux de risque.

Les résultats de l'étude sur l'outil CFRA menée à Washington par Johnson et O'Connor [2008] indiquent que la classification des familles à risque de commettre un abus futur est effectivement associée aux taux de nouveaux signalements d'abus. Par exemple, les familles « à risque très élevé » de récurrence affichent un taux près de 4 fois supérieur à celui des familles « à faible risque » de nouveaux signalements (15,6 % comparés à 4,0 %). De façon générale, l'augmentation du taux de nouveaux signalements d'abus suit la progression de la catégorisation, soit en fonction de chaque niveau de risque (de faible à modéré, de modéré à élevé, et d'élevé à très élevé). Cependant, dans sa version originale, l'outil est moins efficace au moment de distinguer les familles « à risque élevé » de celles « à risque très élevé », si l'on se fie au faible écart entre les taux dans ces deux catégories (13,2 % et 15,6 %). D'après Johnson et O'Connor, une période de suivi de 12 à 24 mois permettrait d'obtenir de meilleurs résultats que dans cette étude, dans laquelle la période de suivi n'est que de 6 mois.

2.2.3.2 Taux de réévaluation des cas d'abus (tous types confondus) selon la classification des familles à risque

Le tableau 2 regroupe les résultats sur les taux de réévaluation des cas d'abus (tous types confondus) selon la classification des familles dans les différents niveaux de risque de récurrence de ce type de maltraitance.

Tableau 2 Taux de réévaluation des cas d'abus (tous types confondus) selon les niveaux de risque

Outil Auteurs (année) Durée de suivi Territoire		Échantillon (taille)	Niveaux de risque (%)					PSEP	DIFR
			faible	modéré	élevé	très élevé	Échantillon total (taux de référence)		
1	FRAAN Johnson et Bogie, 2009 18 mois Caroline du Nord (É-U)	1 294	3,6* 2,6**	5,6* 5,2**	9,7* 15,1**	s. o.	4,4	6,1* 12,5**	--
2	OFRA ²⁴ Shlonsky <i>et al.</i> , 2011 24 mois Ontario (Canada)	7 782	7,10* 7,31**	8,96* 9,56**	10,42* 10,09**	7,27 s. o.	8,81	0,17* 0,03**	0,14**
3	OFRA Lee <i>et al.</i> , 2014 12 mois Ontario (Canada)	8 210	10,4	22,1	34,6	22,2	11,8	12,6	--
4	CFRA Johnson et Wagner, 2003 24 mois Californie (É-U)	5 694	24,5* 28,2**	37,1* 42,8**	48,5* 60,3**	61,3* 74,8**	32,3	36,8* 46,6**	--
5	CFRA Johnson <i>et al.</i> , 2006 18 mois Minnesota (É-U)	11 159	4,3* 4,0**	8,9* 9,2**	13,6* 19,0**	15,6* s. o.	7,4	11,3* 15,0**	--
6	CFRA Johnson <i>et al.</i> , 2007 18 mois Californie (É-U)	10 097	15,9* 12,6**	23,6* 23,0**	30,6* 34,4**	32,6* 49,6**	20,7	16,7* 37,0**	--
7	CFRA Dankert et Johnson, 2014 18 mois Californie (É-U)	11 444	13,4* 11,4**	22,3* 20,9**	35,7* 33,3**	40,2* 45,7**	20,9	26,8* 34,3**	0,54**

* Version originale de l'outil utilisée dans le SPJ

** Version révisée de l'outil (version améliorée proposée par les auteurs)

s. o. : sans objet (ne s'applique pas en raison de la combinaison des niveaux élevé et très élevé.)

-- : non spécifié

PSEP (prognostic separation index) : Indice de distinction pronostique

DIFR (dispersion index for risk) : Indice de dispersion relatif au risque

Les résultats de l'étude de Johnson et Bogie [2009] sur l'outil **FRAAN** indiquent une augmentation graduelle du taux de réévaluation des cas d'abus selon la classification des niveaux de risque. Concernant la catégorie des familles « à risque élevé », le résultat obtenu à l'aide de la version révisée de l'outil s'est quelque peu amélioré (15,1 %) en comparaison du résultat obtenu à l'aide de la version originale (9,7 %), mais seulement en ce qui concerne

²⁴ Dans les études sur l'outil OFRA réalisées par Shlonsky et ses collaborateurs [2011] et par Lee et ses collaborateurs [2014], les données sur les niveaux de risque concernent uniquement l'abus physique.

cette catégorie. Cela fait en sorte de doubler le résultat du PSEP, qui passe ainsi de 6,1 % à 12,5 %. Le taux de référence est cependant très faible, soit 4,4 %.

Les résultats des deux études sur l'outil **OFRA**, soit celle de Shlonsky et ses collaborateurs [2011], et celle de Lee et ses collaborateurs [2014], montrent aussi que le taux de réévaluation des cas d'abus est associé à la catégorisation des familles à risque. Dans la première, les taux de réévaluation des cas obtenus à l'aide des deux versions, originale et révisée, de l'outil sont presque similaires; on rapporte un faible écart selon les niveaux de risque. Les résultats des mesures du PSEP (version originale : 0,17; version révisée : 0,03) et du DIFR (0,14) sont faibles. Il faut dire que la catégorie des familles « à risque très élevé » (7,27 %) ne suit pas la catégorisation du risque de la version originale. Même la réduction de 4 à 3 niveaux de risque de la version révisée ne permet pas de résoudre ce problème. En effet, le résultat de cette catégorie passe de 10,42 % à 10,09 %.

Ce problème est aussi rapporté dans l'étude de Lee et ses collaborateurs, dont les résultats s'appuient sur la même base de données (OCANDS) que celle de Shlonsky et ses collaborateurs. En effet, le taux de réévaluation des cas d'abus physique de la catégorie des familles « à risque très élevé » (22,2 %) est similaire au taux de la catégorie des familles « à risque modéré » (22,1 %); il est moins élevé que le taux de la catégorie « à risque élevé » (34,6 %). À l'exception de la catégorie « à risque très élevé », les résultats relatifs aux autres catégories suivent une progression continue (10,4 %, 22,1 %, 34,6 %). Aussi, la dispersion des catégories situées aux extrêmes de la classification est de loin supérieure (PSEP : 12,6 %) à celle obtenue par Shlonsky et ses collaborateurs. Ces derniers mentionnent d'ailleurs que l'échelle l'abus physique donne de moins bons résultats que l'échelle de la négligence²⁵.

Par ailleurs, une étude de validation de l'outil **CFRA** a été menée par Johnson et Wagner en 2003. Si on se fie aux résultats obtenus au cours des 24 mois qui suivent l'évaluation initiale des 5 694 familles, la validité de l'outil est jugée bonne étant donné la précision de la classification relative aux taux de réévaluation des cas d'abus. En effet, la progression de ces taux suit celle de la catégorisation des niveaux de risque (faible : 24,5 %, modéré : 37,1 %, élevé : 48,5 %, très élevé : 61,3 %) de la version originale de l'outil. Les résultats obtenus à l'aide de la version révisée de l'outil sont encore meilleurs. Par exemple, le taux de réévaluation des cas d'abus se situe à 74,8 % concernant les familles classées « à risque très élevé ». Aussi, les résultats de la mesure PSEP sont les plus élevés de ceux obtenus par l'ensemble des outils évalués, soit 36,8 % à l'aide de la version originale de l'outil et 46,6 % à l'aide de la version révisée. Bref, les deux versions du CFRA classent adéquatement les familles selon différents niveaux de risque, une fois considéré le taux de réévaluation des cas d'abus. La version révisée de l'outil est encore plus efficace que la version originale à ce chapitre.

Dans les 3 autres études sur l'outil CFRA, la durée de suivi de la cohorte observée est un peu moins longue, soit 18 mois, que celle de Johnson et Wagner. Les résultats vont néanmoins dans le même sens, bien qu'ils soient plus faibles, en particulier concernant la catégorie des familles « à risque élevé ».

²⁵ Rappelons que les résultats des deux études sur l'outil OFRA portent uniquement sur l'abus physique, ce qui restreint la comparaison des résultats rapportés dans des études sur des outils évalués relativement à d'autres types de maltraitance en plus de celui-ci.

Prenant appui sur un échantillon composé de 11 159 familles, les résultats de l'étude menée par Johnson et ses collaborateurs en 2006 indiquent une progression des taux de réévaluation des cas d'abus qui suit celle de la catégorisation des différents niveaux de risque de la classification des familles. Les auteurs de cette étude cherchent à améliorer la précision de cette classification adoptée dans les SPJ du Minnesota en proposant une version révisée de l'outil CFRA. À cette fin, ils réduisent le nombre de catégories de la classification, qui passe de quatre à trois, en jumelant la catégorie « à risque élevé » à la catégorie « à risque très élevé ». En conséquence, la version révisée de l'outil accentue l'écart entre les catégories situées aux extrêmes de la classification. En effet, le taux de réévaluation de cas relatif aux familles « à faible risque » est de 4,0 % comparativement à 19,0 % concernant les familles « à risque élevé et très élevé ». Il faut préciser qu'en ce qui concerne l'ensemble de l'échantillon, la représentation des familles « à risque très élevé » dans la version originale de l'outil est beaucoup plus faible (3,4 %) par rapport à celle dans la version révisée (22,1 %), si l'on considère l'ensemble des types de maltraitance et non seulement l'abus physique. Cette situation rend difficile l'interprétation des données relatives à la catégorie des familles « à risque très élevé » selon la classification à quatre niveaux de risque de la version originale. Néanmoins, la validité de l'outil s'avère adéquate, que ce soit dans sa version originale ou dans sa version révisée, laquelle donne des résultats légèrement meilleurs que l'originale.

Par ailleurs, dans leur étude, Johnson et ses collaborateurs signalent qu'au cours de la période de 18 mois suivant l'évaluation, la version originale de l'outil distingue clairement les familles classées selon différents niveaux de risque relativement au taux de réévaluation des cas d'abus. La progression d'un niveau de risque au suivant (faible : 15,9 %, modéré : 23,6 %, élevé : 30,6 %, très élevé : 32,6 %) confirme une assez bonne calibration de l'outil. Les auteurs proposent une version révisée de l'outil étant donné que la version originale distingue moins bien les familles « à risque élevé » de celles « à risque très élevé ». La première permet d'obtenir une plus grande proportion de familles, classées dans ces catégories, qui font l'objet d'une réévaluation de leur cas. Par exemple, les familles à risque très élevé d'abus futurs affichent un taux de réévaluation de leurs cas qui est 4 fois supérieur à celui des familles à faible risque (49,6 % comparativement à 12,6 %). Le résultat de la mesure PSEP est de 20,7 % et la mesure de dispersion DIFR est de 0,58, ce qui indique une bonne capacité de discrimination entre les niveaux de risque de la classification.

L'étude de validation la plus récente de l'outil CFRA est celle de Dankert et Johnson [2014]. Ces auteurs suivent une démarche similaire à celle de Johnson et ses collaborateurs [2007] et les résultats des deux études correspondent. Dans les 18 mois qui ont suivi l'évaluation de 11 444 familles en 2011, tant à l'aide de la version originale de l'outil qu'à l'aide de la version révisée, Dankert et Johnson ont observé un écart marqué entre les niveaux de risque situés aux extrêmes de la classification (selon la mesure PSEP) tout comme entre les niveaux consécutifs (selon la mesure DIFR). À chaque échelon du niveau de risque correspondait un accroissement substantiel du taux de réévaluation des cas d'abus. Par exemple, les auteurs ont constaté que le passage du niveau « à faible risque » au niveau « à risque modéré », et le passage du niveau « à risque modéré » au niveau « à risque élevé » correspondaient à une augmentation de 35 % du taux de réévaluation des cas. Le taux obtenu à l'aide de la version révisée relativement aux familles à risque très élevé (45,7 %) est 4 fois plus élevé que celui relatif aux familles à faible risque (11,4 %). Cela se reflète dans les résultats élevés des mesures PSEP (34,3 %) et DIFR (0,54).

Les résultats relatifs au taux de réévaluation des cas dans les plus récentes études considérées aux fins de la production du présent avis indiquent une validité adéquate de l'outil CFRA. Cependant, sa version adaptée, en usage en Ontario (outil OFRA), donne de moins bons résultats en ce qui a trait à la catégorie des familles « à risque élevé ». La validité de l'outil FRAAN est également adéquate, bien que celui-ci soit avantagé du fait que la classification comporte trois niveaux de risque au lieu des quatre niveaux que comportent les autres outils.

2.2.3.3 Taux de nouvelles allégations fondées d'abus (tous types confondus) selon la classification des familles à risque

Dans leur étude de validation (temporelle) de l'outil **CFRAT**, Coohy et ses collaborateurs [2013] ont observé, 24 mois après l'évaluation initiale des familles, une correspondance étroite entre le taux de nouvelles allégations fondées d'abus et la progression des niveaux de risque selon la classification établie, et ce, tant en ce qui concerne les résultats obtenus à l'aide de la version originale qu'en ce qui concerne les résultats obtenus à l'aide de la version révisée de l'outil. Comme le montre le tableau 3, à l'aide de la version révisée que ces auteurs proposent, l'écart entre les résultats des niveaux de risque passe du simple au double, soit respectivement 3,8 %, 9,2 % et 17,2 %. La catégorie des familles « à risque élevé » est celle où l'écart entre les résultats obtenus à l'aide de la version originale et ceux obtenus à l'aide de la version révisée de l'outil est le plus marqué, soit respectivement 9,7 % et 17,2 %. La mesure PSEP obtenue à l'aide de la version révisée de l'outil indique un écart 2 fois plus élevé (13,4 %), qui tend à polariser davantage les catégories situées aux extrêmes de la classification que celle obtenue à l'aide de la version originale de l'outil (5,7 %).

Tableau 3 Taux de nouvelles allégations fondées d’abus (tous types confondus) selon les niveaux de risque

Outil Auteurs (année) Durée de suivi Territoire		Échantillon (taille)	Niveaux de risque (%)					PSEP	DIFR
			faible	modéré	élevé	très élevé	Échantillon total (taux de référence)		
1	CFRAT Coohey <i>et al.</i> , 2013 24 mois Iowa (É-U)	6 832	4,0* 3,8**	6,2* 9,2**	9,7* 17,2**	s. o.	6,2	5,7* 13,4**	--
2	OFRA Shlonsky <i>et al.</i> , 2011 24 mois Ontario (Canada)	7 782	2,60* 2,47**	3,45* 3,70**	3,89* 4,35**	2,73* s. o.	3,3	0,14* 0,02**	0,22**
3	OFRA Lee <i>et al.</i> , 2014 12 mois Ontario (Canada)	8 210	3,7	10,2	17,3	18,2	9,0	5,3	--
4	CFRA Johnson et Wagner, 2003 24 mois Californie (É-U)	5 694	6,1* 8,8**	10,6* 16,7**	16,8* 28,9*	22,5* 46,2**	9,1	16,4* 37,4**	--
5	CFRA Johnson <i>et al.</i> , 2006 18 mois Minnesota (É-U)	11 159	1,5* 1,3**	3,7* 3,9**	6,4* 9,1**	9,9* s. o.	3,0	8,4* 7,8**	--
6	CFRA Johnson <i>et al.</i> , 2007 18 mois Californie (É-U)	10 097	3,8* 2,9**	6,5* 5,9**	8,0* 10,0**	8,1* 13,0**	5,4	4,3* 10,1**	0,58**
7	CFRA Johnson et O’Connor, 2008 6 mois Californie (É-U)	5 706	0,2	1,2	1,6	2,4	0,9	2,2	--
8	CFRA Dankert et Johnson, 2014 18 mois Californie (É-U)	11 444	2,3* 2,2**	4,9* 4,2**	8,5* 8,1**	8,3* 9,9**	4,4	6,0* 7,7**	0,36**

* Version originale de l’outil utilisée dans le SPJ

** Version révisée de l’outil (version améliorée proposée par les auteurs)

s. o. : sans objet (ne s’applique pas en raison de la combinaison des niveaux élevé et très élevé.)

-- : non spécifié

PSEP (prognostic separation index) : Indice de distinction pronostique

DIFR (dispersion index for risk) : Indice de dispersion relatif au risque

La version révisée de l’outil est, elle aussi, légèrement plus exacte que la version originale si l’on se fie aux valeurs de l’aire sous la courbe ROC (AUC), non illustrée ici. En effet, mesurée à l’aide de la version originale, la possibilité qu’une famille fasse l’objet d’une nouvelle

allégation fondée d'abus est de 60 % (AUC = 0,60) et elle est de 68 % (AUC = .68), soit une valeur légèrement supérieure à la valeur obtenue à l'aide de la version révisée de l'outil. Ces valeurs indiquent une validité prédictive adéquate de l'outil dans le contexte spécifique des SPJ²⁶.

Dans leur étude de validation (interne) prospective de l'outil **OFRA**, Shlonsky et ses collaborateurs [2011] mentionnent qu'au cours des 24 mois qui ont suivi l'évaluation initiale à l'aide de cet outil, l'échantillon composé de 7 782 familles affichait 11,9 % des nouvelles allégations fondées de maltraitance en général. Parmi ces résultats, l'abus physique représente 3,34 % des nouvelles allégations fondées. La catégorie « à risque très élevé » pose problème en raison du petit nombre de cas (n) qui y sont classés (n = 110). En effet, cette catégorie est associée à seulement 2,73 % des nouvelles allégations fondées d'abus physique. De plus, l'écart est très faible entre les catégories « à risque élevé » (3,89 %) et « à risque très élevé » (2,73 %). Le résultat du PSEP (0,14) est faible.

Afin de pallier ces lacunes, les auteurs de l'étude proposent de regrouper la catégorie « à risque élevé » avec la catégorie « à risque très élevé » afin d'obtenir trois niveaux de risque au lieu de quatre. Ce recalibrage de l'outil améliore quelque peu sa structure. Dans le tableau 3, on remarque que dans une classification à trois niveaux de risque, le jumelage des catégories « à risque élevé » et « à risque très élevé » donne un résultat légèrement supérieur (4,35 %) à celui de la catégorie « à risque élevé » de l'outil à quatre niveaux de risque (3,89 %). Cependant, quel que soit le nombre de catégories de l'outil (trois ou quatre), le taux de nouvelles allégations fondées de la catégorie « à risque élevé et très élevé » demeure faible; il égale ou dépasse de peu celui de la catégorie « à risque modéré ». Quant aux mesures de discrimination selon cette classification à trois niveaux de risque, elles sont faibles (DIFR : 0,22; PSEP : 0,02). Bref, si la validité interne de l'outil s'est légèrement améliorée, elle demeure faible en ce qui concerne l'échelle de l'abus physique.

Les résultats du second volet de la validation prospective de l'outil OFRA [Lee *et al.*, 2014] portent sur 33 385 familles ontariennes ayant fait l'objet d'une évaluation concernant un problème de maltraitance (tous types confondus). Dans les 12 mois suivant cette évaluation, les résultats indiquent que, relativement à l'ensemble de cet échantillon, on observe une proportion plus grande de signalements (24,8 %) que de cas réévalués (11,8 %) de même qu'une proportion plus grande de cas réévalués que de cas faisant l'objet de nouvelles allégations fondées (9,0 %). Concernant cette dernière valeur, les auteurs constatent également que la proportion augmente avec les échelons de niveau de risque, mais que la catégorie des familles « à risque élevé » (17,3 %) ne se distingue guère de la catégorie des familles « à risque très élevé » (18,2 %). Lee et ses collaborateurs [2014] admettent que la catégorie « à risque très élevé » ne peut pas être associée à la récurrence étant donné le faible nombre de cas contenus dans cette catégorie (n = 99). En conséquence, les auteurs proposent de combiner les catégories « à risque élevé » et « à risque très élevé » de manière à établir une classification à trois niveaux au lieu de quatre, mais sans tenter une expérimentation en ce sens. En fin de parcours, les auteurs des deux volets (2010 et 2014) de l'étude sur la validité de l'outil OFRA

²⁶ Dans un contexte expérimental, l'aire sous la courbe (AUC) doit être $\geq 0,90$ pour être jugée excellente et se situer entre 0,70 et 0,89 pour être jugée adéquate. Selon Coohy et ses collaborateurs [2013 : 153], dans le contexte des SPJ, cette valeur dépasse rarement 0,70.

concluent à une validité adéquate étant donné le lien établi entre la classification des familles selon les niveaux de risque et les indicateurs d'abus physique futurs.

Dans cinq études de validation de l'outil **CFRA**, il est question du taux d'allégations fondées d'abus.

Les résultats obtenus à ce sujet par Johnson et Wagner [2003] sont aussi significatifs que ceux déjà indiqués relativement aux autres indicateurs d'abus futurs. En effet, la progression du taux de nouvelles allégations fondées d'abus suit la catégorisation des niveaux de risque de la classification élaborée à l'aide de l'outil CFRA. Le taux de nouvelles allégations fondées de la catégorie « à risque élevé » est près de 4 fois plus élevé que celui de la catégorie « à faible risque », soit 22,5 % comparativement à 6,1 %. La version révisée de l'outil accentue encore cet écart, le taux qu'elle permet d'obtenir étant 5 fois plus élevé (46,2 % comparativement à 8,8 %) entre ces 2 catégories de familles à risque, qui se situent aux extrêmes de la classification. La mesure PSEP obtenue à l'aide de la version révisée est 2 fois plus élevée que celle obtenue à l'aide de la version originale de l'outil soit 37,4 % comparativement à 16,7 %.

De leur côté, Johnson et ses collaborateurs [2006] proposent une version révisée de l'outil CFRA qui permet une meilleure répartition des familles à risque en comparaison de la version originale. Le taux de nouvelles allégations fondées triple d'un niveau de risque au suivant; celui des familles « à faible risque » étant de 1,3 %, celui des familles « à risque modéré », de 3,9 %, et celui des familles « à risque élevé et très élevé », de 9,1 %. Le fait de réduire le nombre de catégories (niveaux de risque) de quatre à trois hausse le résultat de la catégorie des familles « à risque élevé », lequel passe ainsi de 6,4 % à 9,1 %. Il faut mentionner que cette catégorie combine les familles « à risque élevé » et les familles « à risque très élevé » de la version originale de l'outil. Donc, le gain obtenu par cette nouvelle catégorisation est étroitement lié au retrait de cette dernière catégorie, dont le taux obtenu à l'aide de la version originale est de 9,9 %. Néanmoins, le résultat de la mesure PSEP obtenu à l'aide de la version révisée est légèrement plus faible soit 7,8 % comparativement à 8,4 %.

Dans leur étude de 2007, Johnson et ses collaborateurs signalent qu'au cours de la période de 18 mois suivant l'évaluation, la version originale de l'outil distingue assez bien les familles, classées selon différents niveaux de risque relativement au taux de nouvelles allégations fondées. L'écart entre les résultats relatifs à chacun des niveaux de risque n'est toutefois pas très marqué. La version révisée de l'outil améliore cette situation, en particulier pour ce qui est du taux de nouvelles allégations fondées d'abus dans la catégorie « à risque élevé » de récidive (13,0 %). La mesure PSEP obtenue à l'aide de la version révisée est de 10,1 % comparativement à 4,3 % à l'aide de la version originale.

L'étude de Johnson et O'Connor [2008] montre que si les taux de nouvelles allégations fondées d'abus suivent effectivement la même progression que la catégorisation (échelons) des niveaux de risque au cours des 6 mois qui suivent l'évaluation des cas à l'aide de l'outil CFRA, ces taux demeurent peu élevés. Le taux de base se situe à seulement 0,9 %. Il faut mentionner que la durée de suivi des familles est de 6 mois, donc la plus courte parmi les études sur cet outil.

Enfin, l'étude de Dankert et Johnson [2014], dont le suivi de l'échantillon des familles se déroule sur une période de 18 mois, donne des résultats un peu plus significatifs. La version originale de l'outil distingue clairement ces familles selon le niveau de risque relatif aux taux de nouvelles allégations fondées d'abus. Ces taux passent du simple au double, sauf dans la catégorie des familles « à risque très élevé », dont le taux est presque similaire (8,3 %) à celui

de la catégorie des familles « à risque élevé » (8,5 %). Le résultat obtenu à l'aide de la version révisée de l'outil est légèrement meilleur : la catégorie « à risque très élevé » affiche ainsi 9,9 %. Toutefois, la mesure PSEP obtenue à l'aide de la version révisée de l'outil est de 7,7 %; elle se démarque peu de celle obtenue à l'aide de la version originale, qui est de 6,0 %.

2.2.3.4 Taux de placement subséquent d'un enfant hors famille en raison de maltraitance (abus et négligence) selon les niveaux de risque

Les résultats des études sur cet indicateur sont présentés dans le tableau 4. Seuls les résultats, qui englobent l'abus et la négligence, obtenus à l'aide de l'outil **CFRA** sont disponibles.

Tableau 4 Taux de placement subséquent d'un enfant en dehors du foyer familial en raison de maltraitance (abus et négligence), selon les niveaux de risque

Outil Auteurs (année) Durée de suivi Territoire		Échantillon (taille)	Niveaux de risque (%)				Échantillon total (taux de référence)	PSEP	DIFR
			faible	modéré	élevé	très élevé			
1	CFRA Johnson et Wagner, 2003 24 mois Californie (É-U)	5 694	1,1* 0,6**	2,9* 1,9**	12,6* 7,4**	31,7* 22,1**	4,7	30,6* 21,5**	--
2	CFRA Johnson <i>et al.</i> , 2007 18 mois Californie (É-U)	10 097	1,8* 2,3**	4,0* 5,9**	8,9* 12,5**	16,5* 18,2**	7,0	14,7* 15,9**	0,74**
3	CFRA Dankert et Johnson, 2014 18 mois Californie (É-U)	11 444	1,2* 1,1**	2,9* 2,3**	6,4* 6,0**	6,5* 8,9**	2,8	5,3* 7,8**	0,154**

* Version originale de l'outil utilisé dans le SPJ

** Version révisée de l'outil (version améliorée proposée par les auteurs)

-- : non spécifié

DIFR (dispersion index for risk) : Indice de dispersion relatif au risque

PSEP (prognostic separation index) : Indice de distinction pronostique

Les résultats de l'étude de Johnson et Wagner [2003] indiquent que le classement est adéquat une fois associé à cet indicateur. Les niveaux de risque sont distincts : concernant la catégorie des familles à « risque très élevé », dans 31,7 % des cas, l'enfant sera placé en dehors du foyer familial en raison de maltraitance. En comparaison, la catégorie « à faible risque » affiche un taux de 1,1 %, la catégorie « à risque modéré », un taux de 2,9 %, et la catégorie « à risque élevé », un taux de 12,6 %. Précisons que la mesure PSEP est très élevée (30,6 %), ce qui traduit cet écart marqué entre les niveaux. Toutefois, la version révisée de l'outil qui est proposée par les auteurs de l'étude n'améliore pas davantage la précision du classement.

Les résultats de l'étude de Johnson et ses collaborateurs [2007] vont dans le même sens, soit un classement assez précis des familles, puisque la progression du taux de placement subséquent d'un enfant en suit les échelons (niveaux de risque). Les résultats obtenus à l'aide de la version révisée de l'outil ne sont que très légèrement meilleurs : la mesure PSEP passe de 14,7 % à 15,9 %, alors que la mesure DIFR est de 0,74, ce qui indique une très bonne

capacité de discrimination entre les niveaux de risque qui composent le classement des familles évaluées.

Enfin, les résultats de l'étude de validation la plus récente de l'outil CFRA, celle de Dankert et Johnson [2014], montrent également un classement assez précis des familles, mais dans de moindres proportions. Alors que la taille de l'échantillon est la plus élevée ($n = 11\,444$) des 3 études analysées ici, le taux de référence est le moins élevé (2,8 %). On observe également que les écarts entre les niveaux de risque sont moins marqués. Cela se reflète dans le résultat de la mesure de discrimination DIFR, qui est de 0,154, donc assez faible. Seul le résultat de la catégorie « à risque très élevé » obtenu à l'aide de la version révisée de l'outil est meilleur, soit 8,9 % comparativement au résultat de 6,5 % obtenu à l'aide de la version originale.

En résumé, la validité de l'outil CFRA est adéquate pour évaluer le taux de placement subséquent d'un enfant en dehors du foyer familial en raison de maltraitance, selon le classement établi. La précision du classement reste constante au fil des validations successives de l'outil, sans pour autant que les résultats ainsi obtenus s'améliorent de façon significative avec les années.

2.2.3.5 Validité prédictive d'outils d'évaluation de la maltraitance (abus et négligence)

Dans certaines études, aucune distinction n'est établie relativement aux résultats parmi les types de maltraitance, c'est-à-dire entre l'abus (tous types confondus) et la négligence. Les résultats de ces études sur la validité des outils considérés sont présentés dans le tableau 5. Ces résultats ne peuvent pas être comparés à ceux des études précédentes, lesquels portent sur l'abus (tous types confondus) ou uniquement sur l'abus physique, dans le cas de l'outil OFRA, mais non sur la négligence.

Tableau 5 Validité prédictive d'outils d'évaluation de la maltraitance (abus et négligence)

Outil Auteurs (année) Durée de suivi Territoire		Échantillon (taille)	Critères de validité prédictive				PSEP
			Se	Sp	Global (Se + Sp)	AUC	
1	FRAAN Wood, 2011 Aucun suivi Bournemouth (Royaume-Uni)	Abus physique et négligence : 58 enfants (dont 31 sont décédés)	88 %	----	----	----	----
2	C-CAPS Ezzo et Young, 2012 Aucun suivi Cleveland (É-U)	118 enfants	100 %	95 %	98 % (groupe 3 exclus) 71 % (y compris le groupe 3)	Groupe 3 : 57 %	----
3	CFRA Loman et Siegel, 2004 24 mois Minnesota (É-U)	15 100 familles	89,7 %	77,2 %	Taux d'erreur global = 33,1 % concernant 83,2 % des 15 100 familles sans services	----	----
4	CFRA Johnson, 2011 Californie (É-U)	7 685 familles, dont 6 543, sans services et 1 142 familles ayant des services	----	----	----	0,63 à p < 0,0005	0,336 (échantillon 1995-1997) 0,289 (échantillon 2000-2002)
5	CFRA Johnson, 2004 6 mois et 24 mois Californie (É-U)	Année 1995 : 2 511 familles Année 2000 : 6 543 familles	Classification des familles à risque de maltraitance future : En 1995 (n = 2 511) (après 24 mois) : En 2000 (n = 6 543) (après 24 mois) : Familles à faible risque : 7,7 % Familles à faible risque : 7,8 % Familles à risque modéré : 13,8 % Familles à risque modéré : 13,3 % Familles à risque élevé : 31,8 % Familles à risque élevé : 26,6 % Familles à risque très élevé : 44,3 % Familles à risque très élevé : 36,7 % En 2000 (n = 6 543) (après 6 mois) : Familles à faible risque : 3,9 % Familles à risque modéré : 5,3 % Familles à risque élevé : 10,8 % Familles à risque très élevé : 25,0 %				

n : nombre

Sp : Spécificité (critère de)

AUC : aire sous la courbe ROC

PSEP (prognostic separation index) : Indice de distinction pronostique

Se : Sensibilité (critère de)

Aux fins de sa thèse de doctorat, Wood [2011] a effectué une étude rétrospective de validation de l'outil **FRAAN** en l'appliquant à une cohorte établie à partir de rapports de révision de cas dont les résultats, portant sur l'analyse des facteurs de risque, ont été comparés aux données de deux études similaires et d'une étude sur des cas d'homicides d'enfants au Royaume-Uni. L'auteure met à l'épreuve la validation interne de cet outil. Elle émet l'hypothèse selon laquelle l'outil permet de prédire avec exactitude tous les cas

récurrents d'abus physique et de négligence ayant été auparavant désignés « à risque élevé » et « à risque très élevé ». Au premier regard, les résultats obtenus indiquent une excellente validité prédictive de l'outil. En effet, à l'aide de cet outil, l'auteure a cerné correctement la plupart des cas d'abus physique et les cas de négligence ayant auparavant été évalués comme étant « à risque élevé » et « à risque très élevé » avec une sensibilité (Se) égale à 88 %, comme l'indique le tableau 5. Toutefois, le critère de spécificité (Sp) n'a pu être évalué, car tous les cas « à risque très élevé » ont eu des conséquences majeures : mortalité ou blessures graves de l'enfant. En effet, l'échantillon de l'étude (cohorte) est composé de 58 enfants, dont 31 sont décédés. L'absence de résultats sur la spécificité ne permet donc pas d'établir l'exactitude de l'outil. De plus, selon Wood, l'outil ne peut distinguer, avec exactitude, les cas de décès d'enfants et les cas où il n'y a pas eu de décès d'enfant²⁷.

La mesure de la validité de l'outil **C-CAPS** par Ezzo et Young [2012] est aussi établie à partir des critères de sensibilité et de spécificité de l'outil. Ainsi, pour réaliser leur étude d'élaboration²⁸ de l'outil C-CAPS, un outil actuariel utilisé seul, ces auteurs disposent d'un échantillon de 118 cas divisés en 3 groupes. Le groupe 1 correspond aux cas de maltraitance non fondés (n = 34); le groupe 2 correspond aux cas fondés de maltraitance (n = 37) et le groupe 3 correspond aux cas d'allégation de maltraitance à évaluer (n = 47). Les auteurs de l'étude concluent à l'excellente capacité de l'outil à distinguer le groupe des enfants qui ne sont pas maltraités (groupe 1) du groupe des enfants maltraités (groupe 2). En effet, le taux de spécificité (Sp) est de 100 % et le taux de sensibilité (Se) est de 95 %, soit un résultat global de 98 %.

Comme l'indique le tableau 5, ce résultat global diminue à 71 % après l'introduction du groupe intermédiaire (groupe 3), qui est composé d'enfants qui font l'objet d'allégations de maltraitance, le groupe le plus représentatif des cas habituellement évalués dans les SPJ. Le groupe 3 est aussi le plus difficile à classer : sa précision (erreur dans la répartition des cas de faux positifs et des cas de faux négatifs) est de 57 %. Les cas ayant justement fait l'objet d'erreurs de classement (11 cas sur 47) sont répartis entre le groupe d'enfants maltraités (7 cas) et le groupe d'enfants non maltraités (4 cas). Les dimensions de l'outil ne permettent donc pas de prédire avec exactitude le classement de ceux-ci. Bref, les résultats sur la validité prédictive de l'outil n'appuient pas les conclusions formulées par ces auteurs.

À partir de deux sous-échantillons, l'un composé de familles ayant reçu des services d'un SPJ et l'autre non, et en considérant une durée de suivi de 6 mois et une autre de 24 mois suivant l'évaluation des cas, Johnson [2004] montre que le CFRA répond à 5 des 6 critères de validité énoncés par Altman et Royston [2000]. Selon Johnson, la structure de l'outil CFRA est encore valide en 2004, même s'il a été élaboré en 1995. Ainsi, l'échantillon de validation, composé de familles évaluées au cours de l'année 2000 est comparé à l'échantillon des familles qui ont été évaluées lors de l'élaboration de l'outil en 1995. Les résultats de cette comparaison entre les familles n'ayant pas reçu de services dans les 24 mois qui ont suivi leur évaluation sont présentés dans le tableau 5.

²⁷ Pecora et ses collaborateurs [2013] mentionnent qu'il existe une différence fondamentale entre la probabilité de récurrence de sévices corporels et le risque d'atteinte grave à l'intégrité physique de l'enfant due aux services corporels. Ces auteurs ajoutent que les outils d'évaluation du risque d'abus n'ont pas été conçus pour prédire le risque d'abus sévères ou mortels envers des enfants.

²⁸ L'annexe G présente une description des types et des stratégies de validation.

De 1995 à 2000, les résultats sont presque similaires : on observe une distinction assez nette entre les niveaux de risque. On observe également une relation positive entre le classement des familles à risque de récurrence et la récurrence de maltraitance future, en 2000 comme en 1995. Cette relation positive tient même après une période de suivi plus courte, soit seulement 6 mois après l'évaluation de l'échantillon destiné à valider l'outil. Les résultats obtenus par Johnson [2004] sur une période de 6 mois vont donc dans le même sens que ceux obtenus au cours de la période de 24 mois. Cependant, ils sont moins tranchés, étant donné ce court laps de temps pour évaluer la récurrence de la maltraitance. Johnson conclut que le CFRA est tout aussi valide (validité interne) que lors de son élaboration en 1995 : la progression des niveaux de risque relativement aux résultats sur la récurrence de la maltraitance est constante.

Pour leur part, Loman et Siegel [2004] soulignent que la validité de cet outil est adéquate pour prédire les nouveaux signalements et la réouverture d'un dossier dans les 24 mois suivant l'évaluation des familles. Bien que la relation entre les niveaux de risque de maltraitance et le taux de nouveaux signalements d'abus soit statistiquement significative ($\chi^2 = 392,7$; $p < 0,0001$), cette relation est faible (d de Somer = 0,116, $p < 0,0001$). Ces auteurs ont alors examiné les faux négatifs (FN) relatifs aux familles de la catégorie « à faible risque » et de la catégorie « à risque modéré » de récurrence qui n'ont pas reçu de services. Comme le montre le tableau 5, la sensibilité (Se) est de 89,7 %, donc tout près de la norme de référence (≥ 90). La spécificité (Sp) est de 77,2 %, soit légèrement sous la norme de référence (≥ 80). Par ailleurs, ils ont examiné les FN relatifs aux familles de la catégorie « à risque élevé » et de la catégorie « à risque très élevé » de récurrence qui n'ont pas reçu de services et qui n'ont pas récidivé. La combinaison des faux positifs (FP) et des FN donne un taux d'erreur global de 33,1 % concernant 83,2 % des 15 100 familles n'ayant pas reçu de services. L'outil CFRA n'a donc pu prédire correctement le risque de récurrence de maltraitance concernant environ 1 famille sur 3 qui n'a pas reçu de services. Cela explique pourquoi ces auteurs estiment que la validité de cet outil est adéquate, mais non excellente.

Enfin, dans une étude plus récente, Johnson [2011] évalue également la validité de l'outil CFRA. Considérant le calibrage de l'outil dans sa version originale, l'auteur explique que la validité de l'outil est adéquate, avec un coefficient de régression de 0,79 (valeur de $p < 0,0005$), ce qui témoigne d'une structure correcte de l'outil, sans être excellente, puisqu'elle est inférieure à 1,0. Quant à la capacité de discrimination de cet outil, l'aire sous la courbe ROC est égale à 0,63 (valeur de $p < 0,0005$), ce qui indique une efficacité moyenne (selon les normes établies, une valeur $\geq 0,70$ est jugée adéquate dans le contexte des SPJ). De plus, Johnson constate un écart statistiquement significatif ($p < 0,0005$) relativement à la maltraitance future lorsque des familles de la catégorie « à faible risque » de récurrence sont comparées aux familles de la catégorie « à risque élevé » de récurrence. Et cela vaut autant pour l'échantillon de 1995-1997, constitué à des fins d'élaboration de l'outil (PSEP = 0,336), que pour l'échantillon de 2000-2002, constitué à des fins de validation de l'outil (PSEP = 0,289). Ces résultats indiquent que la version originale de l'outil permet d'établir une meilleure discrimination entre les niveaux de risque. En effet, l'écart entre les catégories extrêmes de la classification, soit la catégorie « à faible risque » et la catégorie « à risque élevé », est légèrement plus accentué (0,05) que l'écart obtenu à l'aide de la version améliorée de l'outil, laquelle s'appuie sur l'échantillon de validation.

Dans l'ensemble, les résultats des études de validation des outils CFRAT, FRAAN, OFRA et CFRA indiquent une validité adéquate de ces outils pour évaluer le risque d'abus (tous types

confondus) et sa récurrence, considérant le lien établi entre le classement des familles dans différentes catégories (niveaux) de risque et des indicateurs d'abus futur (résultats d'intérêt), dans le contexte spécifique des SPJ. Dans la plupart des études, la version révisée des outils qui est proposée par la plupart des auteurs de ces études permet d'obtenir des résultats d'intérêt légèrement meilleurs que ceux obtenus à l'aide de la version originale de l'outil utilisé par les SPJ. Cette meilleure performance s'explique par le recalibrage de l'outil (réduction de quatre à trois niveaux de risque) et par une plus grande capacité de discrimination, c'est-à-dire une meilleure répartition des sous-groupes dans les niveaux de risque, et un plus grand écart entre les sous-groupes situés aux extrêmes de la classification, soit la catégorie « à faible risque » et la catégorie « à risque très élevé ».

Par ailleurs, concernant l'ensemble des outils évalués, les résultats sur la validité prédictive obtenus à partir des critères Se, Sp et d'aire sous la courbe varient selon les études considérées.

Aucune de ces études ne présente des résultats probants sur la validité prédictive d'un outil d'évaluation du risque d'abus physique.

En résumé, les résultats des études qui portent sur l'efficacité théorique (fiabilité, validité) des outils d'évaluation du risque d'abus physique chez l'enfant permettent de faire les observations suivantes :

- à l'exception de l'outil CAPI, les outils évalués ne permettent pas de distinguer les différents types d'abus (physique, sexuel ou psychologique). Par ailleurs, les deux études sur l'outil OFRA [Lee *et al.*, 2014; Shlonsky *et al.*, 2011] présentent des résultats spécifiques à l'abus physique.
- aucun des outils évalués ne satisfait à la fois aux critères d'une bonne fiabilité et aux critères d'une bonne validité.
- aucun des outils évalués ne permet de prédire si les parents (ou leurs conjoints) vont commettre ou commettre à nouveau un acte d'abus physique envers leur enfant, ou non.
- parmi les études qui portent sur l'outil contextuel CAPI et sur l'outil actuariel FRAAN, certaines concluent que la fiabilité de ces outils est bonne.
- les études sur les outils contextuels CAPI et AAPI-2 indiquent que leur validité n'est pas adéquate pour prédire l'abus physique. Le nombre élevé et la nature des facteurs de risque (facteurs statiques, facteurs dynamiques, facteurs de résilience) qui composent les échelles de ces outils expliquent, en partie, ce résultat.
- les outils actuariels, qui tiennent compte de la présence de facteurs de risque associés à l'abus physique, sont plus efficaces que les outils contextuels pour évaluer le niveau de risque d'abus physique ou sa récurrence.
- la validité prédictive des outils actuariels n'est pas toujours établie selon les critères de sensibilité et de spécificité ainsi que selon la mesure de l'aire sous la courbe. Les résultats obtenus dans certaines des études évaluées indiquent une bonne sensibilité ou une bonne spécificité de ce type d'outil, mais pas les deux à la fois.
- la validité prédictive des outils actuariels est souvent établie en deux temps. Au temps 1, les résultats de l'évaluation initiale du risque (combinaison des résultats globaux obtenus sur l'échelle de l'abus et sur l'échelle de la négligence) que présentent les familles évaluées sont considérés. Ces résultats servent à classer ces familles selon la

catégorisation du risque en divers niveaux (faible, modéré, élevé, très élevé). Au temps 2, la validité de l'outil s'appuie sur la précision de ce classement relativement aux indicateurs de résultats sur l'abus futur : taux de nouveaux signalements, taux de réévaluation de cas, taux de nouvelles allégations fondées, taux de placement subséquent d'un enfant en dehors du foyer familial.

- le calibrage et le recalibrage de l'outil et sa capacité de discrimination entre les niveaux de risque (à l'aide des mesures PSEP et DIFR) sont des éléments qui permettent d'évaluer la validité d'un outil actuariel.
- les auteurs de certaines études de validation de ces outils proposent une version révisée de l'outil original en usage dans certains SPJ. Cette révision consiste à recalibrer l'outil par une diminution du nombre de niveaux de risque, par l'ajout ou le retrait de facteurs de risque ainsi que leur pondération, et par la modification de la valeur limite (seuil) permettant de distinguer les niveaux de risque. Dans la version révisée, malgré un moins grand nombre de familles classées dans les catégories « à risque élevé » et « à risque très élevé », ces familles sont celles qui sont le plus susceptibles de commettre une récidive en matière d'abus.
- les résultats des études retenues indiquent que l'outil CFRA, appliqué dans plusieurs États américains et, dans une moindre mesure, l'outil OFRA – sa version adaptée, en usage en Ontario –, donnent des résultats adéquats sur le plan de la validité prédictive. Celle-ci repose sur l'association du classement des familles selon différents niveaux de risque de récidive d'abus à l'aide d'indicateurs de résultats sur un abus futur.

2.3 Questions 2 et 3 – Efficacité pratique des outils (utilité, applicabilité)

2.3.1 Revue de la littérature

L'efficacité d'un outil d'évaluation du risque d'abus physique envers des enfants ne doit pas être seulement théorique, elle doit également être pratique. À ce titre, l'outil doit être applicable dans les milieux d'intervention concernés et représenter une valeur ajoutée pour la pratique d'évaluation [Altman et Royston, 2000]. Ce passage de l'efficacité théorique à l'efficacité pratique de ce type d'outil a suscité la formulation des deux questions suivantes :

Question 2 - Ces outils dont la l'efficacité est reconnue (validité, fiabilité) **sont-ils utiles** à la prise de décision au moment d'évaluer des enfants qui font l'objet d'un signalement, ou d'un nouveau signalement d'abus physique?

Question 3 - Comment ces outils dont l'efficacité est reconnue s'**appliquent-ils** au contexte d'intervention?

La revue de la littérature grise a permis de retenir neuf documents qui traitent de l'efficacité pratique (utilité, applicabilité) des outils d'évaluation du risque d'abus physique. D'une part, il est question de sa valeur ajoutée à la pratique d'évaluation des cas. C'est sous l'angle de l'utilité ou de la pertinence de l'outil, de son acceptabilité (la manière dont il est perçu par ses utilisateurs) et de son apport à la prise de décision que cette question est soulevée. D'autre part, la question de l'intégration (applicabilité) de l'outil au contexte d'intervention est envisagée sous l'angle des conditions de son implantation. Considérant ces deux

préoccupations, certains atouts et limites associés aux outils de type actuariel, en particulier le CRC-SDM, sont mis en relief à partir de points de vue d'intervenants, d'agents de liaison et de gestionnaires de SPJ qui ont été interrogés sur celles-ci.

2.3.2 Utilité des outils

Concernant la première question, celle de la valeur ajoutée à la pratique d'évaluation des cas, les résultats des études menées par Kim et ses collaborateurs [2008] et par Jones et Beecroft [2008] abordent les atouts et les limites de l'outil CFRA et, plus largement, de l'ensemble intégré CRC-SDM, dont il est une composante. Kim et ses collaborateurs ont réalisé une étude en trois volets. Dans le premier volet, les auteurs dégagent trois principaux atouts de l'outil, à partir des propos de six gestionnaires et d'un superviseur : 1) une évaluation des cas mieux ciblée, plus objective et plus fiable; 2) un meilleur soutien dans la prise de décision des intervenants; et 3) une utilisation plus ciblée des ressources (diminution des réponses immédiates, un moins grand nombre de placements d'enfants, un plus grand nombre de cas orientés vers les ressources communautaires) contribuerait à un cheminement plus fluide des cas à travers le processus d'intervention. À ce propos, Jones et Beecroft [2008] mentionnent que les répondants (intervenants, superviseurs) désirent continuer d'utiliser cet outil, qui permet aux intervenants d'améliorer la fiabilité de leurs évaluations. Leurs superviseurs apprécient justement la constance des décisions plus « objectives » qui sont prises à l'aide du SDM. En effet, l'usage de cet outil permettrait d'obtenir une justification plus claire et défendable au moment d'évaluer l'importance de certaines allégations au sujet de la maltraitance d'enfants, réduisant ainsi l'incertitude liée à la subjectivité du jugement des intervenants. Ces mêmes auteurs décrivent les impacts de l'usage de cet outil sur les cas évalués sous l'angle du volume de cas à traiter, soit :

- une augmentation (surévaluation) de la proportion des familles « à risque élevé et très élevé »;
- une diminution de la proportion d'enfants évalués, dont la situation est jugée « sans danger immédiat »;
- une diminution des cas jugés « de priorité moyenne ».

En contrepartie à ces atouts, Jones et Beecroft [2008] soulignent un usage de l'outil qui varie selon les établissements et selon les utilisateurs, ce qui en limiterait la portée réelle. D'ailleurs, les résultats d'un sondage en ligne, effectué auprès de 84 intervenants et superviseurs ayant utilisé le CRC-SDM, indiquent que ce ne sont pas tous les intervenants qui utilisent l'outil ainsi qu'il est recommandé [Kim *et al.*, 2008]. Certains intervenants y ont recours une fois leur décision prise, pour confirmer leur évaluation, plutôt que de se servir de l'outil pour soutenir celle-ci. D'autres en font un usage détourné, de manière à obtenir les résultats voulus. Gillingham et Humphreys [2010] expliquent ainsi ce comportement :

- une surestimation du risque (trop de faux positifs) : Les familles sont souvent classées dans la catégorie « à risque élevé » en raison du caractère statique des facteurs de risque de l'outil, et ce, même si la situation de ces familles évolue.
- l'absence de considération des conditions de vie des familles : L'outil ne prend pas en compte la pauvreté et le stress que vivent les familles. La complexité de la vie des familles en proie à de multiples problèmes échappe à cet outil, dont les facteurs de risque sont limités.

Pour leur part, Kim et ses collaborateurs [2008] expliquent cet usage détourné de l'outil par le fait que certains intervenants voient dans le recours obligatoire à cet outil une atteinte à leur compétence professionnelle, ou encore le perçoivent comme un outil de reddition de compte au service de l'administration. À ce sujet, Gillingham et Humphreys [2010] précisent que ce type d'outil ne remplace pas l'expertise professionnelle des intervenants. Ils sont d'avis, comme dans la proposition de Munro [2005], qu'il faut miser davantage sur la formation des intervenants pour parfaire leur expertise plutôt que sur ce type d'outils. Ce dernier point est repris par McVeigh Palmer [2011]. Cette auteure précise que si les intervenants trouvent indispensable d'évaluer le risque de maltraitance à partir des facteurs déterminés par l'outil OFRA, ils préfèrent miser sur leur jugement professionnel et sur un outil complémentaire, le modèle *Signs of Safety*, qui est plus axé sur l'intervention. Elle conclut que les caractéristiques de l'outil OFRA (intégré au SDM) contredisent les avantages indiqués par son promoteur (gouvernement ontarien), et feraient abstraction du jugement professionnel des intervenants. D'une part, cet outil serait peu sensible aux ressources dont disposent les familles concernées et les facteurs de protection de celles-ci (ex. : leur réseau social) contre le risque de maltraitance. D'autre part, l'outil ne mesurerait pas adéquatement ce qu'il est censé mesurer, car il permet de détecter davantage de familles à risque élevé de récidive en matière de maltraitance.

2.3.3 Applicabilité des outils

Au sujet de la seconde question, soit l'applicabilité ou les conditions d'implantation d'un outil d'évaluation du risque d'abus physique, les résultats de groupes de discussion menés par Shlonsky et Lambert [2007] auxquels ont participé 92 intervenants dans 3 SPJ en Ontario (Canada) indiquent qu'en ce qui a trait à l'accessibilité et à l'utilisation de l'outil, ces derniers veulent un outil concis qui réduit la paperasse et évite la collecte en double des renseignements sur les cas évalués. Ces intervenants se prononcent en faveur d'un outil facile à interpréter et qui peut tracer un portrait d'ensemble du fonctionnement familial, sans demander trop de temps pour le remplir, ni être trop intrusif afin que les familles coopèrent. Par ailleurs, la formation du personnel et les conditions d'implantation de l'outil représentent deux aspects qui sont couramment mentionnés dans la littérature et présentés ci-dessous.

Formation du personnel

- encourager le personnel à reconnaître la pertinence et l'utilité de l'outil en matière de protection de la jeunesse, par exemple la manière dont l'outil s'intègre à la prise de décision et dans le processus d'intervention en général [Rycus et Hughes, 2003].
- informer le personnel sur ce que signifie la catégorisation du risque et de quelle manière celle-ci guide les actions à entreprendre par la suite [Shlonsky *et al.*, 2011; Shlonsky et Lambert, 2007; Johnson *et al.*, 2006].
- intégrer la démarche, les définitions des termes et les méthodes de calcul du risque à la démarche courante de révision des cas par les superviseurs et lors des rencontres d'équipe (ex. : exercices réalisés à l'aide de capsules cliniques, discussion sur des situations spécifiques) [Johnson *et al.*, 2006];
- aborder, si l'outil fait partie d'un ensemble SDM, les différents aspects (et réponses possibles) couverts par les composantes de ce modèle, leurs fonctions et leurs limites de manière à lever l'ambiguïté de la prise de décision qui en découle (biais et erreurs possibles) [Shlonsky et Lambert, 2007].

- partager la terminologie sur tout ce qui concerne ce type d'outil (notions, items, etc.) entre les membres qui composent le personnel du réseau de services concernés [Rycus et Hughes, 2003].

Mise en place de conditions générales d'implantation

- considérer que l'implantation de l'outil nécessite du temps, notamment en ce qui concerne la formation du personnel, afin de répondre aux politiques de l'établissement, et pour rendre les anciens dossiers conformes aux normes courantes [Jones et Beecroft, 2008; Kim *et al.*, 2008].
- favoriser l'implantation de l'outil dans les milieux d'intervention par le soutien des gestionnaires, un engagement constant de la part des superviseurs, l'encadrement du personnel concerné et la tenue de consultations (audits) sur l'évolution du processus de prise de décision à l'aide de ce type d'outil [Barlow *et al.*, 2012; Johnson *et al.*, 2006; D'Andrade *et al.*, 2005].
- définir clairement les procédures et les usages de ce type d'outil dans les milieux d'intervention au moyen de lignes directrices et d'une entente entre les établissements visés [Rycus et Hughes, 2003].
- encourager le personnel à utiliser l'outil conformément aux directives de ses concepteurs afin d'assurer la fiabilité des évaluations (l'outil ne sert pas à recueillir des données, mais à soutenir la prise de décision). Cette implantation ne peut se faire sans l'accompagnement de spécialistes de cet outil afin de l'adapter aux particularités du territoire où il sera implanté [Rycus et Hughes, 2003].
- s'assurer que l'outil ne sert pas à d'autres fins que celles pour lesquelles il a été conçu, à savoir estimer le risque d'abus physique (ou de maltraitance) envers des enfants et sa probabilité de récurrence ainsi que pour faciliter la prise de décision en ce sens [Shlonsky *et al.*, 2011; Rycus et Hughes, 2003].
- s'assurer que les résultats de l'évaluation à l'aide de l'outil donnent accès à des services en conséquence, selon le classement des familles à risque (le CRC-SDM inclut un cadre d'optimisation de l'offre de services et l'intensité requise de ces services) [Rycus et Hughes, 2003].
- Impliquer le personnel dans l'implantation graduelle de l'outil et son adaptation locale [Kim *et al.*, 2008; Johnson *et al.*, 2006].
- réaliser un projet pilote pour tester l'outil avant son implantation, afin de mesurer l'écart entre les résultats de l'évaluation du risque obtenus à l'aide de l'outil et la politique interne de l'organisation (établissement) en matière de gestion du risque [Barlow *et al.*, 2012; Shlonsky *et al.*, 2011].
- revalider et recalibrer périodiquement l'outil afin de s'assurer de l'exactitude du classement des familles à risque de récurrence [Shlonsky et Lambert, 2007; Johnson *et al.*, 2006].

2.4 Consultation des milieux d'intervention

2.4.1 Groupes de discussion focalisée – Synthèse des résultats²⁹

Les préoccupations soulevées dans la littérature grise au sujet de l'efficacité pratique d'outils d'évaluation du risque d'abus physique (et de la maltraitance en général) ont servi à formuler les questions du guide d'entretien de quatre groupes de discussion focalisée, plutôt que les trois prévus initialement³⁰.

Un premier groupe rassemblait des intervenantes de deux CSSS qui travaillent auprès de la clientèle « enfants et adolescents » (éducatrices spécialisées, travailleuses sociales, psychologues et infirmières). Un deuxième groupe rassemblait des intervenantes (éducatrices spécialisées et travailleuses sociales des secteurs de l'urgence sociale, de la prévention et de la vérification sur le terrain (PVT), de la réception et du traitement des signalements (RTS) et de l'évaluation) en provenance de deux CJ. Un troisième et un quatrième groupe rassemblaient des gestionnaires (chefs de service à l'évaluation, chefs de service à l'urgence sociale, directeur adjoint de la protection de la jeunesse) et des agents de liaison de deux CJ. Au total, 21 personnes ont participé à ces groupes de discussion focalisée (voir l'annexe H).

Outre des questions initiales sur la nécessité d'évaluer le risque d'abus physique chez les enfants, sur la connaissance de ce type d'outil et son usage éventuel, les participants ont été questionnés dans le but d'obtenir des réponses aux sous-questions suivantes :

- Est-ce que l'outil pourrait améliorer l'évaluation des cas (utilité, pertinence, etc.)?
- Est-ce que l'outil soutiendrait la prise de décision?
- Quels seraient les critères d'acceptabilité de l'outil pour ses utilisateurs?
- Quelles seraient les conditions d'implantation de l'outil?

2.4.1.1 Valeur ajoutée (utilité, pertinence, acceptabilité) de l'outil d'évaluation du risque

Les résultats obtenus à partir de la synthèse des propos des participants à ces groupes de discussion indiquent un intérêt pour ce type d'outil, à certaines conditions incontournables. Selon ces participants, l'enfant qui présente un risque d'être victime d'abus est celui qui en tirerait le plus grand avantage pour sa santé et son bien-être futur. Les intervenants en CJ ou en CSSS en tireraient également avantage dans l'exercice de leur travail auprès de la clientèle, mais selon un usage différent.

Présentement, en l'absence d'un outil d'évaluation, le recours aux collègues pour juger s'il y a lieu de signaler, à un CJ, un cas (de risque) d'abus physique est une pratique courante des intervenants des CSSS. Selon ces derniers, puisqu'il n'y a pas toujours de mécanisme de consultation officiel entre les CSSS et les CJ, l'information transmise entre ces établissements sur une situation d'abus physique peut s'avérer plus ou moins complète. De leur côté, les

²⁹ Cette section est rédigée à partir de la synthèse, réalisée par Isabelle Beaumier (CIQ-IU), des propos des participants à ces groupes de discussion focalisée qu'elle a également coanimés.

³⁰ En raison d'un problème de disponibilité des gestionnaires de deux CJ, deux groupes de discussion focalisée ont été réalisés avec ces gestionnaires plutôt qu'un seul.

milieux scolaires auraient tendance à s'adresser aux CSSS plutôt qu'aux CJ lorsqu'il s'agit de signaler un cas de ce type. En CJ, le jugement de l'intervenant et le soutien apporté par les spécialistes en activités cliniques, les superviseurs, les coordonnateurs ou les chefs de service et les experts indépendants (ex. : des médecins) sont les principaux recours pour appuyer la prise de décision d'un cas de ce type. L'usage obligatoire du SSP dans les CJ n'est pas toujours apprécié par les intervenants, au point que certains d'entre eux cherchent à en contourner les normes d'usage. Plusieurs croient que cet outil devrait servir uniquement aux nouveaux intervenants, par exemple au cours de leur première année de service.

Des intervenants sont d'avis que l'évaluation du risque d'abus physique n'est pas « tangible », il s'agit davantage d'une présomption qui donne lieu à différentes façons d'envisager ce problème. Cette situation ouvre la porte aux interprétations multiples du problème et, par conséquent, à l'émergence de « zones grises » en matière d'intervention. Pour cette raison, l'usage d'un outil d'évaluation du risque d'abus physique ne peut se faire sans disposer d'une définition partagée de l'abus physique (et autres types de maltraitance). Lever cet obstacle du partage d'une définition commune du risque d'abus physique est justement l'un des apports possibles de ce type d'outil. Une formation spécifique sur ce type de maltraitance comblerait, en partie, cette lacune. Les autres atouts qu'il représente pour le personnel des CJ et des CSSS sont résumés³¹ comme suit :

Participants des CJ

« Un outil destiné à cette fin peut donc aider à uniformiser les perceptions, à “objectiver” les situations et à standardiser la pratique. Il peut aider plus spécifiquement à dénouer des situations complexes où les données n'indiquent pas clairement la présence d'abus physique, mais où le doute persiste. Intervenants et gestionnaires estiment qu'un tel outil peut aider à détecter les cas d'abus physique, en plus de permettre d'obtenir davantage d'information et d'approfondir certaines situations où le risque d'abus physique est très élevé. On souligne également la possibilité d'éviter qu'un enfant soit de nouveau victime d'abus physique (récurrence), de permettre d'évaluer ce risque dans la fratrie et de fermer un dossier avec une plus grande certitude et sans être intrusif. Par ailleurs, plusieurs croient que l'outil procurerait à l'intervenant davantage de “crédibilité” auprès du client, en plus de l'aider à articuler son raisonnement au tribunal. Également, ce type d'outil pourrait aider les nouveaux intervenants, en leur permettant d'acquérir plus d'assurance et d'autonomie dans leur pratique d'évaluation ».

Participants des CSSS

« [...] l'accent est mis sur le potentiel de détection des cas d'abus physique grâce à ce type d'outil et le fait de pouvoir agir en prévention. On croit que cela peut aider à uniformiser les processus, à utiliser un langage commun et à mieux connaître les indices de présence, de risque ou de récurrence de l'abus physique. Les intervenantes rapportent qu'un tel outil pourrait aider à divulguer les éléments pertinents lors d'un signalement et potentiellement, éviter de signaler inutilement. Également, l'outil représente une occasion de travailler en partenariat et en continuité avec les CJ. D'après ces participants, les intervenants des CJ sont

³¹ Ces extraits sont tirés de la synthèse réalisée par Isabelle Beaumier (CJQ-IU).

les premiers à tirer profit d'un tel outil, suivi des intervenants en CSSS et, en troisième lieu, du personnel en provenance du milieu scolaire ».

Certaines limites associées à l'utilisation de ce type d'outils ont également été mentionnées, dont les suivantes :

- une évaluation « en vase clos », loin des discussions cliniques et sans le recours au jugement professionnel;
- l'usage obligatoire de l'outil et les résistances possibles qui en résulteraient;
- le danger de chevauchement ou de répétition des mesures effectuées à l'aide d'outils existants (ex. : outils d'autres types, grilles maison);
- l'absence de prise en considération de facteurs dynamiques par l'outil et l'accent mis sur des critères statiques (ex. : les condamnations des parents, le nombre de signalements antérieurs à un SPJ) qui écartent tout changement survenu dans la situation familiale;
- la charge de travail supplémentaire occasionnée par l'introduction d'un nouvel outil;
- la crainte de représailles de la part du parent, dans la mesure où le questionnaire de l'outil serait rempli avec ce dernier.

2.4.1.2 Applicabilité : les conditions d'implantation de l'outil

Au sujet du moment opportun de l'application de l'outil, relativement aux étapes d'intervention en vertu de la LPJ du Québec, le personnel des CJ est d'avis que ce type d'outil devrait être utilisé :

- à l'étape de l'évaluation, où les situations sont parfois complexes et où la prise de décision est difficile;
- à l'étape de l'orientation et du choix du régime et des mesures;
- à d'autres étapes du processus d'intervention, notamment avant la fermeture du dossier concernant l'application des mesures, pour évaluer le risque de récurrence de l'abus physique et afin de vérifier si l'intervention réalisée a permis de diminuer le degré de risque d'abus physique, ce qui aiderait le réviseur quant à l'orientation à prendre dans le dossier.

Selon le personnel des CSSS, le recours à ce type d'outil dans leur établissement serait approprié à l'étape de l'accueil, à l'étape de l'évaluation (collecte des renseignements auprès de la clientèle) et même au guichet d'accès aux services afin de mieux cerner les besoins, en particulier ceux de la clientèle du programme SIPPE³². Cet outil pourrait s'avérer utile dans le contexte du plan de services individualisé (PSI), au moment de déterminer les besoins d'une intervention relative au risque d'abus physique.

³² En mars 2004, le ministère de la Santé et des Services sociaux (MSSS) diffuse officiellement le document Les services intégrés en périnatalité et pour la petite enfance à l'intention des familles vivant en contexte de vulnérabilité – Cadre de référence [MSSS, 2004]. Ces services sont le résultat de l'intégration de deux programmes offerts par les équipes Famille-Enfance-Jeunesse (FEJ) en CLSC : Naître égaux – Grandir en santé (NE-GS) et le Programme de soutien aux jeunes parents (PSJP). Les services intégrés en périnatalité et pour la petite enfance (SIPPE) à l'intention des femmes enceintes, des bébés à naître, des parents et des enfants d'âge préscolaire qui ont dans une situation les rendant vulnérables comprennent deux composantes : l'accompagnement des familles et le soutien à la création d'environnements favorables à la santé et au bien-être de ces familles.

En ce qui a trait aux conditions propices à l'intégration de ce type d'outil dans les milieux d'intervention, les intervenants des CJ insistent sur sa simplicité (un outil concis, précis, qui offre une interprétation sans équivoque des données), sur son libre usage et sur le fait qu'il permet une évaluation du risque sans pour autant faire abstraction des caractéristiques de la situation familiale. Pour leur part, les gestionnaires des CJ insistent sur la pertinence de ce type d'outil pour soutenir la réflexion et la prise de décision sur un cas; un outil qui se démarque par rapport aux outils existants et qui est adapté aux caractéristiques du client (ex. : membre d'une communauté culturelle).

Tous les participants aux groupes de discussion focalisée s'entendent sur la nécessité d'une formation visant à mieux connaître et à adopter l'outil et sur l'importance d'assurer un suivi afin de vérifier l'évolution de son utilisation et de ses impacts suivant son implantation, selon des lignes directrices clairement définies. Selon eux, des personnes pivots (chef de service, intervenants expérimentés, spécialistes en activités cliniques, etc.) devraient s'assurer de l'application adéquate de l'outil. Certains intervenants mentionnent qu'il faut éviter de choisir un outil « autogéré », car plusieurs parents éprouvent des difficultés de compréhension. Ils pourraient aussi être tentés de répondre aux questions que comprend l'outil de manière à éviter l'ouverture d'un dossier ou encore cesser d'utiliser les services qui leur sont déjà offerts.

Selon certains intervenants de CSSS, le succès de l'implantation de ce type d'outil réside dans le partenariat établi avec les CJ, les milieux scolaires; le choix de l'outil, les orientations et les résultats escomptés devraient alors faire l'objet d'une entente entre les CSSS et les CJ. Si l'outil permettait aux intervenants des CJ d'évaluer le risque d'abus physique, il servirait aux intervenants des CSSS à orienter l'intervention. Aussi, un tel outil permettrait l'utilisation d'un langage et de critères communs relatifs aux signalements. Enfin, la présentation de l'outil devrait être adaptée aux besoins des intervenants des CSSS et prendre, par exemple, la forme d'un aide-mémoire simple et concis. Ce dernier ferait référence à des données objectives qui permettent de repérer des indices associés au risque et à la récurrence de l'abus physique.

En résumé, aux yeux des participants, la valeur ajoutée de cet outil sur le plan de son acceptabilité par les milieux d'intervention, dans l'objectif d'améliorer l'évaluation des cas et de soutenir la prise de décision qui s'ensuit, se révèle par :

- l'avantage pour la santé et bien-être l'enfant;
- une définition, un langage et des critères communs d'évaluation du risque d'abus physique (pour contrer les interprétations variables du phénomène);
- l'objectivation des situations (données « tangibles », articulation du raisonnement, transparence de l'argumentation) et une documentation plus complète des cas évalués;
- une standardisation de la pratique par l'uniformisation du processus d'évaluation (éviter les « zones grises »);
- l'acquisition d'une plus grande assurance dans la pratique d'évaluation (dénouer les situations complexes, statuer sur l'état de compromission ou non);
- une meilleure crédibilité de la démarche auprès des tribunaux;
- l'orientation de la prise de décision en faveur d'une priorisation des services et de l'intensité de l'offre de service.

En ce qui concerne les limites associées à cet outil dans la littérature, les participants ont mis l'accent sur les points suivants :

- l'outil ne doit pas se substituer au jugement professionnel;
- l'outil ne met pas suffisamment l'accent sur la situation familiale (contexte de vie) et les facteurs de résilience des familles évaluées;
- la charge de travail supplémentaire occasionnée par l'introduction d'un nouvel outil;
- les résistances possibles à utiliser un outil de ce type si son usage devenait obligatoire.

Au moment d'aborder l'applicabilité (conditions d'implantation) de ce type d'outil, les intervenants de CJ se sont prononcés en faveur d'un outil simple à utiliser, concis et qui pourrait être utilisé à différentes étapes du processus d'intervention, sans pour autant que son usage soit obligatoire, contrairement au SSP. La formation et la supervision des utilisateurs de l'outil ainsi que le soutien fourni par les gestionnaires des établissements concernés reviennent également dans leurs propos. De leur côté, les gestionnaires de CJ ont suggéré que l'outil soutienne la réflexion menant à la prise de décision sur les cas évalués, se démarque par rapport aux outils existants (éviter le chevauchement et les répétitions inutiles) et soit adapté aux caractéristiques du client (ex. : un membre d'une communauté culturelle). Dans l'ensemble, les participants se sont prononcés en faveur d'un outil qui ne se substitue pas au jugement professionnel, qui requiert une formation visant à le connaître suffisamment pour évaluer la nature et l'intensité des services à offrir à la clientèle. Enfin, parmi les conditions favorables à l'intégration de ce type d'outil, la nécessité d'une entente entre CSSS et CJ sur le choix de l'outil, les orientations et les résultats escomptés est soulignée.

2.5 Question 4 – Consultations des parties prenantes

Un comité de suivi et des experts indépendants ont accompagné l'INESSS dans les différentes étapes menant à la réalisation de cet avis. Composé de 14 représentants de divers organismes gouvernementaux, d'établissements et d'ordres professionnels, le comité de suivi s'est rencontré à 4 reprises; 3 experts indépendants ont aussi contribué, par l'entremise d'échanges de courriels, à la production du présent avis. Au cours des rencontres, les membres du comité de suivi ont soulevé un certain nombre de préoccupations sur les outils d'évaluation du risque d'abus physique.

Une première préoccupation est que ce type d'outil puisse servir aux intervenants des CJ et à ceux des CSSS, de manière à faciliter la continuité des services offerts à la clientèle qui passe par ces deux types d'établissements et parfois de l'un à l'autre. Or, les outils d'évaluation du risque d'abus physique repérés dans la littérature, surtout les outils actuariels, sont rarement utilisés ailleurs que dans les SPJ, lesquels s'apparentent, dans le contexte québécois, aux CJ.

Une seconde préoccupation porte sur le moment opportun où ce type d'outil peut être utilisé. À ce sujet, la revue systématique de la littérature scientifique a permis d'apprendre qu'un système d'évaluation intégrée SDM comprend plusieurs outils qui sont adaptés aux différentes étapes d'une intervention. Dans la littérature, l'outil d'évaluation du risque d'abus physique est surtout utilisé à l'étape de l'évaluation des cas.

Une troisième préoccupation concerne le danger de chevauchement ou de répétitions inutiles dans l'usage d'outils d'évaluation, voire un double emploi. Ce danger dépend des outils déjà utilisés dans un SPJ et de leur validité.

Une quatrième préoccupation concerne le rôle du jugement professionnel dans la prise de décision, plutôt que d'avoir recours à un outil d'évaluation du risque d'abus physique. Sur ce point, outre l'intégration de ce type d'outil à un SDM, l'outil actuariel a été conçu pour être accompagné d'un outil qui prend en compte le contexte familial de la clientèle.

Une cinquième préoccupation porte sur l'efficacité de ce type d'outil et son application en fonction de l'origine ethnique de la clientèle. Plusieurs études de validation de ce type d'outil intègrent l'origine ethnique en tant que variable à examiner au moment de tester son efficacité, car elle est susceptible d'influencer le classement du risque d'abus physique. Les résultats des études montrent que la version révisée d'outils actuariels, qui est parfois proposée par les auteurs de ces études, contribue à améliorer cette situation.

Une sixième préoccupation concerne l'augmentation possible de la tâche de l'intervenant suivant l'ajout d'un outil de ce type. À ce sujet, la littérature consultée indique qu'en contrepartie d'une augmentation du nombre d'ouvertures de dossier concernant les familles classées « à risque élevé » ou « à risque très élevé » de récurrence en matière d'abus physique, on assiste à une diminution du nombre de dossiers à traiter concernant les familles classées « à faible risque » et « à risque modéré ».

Une septième préoccupation porte sur l'adaptation d'un outil d'évaluation du risque d'abus physique au contexte québécois. Certains membres du comité soulèvent le problème de l'équité des ressources entre les CJ, du besoin de formation et de soutien des intervenants si un outil de ce type était implanté. Sur ce point, les membres du comité indiquent la nécessité de documenter les conditions d'implantation de ces outils.

De leur côté, les experts indépendants consultés sont d'avis que d'autres études importantes auraient pu être incluses dans la revue systématique, dont celles qui abordent la validité prédictive d'outils actuariels par rapport aux décisions intuitives et aux pratiques parentales relatives à l'abus physique (ex. : les études sur l'outil Parent-Child Conflict Tactics Scales). Aussi, au Québec, l'application d'un outil actuariel intégré à un ensemble de type SDM poserait un problème majeur étant donné : 1) l'usage obligatoire du SSP et 2) la structure et l'organisation des services de protection de l'enfance qui sont spécifiques de cette province. Une réflexion approfondie sur ces deux points est donc proposée.

3 DISCUSSION

3.1 Efficacité théorique (fiabilité, validité) des outils identifiés

3.1.1 Fiabilité des outils

Parmi les 21 études retenues en vue de répondre à la question sur l'efficacité théorique (fiabilité, validité), seulement 9 études contiennent des mesures de fiabilité. Ces études portent sur 6 des 7 outils évalués; la fiabilité de l'outil CFRAT n'a pas été mesurée par Coohy et ses collaborateurs [2013]. Le fait que des mesures de fiabilité n'accompagnent pas les mesures de validité d'un outil peut étonner, surtout si l'outil a fait l'objet d'ajout ou de retrait de facteurs de risque après un exercice de validation ou de revalidation de ces outils, depuis les douze dernières années. Il faut dire que la mesure de la fiabilité est effectuée lors de la conception de l'outil. Or, la période couverte par les études à retenir aux fins de la revue systématique de la littérature scientifique en vue de produire le présent avis se situe au-delà de la période d'élaboration des outils évalués. En effet, la plupart de ces outils ont été conçus au cours des années 1980 et des années 1990, alors que cet avis porte sur les études publiées de 2003 à 2013 (incluant une mise à jour de la revue de la littérature en 2014). D'ailleurs, les résultats sur la fiabilité et la validité d'un certain nombre d'outils datent, pour la plupart, des années 1980 aux années 2000. De plus, la fiabilité de certains outils a été mesurée dans les milieux d'intervention qui les utilisent, mais les résultats de ces mesures ne sont pas toujours accessibles.

Cela dit, les résultats de deux études où la fiabilité de l'outil a été mesurée indiquent une bonne stabilité et une bonne validité interne du CAPI (outil contextuel). Une autre étude montre la fiabilité du FRAAN (outil actuariel) au Royaume-Uni, alors que cet outil a été élaboré aux États-Unis, ce qui laisse supposer une capacité de généralisation de son application (validité externe).

3.1.2 Validité des outils

3.1.2.1 Précision sur la validité prédictive des outils d'évaluation du risque d'abus physique

Aucun des outils évalués ne permet de prédire avec suffisamment d'exactitude qu'un parent (ou son conjoint) ayant déjà abusé physiquement de son enfant commettra à nouveau un tel acte dans le futur. En fait, la probabilité qu'un parent récidive est plus faible que celle qu'il ne récidive pas.

Le faible de taux de prévalence (taux de référence) de l'abus physique envers des enfants dans la population en général et, en particulier, les cas d'atteinte physique grave à l'intégrité physique de l'enfant, ajoute à cette difficulté. Cela peut expliquer, en partie, l'absence, dans les études considérées, d'une distinction claire entre les différents types de problèmes de maltraitance que ce type d'outil prend en compte : l'abus (physique, sexuelle, psychologique) et la négligence. Non seulement ces problèmes sont interreliés, car à plus ou moins long terme, l'abus mène à la négligence et vice versa, mais la proportion des cas d'abus physique

est habituellement bien inférieure à la proportion des cas de négligence. À l'exception de deux études qui portent sur l'outil OFRA, peu d'auteurs s'emploient à distinguer l'abus physique des autres types d'abus dans leurs études. D'ailleurs, dans le cas des outils actuariels, les résultats globaux sont ceux obtenus sur l'abus, tous types confondus, et sur la négligence. Cela assure un taux de référence suffisant pour effectuer des calculs statistiques à partir de bases de données des SPJ, d'où proviennent d'ailleurs les résultats de la plupart des études évaluées aux fins de la production du présent avis.

À cela s'ajoute le fait qu'aucun outil de référence ne permet de comparer l'efficacité d'un autre outil dans le domaine de l'évaluation du risque d'abus physique à l'égard des enfants.

3.1.2.2 Validité prédictive et prise de décision

En ce qui a trait au lien entre la validité prédictive et la prise de décision (dans un contexte d'incertitude) résultant d'une évaluation des caractéristiques de la clientèle en vue de déterminer le risque d'un abus physique futur, l'erreur associée à cette prise de décision consiste :

- a) soit à juger indûment des familles n'ayant pas commis un acte d'abus physique envers l'enfant comme étant à risque de récurrence (faux positifs);
- b) soit à juger indûment des familles ayant commis un acte d'abus physique envers l'enfant comme n'étant pas à risque de récurrence (faux négatifs).

Ce contexte est illustré dans la figure 2.

Figure 2 Validité prédictive et prise de décision

Événement Prise de décision	Famille ayant commis un acte d'abus physique envers l'enfant	Famille n'ayant pas commis un acte d'abus physique envers l'enfant
Famille à risque de récurrence (+)	Vrais positifs (VP) (Se $\geq$ 0,90)	Faux positifs (FP)
Famille n'étant pas à risque de récurrence (-)	Faux négatifs (FN)	Vrais négatifs (VN) (Sp $\geq$ 0,80)

Sensibilité (Se) $VP/(VP+FN)$ Spécificité (Sp) $VN/(VN+FP)$

La *validité prédictive* (VP) de ce type d'outil est établie à partir de deux critères. La *sensibilité* (Se) de l'outil, qui réside dans sa capacité à déterminer correctement si une famille (parent ou son conjoint) a commis un acte d'abus physique envers son enfant et est à risque de récurrence dans une proportion de 90 % et plus (Se $\geq$ 0,90), soit au moins 9 familles sur 10. Par ailleurs, la *spécificité* (Sp) de l'outil, qui réside dans sa capacité à déterminer correctement si une famille (parent ou son conjoint) n'a pas commis un acte d'abus physique envers l'enfant et n'est pas à risque de récurrence dans une proportion de 80 % et plus (Sp $\geq$ 0,80), soit au moins 8 familles sur 10.

Aucun des résultats obtenus sur la validité des sept outils évalués ne satisfait à la fois à ces deux critères psychométriques, reconnus en matière de détection ou d'évaluation diagnostique [Noonan *et al.*, 2013], qui permettraient de prédire le risque de récurrence d'un abus physique à l'égard des enfants.

Plusieurs des études menées sur la validation de ces outils concluent pourtant que leur validité prédictive s'avère suffisamment adéquate pour évaluer le risque d'un abus physique futur. Comment expliquer cela?

Une première réponse possible à cette question concerne le contexte d'application de ces critères de validité prédictive. Selon White et Walsh [2006 : 10], aucun outil de ce type ne peut être fiable à 100 %, notamment en raison de la rareté des cas d'atteinte grave ou de mortalité de l'enfant résultants d'un abus physique. Ces auteurs situent entre 70 % et 80 % l'exactitude des outils de type actuariel. D'autre part, Coohy [2013 : 153-4] mentionne que l'exactitude de ce type d'outil ne dépasse guère 70 % dans les études où les données recueillies proviennent de bases de données des SPJ. Il ajoute que les résultats obtenus à l'aide de tels outils permettent de classer les familles évaluées selon trois ou quatre niveaux de risque. Cette démarche vise à soutenir la prise de décision sur l'offre de service, ce que les critères d'exactitude (Se, Sp, aire sous la courbe) ne permettent pas de déterminer.

Par ailleurs, afin d'améliorer la validité d'un outil déjà en usage dans un SPJ, et de s'assurer de sa conformité par rapport aux caractéristiques des populations locales, certains auteurs proposent une version révisée de l'outil. Il s'agit de recalibrer l'outil en remplaçant certains facteurs de risque qui le composent et qui sont moins associés aux résultats d'intérêt. Une autre stratégie consiste à diminuer le nombre de niveaux de risque, en le faisant passer de quatre catégories (faible, modéré, élevé, très élevé) à trois catégories (faible, modéré, élevé). Cette démarche s'explique en raison du faible écart, souvent observé, entre les résultats de la catégorie « à risque élevé » et les résultats de la catégorie « à risque très élevé ». La modification apportée en ce sens améliore effectivement la validité de l'outil [Coohy *et al.*, 2013; Johnson *et al.*, 2007; Johnson *et al.*, 2006; Johnson et Wagner, 2003]. En effet, un outil ainsi recalibré classe un moins grand nombre de familles dans la catégorie à « risque élevé » et « à risque très élevé », et augmente l'écart entre les niveaux de risque de la classification. En conséquence, les résultats obtenus à l'aide de la version révisée de l'outil indiquent que les familles classées dans la catégorie « à risque élevé et très élevé » de récurrence d'un abus physique à l'égard de leur enfant sont effectivement celles qui récidivent. Ces familles font l'objet soit de nouveaux signalements, de réévaluations de leur cas, de nouvelles allégations fondées, ou leur enfant fait l'objet d'un placement subséquent en dehors du foyer familial.

Enfin, la version révisée de l'outil assure une meilleure représentativité des familles d'origine ethnique distincte. En effet, cette caractéristique change parfois les résultats de la classification du risque d'abus physique. Par exemple, aux États-Unis, la surreprésentation ou la sous-représentation, selon le cas, des personnes issues de communautés autochtones dans les catégories « à risque élevé » et les catégories « à risque très élevé » de commettre un abus physique dans le futur, soulève le problème d'une répartition équitable des familles d'origine ethnique distincte selon les différentes catégories de risque [Johnson, 2011; Johnson, 2004; Loman et Siegel, 2004]. Or, une version révisée d'un outil permet habituellement d'améliorer sensiblement les résultats de l'outil afin de résoudre ce problème.

Dans l'ensemble, les résultats des études de validation font état – tant en ce qui concerne la version originale de l'outil, en usage dans un SPJ, qu'en ce qui concerne la version révisée proposée par les auteurs de la validation – d'une bonne association entre les résultats globaux du classement des familles à risque et la probabilité d'un abus futur concernant les familles n'ayant pas reçu de services d'un SPJ au cours des 6 à 24 mois, suivant leur évaluation initiale à l'aide de l'outil. La proportion des familles jugées à risque de commettre un abus dans le futur suit effectivement la catégorisation des niveaux de risque.

On constate également que la durée de la période de suivi des échantillons de population étudiés influence les résultats obtenus; plus la durée de suivi est longue, meilleurs sont les résultats, car plus le risque de récurrence d'un abus physique augmente (surtout si l'enfant est en bas âge). En effet, le taux moyen de nouveaux signalements est souvent très faible lorsqu'il est calculé sur une courte période de temps. Aussi, certains auteurs d'études de validation mentionnent qu'une période de suivi d'au moins 24 mois est préférable pour évaluer la validité de l'outil.

En résumé, les résultats rapportés dans les études de validation d'outils d'évaluation du risque d'abus physique envers les enfants montrent la pertinence d'utiliser un outil de type actuariel plutôt qu'un outil de type contextuel pour évaluer le risque d'abus physique. De façon générale, l'outil actuariel est concis, améliore la précision de l'évaluation des cas, documente davantage ces derniers et, surtout, permet de cibler de façon objective les familles les plus à risque de récurrence. Il permet également d'orienter les efforts, les ressources et les services à offrir en conséquence. Le soutien à la prise de décision est davantage appuyé par l'organisation qui, dans une perspective de gestion du risque, tire profit de l'utilisation de ce type d'outil. Mentionnons qu'à défaut d'offrir une excellente validité prédictive, la plupart des outils actuariels considérés dans le présent avis ont une validité suffisamment adéquate pour justifier leur usage dans un processus d'évaluation globale.

L'outil CFRA ayant fait l'objet de validations successives et d'une large diffusion aux États-Unis et à l'étranger (Canada, Australie) lui assurant une certaine validité externe, les mesures de calibrage et de discrimination de cet outil donnent des résultats probants. La version révisée de cet outil qui est proposée dans certaines études permet d'obtenir des résultats légèrement supérieurs concernant la relation entre le classement des familles à risque et les indicateurs d'abus physique futur, en accentuant la proportion des familles dans les deux catégories extrêmes de la classification, soit « à faible risque » et « à risque très élevé ». La plus récente étude de validation de cet outil [Dankert et Johnson, 2014] fournit des résultats probants en ce qui concerne le taux de réévaluation de cas et le taux de placement subséquent d'un enfant en dehors du foyer familial. Sans atteindre des résultats aussi significatifs, l'outil OFRA, qui est la version de l'outil CFRA adaptée au contexte de l'Ontario et en usage dans cette province, montre également une validité adéquate.

3.2 Efficacité pratique (utilité, applicabilité) des outils

Les participants aux groupes de discussion focalisée se disent favorables à l'implantation d'un outil d'évaluation du risque d'abus physique à l'égard des enfants. Cet intérêt s'explique par l'harmonisation des pratiques d'évaluation que permet ce type d'outil et par la prise de décision qui en résulte. À leurs yeux, disposer d'un tel outil favoriserait la convergence dans la manière de concevoir l'abus physique, de détecter ses manifestations, de standardiser le processus d'évaluation et de soutenir la prise de décision. L'usage de cet outil pourrait rassurer l'intervenant, appuyer son argumentation menant à la prise de décision de manière à adapter l'offre de service en fonction du niveau de risque des familles évaluées. L'objectivité de la démarche d'évaluation à l'aide de l'outil augmenterait la crédibilité de l'évaluation réalisée par l'intervenant. Celui-ci se verrait appuyé dans cette démarche par le personnel de son milieu d'intervention et celui d'autres organisations (CJ, CSSS, partenaires), une fois l'outil diffusé auprès de ces derniers. Enfin, la réduction de la récurrence de l'abus physique, dans une perspective de gestion du risque, s'avère un objectif important à atteindre de l'avis de ces acteurs sur le terrain.

En cela, tant les résultats témoignant de la validité adéquate de certains outils actuariels décrits dans le présent avis que les points de vue exprimés par les acteurs interrogés dans les études qualitatives consultées répondent à ces attentes. En effet, ces outils peuvent contribuer à une meilleure gestion du risque en soutenant la prise de décision de manière à diriger les ressources et à orienter l'offre de service vers les familles où le risque de commettre à nouveau un abus physique à l'égard d'un enfant est le plus élevé.

En contrepartie des résultats sur l'efficacité théorique des outils actuariels, l'outil CFRA, ou l'une de ses versions adaptées a été la cible de la plupart des critiques qui portent sur leur efficacité pratique (utilité, applicabilité), dans la littérature recensée. En réaction aux impacts de cet outil sur la prise de décision, certains de leurs utilisateurs soulèvent le problème de la surévaluation de la proportion des familles classées dans les catégories « à risque élevé » et « à risque très élevé » de récurrence, ou encore le fait que l'outil ne prend pas en compte les conditions de vie des familles évaluées, le contexte familial et son évolution. Outre la charge de travail supplémentaire et le temps consacré à mieux connaître et à adopter l'outil, la technicité de la pratique d'évaluation qui lui est associée par certains intervenants ainsi que la remise en question du jugement professionnel de ces derniers et de leurs compétences professionnelles qu'introduit son usage obligatoire sont décriées par ces intervenants. Un mauvais usage ou un détournement d'usage de l'outil représente alors la contrepartie à son usage obligatoire. Suivant son implantation, on assiste également à un usage de l'outil qui varie selon ses utilisateurs et les établissements, ce qui en limite la portée. Précisons que ces critiques s'appliquent plus particulièrement si l'outil actuariel est utilisé seul et d'une façon qui n'est pas conforme aux procédures et aux normes d'usage définies par ses concepteurs.

Ces préoccupations soulevées dans la littérature rejoignent parfois celles exprimées par les participants à l'un ou l'autre des 4 groupes de discussion focalisée. Cet exercice a effectivement permis d'aborder les conditions d'implantation et d'application d'un tel outil dans le contexte organisationnel et légal des SPJ du Québec. D'une part, ces participants ont soulevé les mêmes préoccupations que celles repérées dans la littérature, soit l'usage obligatoire de l'outil, la possibilité de faire abstraction du jugement professionnel lors de l'évaluation de cas, l'omission de tenir compte de l'évolution du contexte familial, la charge de travail supplémentaire pour mieux connaître et adopter l'outil, la nécessité d'obtenir une formation, la présence de personnes pivots pour s'assurer de l'application adéquate de l'outil, de même que l'assurance du soutien des gestionnaires. Les participants ont également soulevé d'autres préoccupations qui s'appliquent spécifiquement au contexte québécois, à savoir :

- les moments ou les étapes du processus d'intervention auxquels cet outil pourrait être utilisé en vertu de la LPJ;
- les conditions de son intégration possible parmi d'autres outils cliniques utilisés dans les CJ du Québec, mais aussi étant donné l'usage obligatoire du SSP dans ce type d'établissement;
- le suivi de l'implantation de l'outil selon des lignes directrices clairement établies dans les établissements concernés (CSSS et CJ);
- la signature d'une entente entre les CSSS et les CJ et leurs partenaires concernant les conditions d'usage de ce type d'outil.

Appelées à délibérer sur la possibilité d'adapter un outil d'évaluation du risque d'abus physique envers des enfants au contexte légal québécois, afin notamment de répondre à la question 4 du présent avis, les membres du comité de suivi ont pris connaissance des réponses aux trois questions précédentes. Ils ont également considéré les suggestions suivantes de l'INESSS : a) réaliser un projet pilote afin de déterminer la faisabilité d'intégrer un outil actuariel, comme l'outil CFRA - sa version adaptée (OFRA) en usage en Ontario pourrait servir d'outil de référence lors de l'expérimentation de ce type d'outil au Québec et, b) intégrer un outil de ce type à une ou plusieurs autres composantes d'un système d'évaluation intégré, tel que le SDM (Structure Decision Making), afin qu'il soit obligatoirement accompagné d'un outil contextuel, lequel prend en compte le contexte des familles évaluées. En conséquence, les membres du comité de suivi se sont penchés sur les suites à donner aux recommandations formulées ci-après par l'INESSS.

De leur côté, les experts indépendants consultés ont insisté sur l'importance de tenir compte de la spécificité de la structure et de l'organisation des SPE au Québec avant d'envisager l'adaptation d'un outil ayant fait ses preuves à l'étranger. Ces considérations s'avèrent d'autant plus pertinentes que selon Trocmé, Collin-Vézina et leurs collaborateurs [2014], le Québec a le système d'évaluation de signalements le plus efficace du Canada. Selon eux, un problème de responsabilité associée à la gestion du risque, dans d'autres provinces canadiennes (Ontario, Alberta, Manitoba, Saskatchewan), semble avoir eu des répercussions importantes sur le processus de prise de décision, à savoir un taux d'incidence des évaluations suivant un signalement et un taux d'incidence de fermeture de dossiers qui, dans ces provinces, se situent bien au-delà des taux observés au Québec.

3.3 Forces et limites de l'avis

La tenue de groupes de discussion focalisée auprès d'intervenants, d'agents de liaison et de gestionnaires des milieux d'intervention concernés par l'utilisation des outils évalués s'avère une de ses forces du présent avis. La représentativité des différentes fonctions occupées par ces participants en provenance de CSSS et de CJ a permis d'obtenir des points de vue variés sur l'efficacité pratique (utilité, applicabilité) des outils d'évaluation du risque d'abus physique. Ces points de vue ont fourni un ancrage précieux aux résultats de la revue de la littérature grise sur ce type d'outil en les situant sur le terrain des pratiques d'évaluation exercées par les établissements (CSSS, CJ) québécois en matière de protection de la jeunesse.

La délibération des membres du comité de suivi sur les résultats des revues de littérature et sur les groupes de discussion focalisée s'est également avérée fructueuse. Elle a permis de confirmer la pertinence de cet avis et d'y donner suite en envisageant la possibilité d'entreprendre un projet pilote concernant l'outil recommandé par l'INESSS, selon certaines conditions préalables.

Par ailleurs, les commentaires exprimés par les experts indépendants au cours de la démarche de réalisation du projet d'évaluation ont suscité des réflexions sur les choix d'ordre méthodologique. Ainsi, la revue systématique de la littérature scientifique a porté sur des outils d'évaluation du risque d'abus physique dont la validité est évaluée relativement à des indicateurs de récurrence. Cela a pu avoir pour effet d'exclure certaines études sur des outils qui évaluent la violence domestique exercée entre conjoints et entre ces derniers et leurs enfants. C'est le cas, par exemple, de l'outil Parent-Child Conflicts Tactics Scales (une version révisée, en 1998, du Conflict Tactics Scales, qui date de 1979) qui s'apparente, en raison de sa

capacité diagnostique, aux outils CAPI et AAPI-2. Utilisé dans plusieurs pays, cet outil a été traduit en français; il figure même dans un rapport de l'Institut de la statistique du Québec (ISQ) [2013].

À cela s'ajoute le fait que certaines études pertinentes n'ont pu être retenues parce qu'elles ont été réalisées avant 2003 et ne correspondaient donc pas aux critères de cet avis concernant la période à couvrir.

Une autre limite de cet avis est due à l'adaptation locale (selon les régions, les États, les territoires, etc.) d'outils de type actuariel. Par exemple, le fait qu'on utilise différemment le système d'évaluation intégré CRC-SDM d'un établissement à un autre, qu'il soit adapté aux caractéristiques de la population locale et que, parfois, seulement certaines de ses composantes soient utilisées, rend difficile l'évaluation de sa validité [Kim *et al.*, 2008]. Cette situation rend tout aussi ardue la tâche de comparer les outils de type actuariel entre eux. Ainsi, des versions adaptées des outils FRAAN et CFRA se distinguent de ceux-ci uniquement par leur adaptation au contexte local, qui consiste à ajouter ou à retirer un ou plusieurs des facteurs de risque qui composent les échelles de l'abus et de la négligence. Cela a fait en sorte de regrouper les versions adaptées de ces outils sous leur appellation d'origine (FRAAN, CFRA) plutôt qu'en indiquant le territoire administratif où ils sont utilisés. D'ailleurs, l'outil CFRA est l'une des composantes du système d'évaluation intégré CRC-SDM, lequel est largement utilisé aux États-Unis. Le nombre d'études de validation de ce type d'outil, et de ses adaptations locales, dépasse le nombre d'études de validation des autres outils recensés. Cela a sans doute influencé la prédominance des publications sur cet outil dans la revue de la littérature scientifique ayant servi à répondre à la question 1 de cet avis.

Enfin, la majorité des études recensées sur ce type d'outil ne porte pas uniquement sur le risque d'abus physique et sa récurrence, puisque la prédiction englobe l'abus en général et, parfois, la négligence. Le taux de prévalence de l'abus physique dans la population en général, mais également le taux d'occurrence et le taux de récurrence de l'abus physique qui figurent dans les bases de données des SPJ indiquent que ce problème occupe la deuxième place, assez loin derrière le problème de la négligence infantile, laquelle occupe le premier rang des types de maltraitance au Québec, comme c'est le cas à l'étranger. Le choix de s'en tenir à l'abus (tous types confondus) limite donc la portée des résultats, des conclusions et des recommandations présentés dans cet avis.

CONCLUSIONS ET RECOMMANDATIONS

La demande adressée à l'INESSS par l'ACJQ consistait à déterminer et à évaluer l'efficacité, sur le plan scientifique, d'outils d'évaluation du risque d'abus physique envers les enfants. Cette demande a donné lieu à la formulation des questions suivantes sur l'efficacité théorique et sur l'efficacité pratique de ce type d'outils :

- 1) Quelle est l'efficacité reconnue, sur le plan scientifique, des outils d'évaluation du risque d'abus physique envers à l'égard des enfants?
- 2) Les outils dont l'efficacité est reconnue sont-ils utiles à la prise de décision au moment d'évaluer des enfants qui font l'objet d'un signalement ou d'un nouveau signalement d'abus physique?
- 3) Comment ces outils dont l'efficacité est reconnue s'intègrent-ils au contexte d'intervention?
- 4) Les outils dont l'efficacité est reconnue sont-ils adaptables au contexte légal québécois?

Les données scientifiques provenant de 21 études de validation d'outils d'évaluation du risque d'abus physique et parfois, de la négligence parmi les familles à risque ayant déjà commis ces actes ont été analysées. Par la suite, les données provenant de 9 publications qui traitent de l'efficacité pratique de ce type d'outils ont permis de cerner un certain nombre de préoccupations qui ont été abordées lors de groupes de discussion focalisée auxquels ont participé des intervenants, des agents de liaison et des gestionnaires en provenance de CSSS ou de CJ du Québec. Enfin, une délibération des membres du comité de suivi en vue de produire le présent avis a permis d'évaluer la possibilité d'adapter un outil de ce type au contexte légal québécois.

En réponse à la question 1, qui porte sur l'efficacité théorique (fiabilité, validité) de ce type d'outil, l'INESSS conclut ce qui suit :

- à l'exception de l'outil CAPI (Child Abuse Potential Inventory), les outils évalués ne permettent pas de distinguer les différents types d'abus (physique, sexuel ou psychologique).
- aucun des outils évalués ne satisfait à la fois aux critères d'une bonne fiabilité et d'une bonne validité. La fiabilité des outils CAPI et FRAAN (Family Risk of Abuse and Neglect) est bonne. La validité de l'outil CFRA (California Family Risk Assessment) est jugée adéquate. Sa version adaptée, l'outil OFRA (Ontario Family Risk Assessment) est en usage en Ontario.
- aucun des outils recensés ne permet de prédire si les parents (ou leurs conjoints) vont commettre (à nouveau) un abus physique envers leur enfant ou non.
- les outils actuariels, qui tiennent compte de la présence de facteurs de risque associés à l'abus physique, sont plus efficaces (fiabilité, validité) que les outils contextuels pour évaluer le niveau de risque d'abus physique ou sa récurrence;
- lorsqu'il fait partie d'un système d'évaluation intégré, tel que le SDM (Structure Decision Making) conçu par le CRC (Children's Research Center) du National Council on Crime and

Delinquency (NCCD), l'outil actuariel fournit une meilleure évaluation du risque d'abus physique que lorsqu'il est utilisé seul. Étant donné la présence d'autres outils d'aide à la décision dans les établissements visés, il y aurait lieu d'examiner comment il peut s'insérer parmi ceux déjà en place.

En réponse aux questions 2 et 3, qui portent sur l'efficacité pratique (utilité, applicabilité) de ce type d'outil, l'INESSS conclut ce qui suit :

- l'ensemble des acteurs sur le terrain (CJ et CSSS) consultés a manifesté un intérêt pour l'ajout d'un outil d'évaluation actuariel du risque d'abus physique à la pratique d'évaluation actuelle.
- selon ces acteurs, un outil de ce type serait utile dans la mesure où il permettrait :
 - de standardiser le processus d'évaluation des cas;
 - de s'entendre sur un langage et des critères d'évaluation communs pour mieux saisir la problématique de l'abus physique;
 - d'harmoniser les pratiques d'évaluation des cas et le processus d'intervention qui en découle par le personnel concerné des CJ et des CSSS;
 - de soutenir la prise de décision sur la priorité et l'intensité des services à offrir à la clientèle visée.
- Selon ces acteurs, l'applicabilité de ce type d'outil dépendrait des conditions suivantes :
 - le moment ou l'étape du processus d'intervention où il est utilisé;
 - son caractère distinct par rapport aux autres types d'outils déjà en usage dans les SPJ au Québec (éviter le chevauchement et la répétition inutile des outils);
 - sa capacité à prendre en compte les caractéristiques du client (ex. : membre d'une communauté culturelle) et l'évolution de sa situation;
 - une formation commune du personnel concerné par cet outil;
 - la possibilité d'en faire un point d'ancrage des pratiques d'évaluation des cas pour le personnel des CSSS, des CJ, et leurs partenaires, afin de mieux intervenir auprès des familles visées;
 - la délimitation précise de la responsabilité du mandat et la désignation du mandataire en matière de gestion du risque et de protection de l'enfance.

Recommandations³³

Suivant la définition du NHMRC (National Health and Medical Research Council) des niveaux de recommandation (de A à D) qui est présentée dans l'annexe J, la recommandation 1 correspond au niveau C : l'ensemble des éléments de preuve considérés permet de recommander un outil actuariel d'évaluation du risque d'abus physique, l'outil CFRA (California Family Risk Assessment), bien que l'application de cet outil demeure sujette à certaines conditions préalables à son adaptation au système légal québécois en matière de protection de la jeunesse.

³³ Tel que mentionné dans l'Avant-propos, depuis le 1^{er} avril 2015, les centres jeunesse du Québec et les équipes dédiées au programme services « Jeunes en difficulté », en première ligne dans les CSSS, sont regroupés sous la direction « Programme jeunesse » à l'intérieur des CISSS et des CIUSSS.

Recommandation 1

L'INESSS recommande aux centres jeunesse du Québec d'intégrer un outil actuariel dans le processus d'évaluation du risque d'abus physique que présentent les familles faisant l'objet d'un signalement.

- Cet outil peut servir, en plus du jugement de l'intervenant, à soutenir l'évaluation du niveau de risque d'abus physique que présentent les familles en fonction de la présence ou de l'absence de certains facteurs de risque qui ont été associés à l'abus physique dans la littérature scientifique. Cet outil ne permet pas de prédire l'occurrence ou la récurrence de ce phénomène pour une famille en particulier;
- L'évaluation du risque de maltraitance que présentent les familles ne devrait pas reposer uniquement sur un outil actuariel. Cette évaluation doit s'appuyer sur un ensemble d'outils d'évaluation, tel que le SDM (Structured Decision Making), élaboré par le CRC (Children's Research Center) du National Council on Crime and Delinquency (NCCD), afin de tenir compte de la situation, des besoins et du contexte des familles évaluées.

Recommandation 2

Avant de procéder à l'intégration proprement dite d'un outil actuariel dans le processus d'évaluation du risque, l'INESSS recommande aux centres jeunesse du Québec de considérer et de documenter les éléments suivants :

- déterminer si les facteurs de risque, notamment aux États-Unis, et dont tiennent compte les outils actuariels évalués, sont les mêmes que ceux associés à la maltraitance au Québec;
- situer à quel moment un outil de ce type doit être utilisé dans l'ensemble du processus d'évaluation des familles dans les centres jeunesse du Québec, et selon quels critères (clientèle, étape et objectifs). À l'origine, ce type d'outil a été conçu pour être utilisé à l'étape « Évaluation de la situation »;
- évaluer la pertinence et la faisabilité d'adapter au contexte québécois la version adaptée du CFRA, soit l'outil OFRA, en usage en Ontario et d'en faire la traduction en français, conformément aux procédures recommandées concernant la traduction des outils psychométriques;
- évaluer le degré de correspondance entre le SSP (Système de soutien à la pratique), conçu en fonction des lois québécoises et dont l'usage est obligatoire au Québec, et le système d'évaluation intégré SDM, afin de voir si les différentes composantes (modules) du SDM pourraient compléter celles du SSP.

Recommandation 3

Dans un contexte où les résultats de cette évaluation mèneraient à recommander le déploiement à grande échelle de l'outil suggéré, l'INESSS recommande aux centres jeunesse de s'assurer que les conditions de succès suivantes sont en place :

- pouvoir compter sur un engagement organisationnel (soutien du personnel, responsabilité des instances des établissements concernés par la gestion du risque et la protection de l'enfance, etc.) afin d'appuyer les changements apportés aux pratiques d'intervention;

- définir clairement les procédures et les usages de l'outil au moyen de lignes directrices.
- offrir une formation à l'ensemble du personnel concerné de manière à favoriser une meilleure compréhension de la terminologie, de la finalité, des critères d'usage et des résultats obtenus à l'aide de ce type d'outil;
- s'assurer que l'application de l'outil et les résultats de l'évaluation obtenus à l'aide de celui-ci sont adaptés aux politiques des CSSS et des centres jeunesse, et aux ressources disponibles dans ceux-ci, afin de pouvoir répondre à la demande de services;
- se doter de moyens pour évaluer et suivre l'implantation de l'outil (utilisation, impacts);
- assurer une veille des études sur ce type d'outil.

RÉFÉRENCES

* indique les références bibliographiques incluses dans la revue systématique

Agar SE. The development of general risk assessment guidelines for physical child abuse and neglect [thèse]. Burnaby, BC : Simon Fraser University; 2003.

Agence d'évaluation des technologies et des modes d'intervention en santé (AETMIS). Guide pratique en évaluation des technologies et des modes d'intervention en santé à l'intention des chercheurs de l'AETMIS. Chapitre 8 : Évaluation des tests diagnostiques. Guide préparé par Pierre Dagenais et Alicia Framarin. Montréal, Qc : AETMIS; 2008.

Altman DG et Royston P. What do we mean by validating a prognostic model? Stat Med 2000;19(4):453-73.

Association des centres jeunesse du Québec (ACJQ). Avec l'énergie du premier jour. Bilan des directeurs de la protection de la jeunesse / directeurs provinciaux 2014. Québec, Qc : ACJQ; 2014. Disponible à : <http://www.centrejeunessedequébec.qc.ca/publications/Bilan%20DPJ/Bilan%20DPJ%202013-2014.pdf>.

Baird C. Comparative study of the use and effectiveness of different risk assessment models in CPS decision making processes. NDACAN Dataset Number 84. Madison, WI : National Council on Crime and Delinquency; 2002. Disponible à : http://www.ndacan.cornell.edu/datasets/pdfs_user_guides/084user.pdf.

*Barber JG, Shlonsky A, Black T, Goodman D, Trocmé N. Reliability and predictive validity of a consensus-based risk assessment tool. Journal of Public Child Welfare 2008;2(2):173-95.

*Barlow J, Fisher JD, Jones D. Systematic review of models of analysing significant harm. Research Report DFE-RR199. Londres, Angleterre : Department for Education; 2012. Disponible à : https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/183949/DFE-RR199.pdf.

Baumann DJ, Dagleish L, Fluke J, Kern H. The decision-making ecology. Washington, DC : American Humane Association; 2011. Disponible à : <http://www.americanhumane.org/assets/pdfs/children/cprc-dme-monograph.pdf>.

Bolton A et Lennings C. Clinical opinions of structured risk assessments for forensic child protection: The development of a clinically relevant device. Children and Youth Services Review 2010;32(10):1300-10.

*Chaffin M et Valle LA. Dynamic prediction characteristics of the Child Abuse Potential Inventory. Child Abuse Negl 2003;27(5):463-81.

*Connors NA, Whiteside-Mansell L, Deere D, Ledet T, Edwards MC. Measuring the potential for child maltreatment: The reliability and validity of the Adult Adolescent Parenting Inventory—2. Child Abuse Negl 2006;30(1):39-53.

Contandriopoulos A-P, Champagne F, Potvin L, Denis J-L, Boyle P. Savoir préparer une recherche : la définir, la structurer, la financer. Montréal, Qc : Presses de l'Université de Montréal; 1990.

- *Coohey C, Johnson K, Renner LM, Easton SD. Actuarial risk assessment in child protective services: Construction methodology and performance criteria. *Children and Youth Services Review* 2013;35(1):151-61.
- *D'Andrade A, Austin MJ, Benton A. Risk and safety assessment in child welfare: Instrument comparisons. *J Evid Based Soc Work* 2008;5(1-2):31-56.
- *Dankert EW et Johnson K. Risk assessment validation: A prospective study. California Department of Social Services, Children and Family Services Division. Madison, WI : National Council on Crime and Delinquency (NCCD) Children's Research Center; 2014. Disponible à : http://www.nccdglobal.org/sites/default/files/publication_pdf/risk-assessment-validation.pdf.
- Desrosiers J et Larivière N. Le groupe de discussion focalisée : application pour recueillir des informations sur le fonctionnement au quotidien des personnes avec un trouble de la personnalité limite. Dans : Corbière M et Larivière N, réd. *Méthodes qualitatives, quantitatives et mixtes : dans la recherche en sciences humaines, sociales et de la santé* Québec, Qc : Presses de l'Université du Québec; 2014 : 257-304.
- European Centre for Disease Prevention and Control. Evidence-based methodologies for public health – How to assess the best available evidence when time is limited and there is lack of sound evidence. Stockholm, Suède : ECDC; 2011. Disponible à : http://ecdc.europa.eu/en/publications/Publications/1109_TER_evidence_based_methods_for_public_health.pdf.
- *Ezzo F et Young K. Child maltreatment risk inventory: Pilot data for the Cleveland Child Abuse Potential Scale. *J Fam Viol* 2012;27(2):145-55.
- Fitch D. Examination of the Child Protective Services decision-making context with implications for decision support system design. *Journal of Social Service Research* 2006;32(4):117-34.
- Fortin F. Propriétés métrologiques des instruments de mesure (fidélité-validité). *Recherche en Soins Infirmiers* 1994;39:58-62.
- Frenette M, Abriat P, Dufault D. L'intelligence artificielle et l'évaluation sociale en protection de l'enfance. Dans : Gabel M et Durning P, réd. *Évaluation(s) des maltraitances : rigueur et prudence*. Paris, France : Fleurus; 2002 : 311-66.
- Fuller TL et Nieto M. Illinois Child Endangerment Risk Assessment Protocol: FY10 Annual Evaluation. Urbana, IL : Children and Family Research Center, University of Illinois at Urbana-Champaign; 2010. Disponible à : http://cfrc.illinois.edu/pubs/rp_20100401_IllinoisCERAPAnnualEvaluationFY2010.pdf.
- *Garrison M. Taking the risks out of child protection risk analysis. *Journal of Law and Policy* 2012;21(1):5-35.
- *Gillingham P et Humphreys C. Child protection practitioners and decision-making tools: Observations and reflections from the front line. *Br J Soc Work* 2010;40(8):2598-616.
- Guay J-P. Prédiction actuarielle et prédiction clinique : le dernier souffle d'une pratique traditionnelle. *Revue Internationale de Criminologie et de Police Technique et Scientifique* 2006;59:149-64.

- Hélie S et Laurier C. Récurrence de la violence, subie ou agie, après une première intervention en protection de la jeunesse : ampleur du risque et facteurs associés. Québec, Qc : Association des centres jeunesse du Québec; 2012. Disponible à : <http://www.centrejeunessedequébec.qc.ca/recherche/RC/Publications%20de%20la%20recherche/Récurrence%20de%20la%20violence%20après%20une%20première%20intervention%20en%20protection%20de%20la%20jeunesse.pdf>.
- Institut de la statistique du Québec (ISQ). La violence familiale dans la vie des enfants du Québec, 2012. Les attitudes parentales et les pratiques familiales. Québec, Qc : ISQ; 2013. Disponible à : <http://www.stat.gouv.qc.ca/statistiques/sante/environnement-social/violence-familles/violence-familiale-2012.pdf>.
- Institut national d'excellence en santé et en services sociaux (INESSS). Cadre d'élaboration des guides de pratique dans le secteur des services sociaux. Document rédigé par Sylvie Beauchamp, Martin Drapeau, Carmen Dionne et Jean-Pierre Duplantie. Québec, Qc : INESSS; 2015. Disponible à : https://www.inesss.qc.ca/fileadmin/doc/INESSS/Rapports/ServicesSociaux/INESSS_Cadre_elaboration_guides_pratique_servicessociaux.pdf.
- Institut national d'excellence en santé et en services sociaux (INESSS). Efficacité des interventions en matière de négligence auprès des enfants, des familles et des communautés autochtones. Rapport rédigé par Carrie Anna McGinn et Jean Damasse. ETMIS 2014;10(2):1-38.
- Institut national d'excellence en santé et en services sociaux (INESSS). Les normes de production des revues systématiques : guide méthodologique. Document rédigé par Valérie Martin et Jolianne Renaud sous la direction de Pierre Dagenais. Québec, Qc : INESSS; 2013. Disponible à : https://www.inesss.qc.ca/fileadmin/doc/INESSS/DocuMetho/INESSS_Normes_production_revues_systematiques.pdf.
- *Johnson K et Bogie A. Risk assessment validation: A prospective study. North Carolina Department of Health and Human Services, Division of Social Services. Madison, WI : National Council on Crime and Delinquency (NCCD) Children's Research Center; 2009. Disponible à : http://www.nccdglobal.org/sites/default/files/publication_pdf/nc_risk_asst_validation_final_2009_2.pdf.
- *Johnson K et O'Connor D. Post-implementation examination of a risk assessment's ability to classify families by the likelihood of subsequent child protective services involvement: A report for Washington Children's Administration. Madison, WI : National Council on Crime and Delinquency (NCCD) Children's Research Center; 2008. Disponible à : <https://www.dshs.wa.gov/sites/default/files/CA/pub/documents/SDMpostimplementation-Report.pdf>.
- *Johnson K et Wagner D. Risk assessment revalidation: A prospective study. California Structured Decision Making. Madison, WI : National Council on Crime and Delinquency (NCCD) Children's Research Center; 2003. Disponible à : http://www.nccdglobal.org/sites/default/files/publication_pdf/cacps2003riskvalidationreport.pdf.
- Johnson K, O'Connor D, Berry S, Ramelmeier D, Pecora PJ. Structuring the decision to accept a child protection report. Journal of Public Child Welfare 2012;6(2):191-205.

- *Johnson K, Wagner D, Scharenbroch C. Risk assessment validation: A prospective study. California Department of Social Services, Children and Family Services Division. Madison, WI : National Council on Crime and Delinquency (NCCD) Children's Research Center; 2007. Disponible à : http://www.nccdglobal.org/sites/default/files/publication_pdf/ca2007riskassessmentvalidation_rpt.pdf.
- *Johnson K, Wagner D, Scharenbroch C, Healy T. Risk assessment validation: A prospective study. Minnesota Department of Human Services. Madison, WI : National Council on Crime and Delinquency (NCCD) Children's Research Center; 2006. Disponible à : http://www.nccdglobal.org/sites/default/files/publication_pdf/mn_2006_ra_validation.pdf.
- *Johnson W. Effectiveness of California's child welfare. Structured Decision-Making (SDM) model: A prospective study of the validity of the California Family Risk Assessment. Oakland, CA : Alameda County Social Services Agency; 2004. Disponible à : http://www.nccdglobal.org/sites/default/files/publication_pdf/ca_sdm_model_feb04.pdf.
- *Johnson WL. The validity and utility of the California Family Risk Assessment under practice conditions in the field: A prospective study. *Child Abuse Negl* 2011;35(1):18-28.
- *Jones B et Beecroft E. The impacts of actuarial risk assessment on child protective services in Virginia. Richmond, VA : Office of Research, Virginia Department of Social Services; 2008. Disponible à : https://www.dss.virginia.gov/files/about/reports/children/cps/all_other/2008/SDM_impacts_2008-02-19.pdf.
- *Kim AK, Brooks D, Kim H, Nissly J. Structured Decision Making[®] and child welfare service delivery project. Berkeley, CA : University of California at Berkeley, California Social Work Education Center; 2008. Disponible à : <https://web.csulb.edu/projects/ccwrl/Brooks.pdf>.
- Kmet LM, Lee RC, Cook LS. Standard quality assessment criteria for evaluating primary research papers from a variety of fields. HTA Initiative # 13. Edmonton, AB : Alberta Heritage Foundation for Medical Research (AHFMR); 2004. Disponible à : <http://www.biomedcentral.com/content/supplementary/1471-2393-14-52-S2.pdf>.
- Knoke D et Trocmé N. L'évaluation du risque en protection de l'enfance. Feuille d'information CEPB #18F. Toronto, ON : Centre d'excellence pour la protection et le bien-être de l'enfance (CEPB); 2004. Disponible à : <http://cwrp.ca/sites/default/files/publications/fr/RiskAssessment18F.pdf>.
- *Lee B, Black T, Shlonsky A, Fallon B. Validation of the Ontario Family Risk Assessment instrument for child maltreatment recurrence: Physical abuse. Toronto, ON : Ontario Child Abuse and Neglect Data System (OCANDS); 2014. Disponible à : <https://www.ocands.org/sites/default/files/OCANDS%20Information%20Sheet%203.pdf>.
- *Leschied AW, Chiodo D, Whitehead PC, Hurley D, Marshall L. The empirical basis of risk assessment in child welfare: The accuracy of risk assessment and clinical judgment. *Child Welfare* 2003;82(5):527-40.
- Letts L, Wilkins S, Law M, Stewart D, J. B, Westmorland M. Critical review form – Qualitative studies (Version 2.0). Hamilton, ON : McMaster University; 2007. Disponible à : <https://www.canchild.ca/en/canchildresources/resources/qualform.pdf>.

- *Loman LA et Siegel GL. An evaluation of the Minnesota SDM Family Risk Assessment. St. Louis, MO : Institute of Applied Research; 2004. Disponible à : <http://www.iarstl.org/papers/FinalFRARReport.pdf>.
- *McVeigh Palmer M. Actuarial risk assessment and Ontario Child Welfare Transformation: A paradox of purpose. Hamilton, ON : McMaster University; 2011. Disponible à : <https://macsphere.mcmaster.ca/bitstream/11375/11383/1/fulltext.pdf>.
- Ministère de la Santé et des Services sociaux (MSSS). Manuel de référence sur la protection de la jeunesse. Québec, Qc : MSSS; 2010. Disponible à : <http://publications.msss.gouv.qc.ca/acrobat/f/documentation/2010/10-838-04.pdf>.
- Ministère de la Santé et des Services sociaux (MSSS). Les services intégrés en périnatalité et pour la petite enfance à l'intention des familles vivant en contexte de vulnérabilité – Cadre de référence. Québec, Qc : MSSS; 2004. Disponible à : <http://publications.msss.gouv.qc.ca/acrobat/f/documentation/2004/04-836-02W.pdf>.
- Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: Application and impact of prognostic models in clinical practice. *BMJ* 2009;338:b606.
- Munro E. What tools do we need to improve identification of child abuse? *Child Abuse Review* 2005;14(6):374-88.
- National Health and Medical Research Council (NHMRC). NHMRC additional levels of evidence and grades for recommendations for developers of guidelines. Canberra, Australie : NHMRC; 2009. Disponible à : https://www.nhmrc.gov.au/_files_nhmrc/file/guidelines/developers/nhmrc_levels_grades_evidence_120423.pdf.
- Noonan VK, Miller WC, Eng JJ, Chan CW. Chapter 28 : Outcome measures. Version 4.0. Vancouver, BC : Spinal Cord Injury Research Evidence (SCIRE); 2013. Disponible à : <http://www.scireproject.com/sites/default/files/SCIRE4-OM-Full-Ch.pdf>.
- Pecora PJ, Chahine Z, Graham JC. Safety and risk assessment frameworks: Overview and implications for child maltreatment fatalities. *Child Welfare* 2013;92(2):143-60.
- Price-Robertson R et Bromfield L. Risk assessment in child protection. NCPC resource sheet. Melbourne, Australie : Australian Institute of Family Studies; 2011. Disponible à : <https://aifs.gov.au/cfca/sites/default/files/publication-documents/rs24.pdf>.
- *Regehr C, Bogo M, Shlonsky A, LeBlanc V. Confidence and professional judgment in assessing children's risk of abuse. *Research on Social Work Practice* 2010;20(6):621-8.
- *Rycus JS et Hughes RC. Issues in risk assessment in child protective services: Policy white paper. Columbus, OH : North American Resource Center Child Welfare; 2003. Disponible à : <http://www.ihs-trainet.com/assets/RAPdf.pdf>.
- Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol* 2009;62(10):1013-20.
- Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10.

- *Shlonsky A et Lambert L. The perceived utility of child maltreatment risk assessment and clinical assessment tools. Toronto, ON : Bell Canada Child Welfare Research Unit; 2007. Disponible à : http://cwrp.ca/sites/default/files/publications/en/Perceived_Utility-RiskAssessment.pdf.
- *Shlonsky A et Wagner D. The next step: Integrating actuarial risk assessment and clinical judgment into an evidence-based practice framework in CPS case management. *Children and Youth Services Review* 2005;27(4):409-27.
- *Shlonsky A, Shin TM, Wong B, Sawh P, Lee B, Eisner A, Tan J. Prospective validation of the new Ontario decision support system: Phase IV final report. Toronto, ON : University of Toronto; 2011.
- Trocmé N, Collin-Vézina D, Roy C, Chamberland C, Desbiens N, Dufour S, et al. Projet de loi 10 et services de protection de la jeunesse : des impacts à considérer. Mémoire déposé dans le cadre des consultations publiques sur le projet de loi 10. Montréal, Qc : Université McGill; 2014. Disponible à : https://www.mcgill.ca/socialwork/files/socialwork/2014_impactsloi10.pdf.
- Trocmé N, Fallon B, MacLaurin B, Sinha V, Black T, Fast E, et al. Étude canadienne sur l'incidence des signalements de cas de violence et de négligence envers les enfants – 2008 : Données principales. Ottawa, ON : Agence de la santé publique du Canada (ASPC); 2010.
- Van Driel M. Glossaire des termes utilisés en Evidence-Based Medicine. 1^{ère} édition en français. Gand, Belgique : Minerva; 2004. Disponible à : <http://catalogue.cssslaval.qc.ca/GEIDFile/43199.PDF?Archive=191380291956&File=43199.PDF>.
- *Walker CA et Davies J. A critical review of the psychometric evidence base of the Child Abuse Potential inventory. *J Fam Viol* 2010;25(2):215-27.
- *Wells BM, Crouch JL, Schubert R, Irwin LM, Risser HJ, Skowronski JJ. Revisiting the issue of the Child Abuse Potential Inventory's internal consistency in adolescent samples. *J Adolesc Health* 2011;48(4):351-7.
- White A et Walsh P. Risk assessment in child welfare: An issues paper. Ashfield, Australie : Centre for Parenting and Research; 2006. Disponible à : http://www.community.nsw.gov.au/docswr/_assets/main/documents/research_riskassessment.pdf.
- *Wood H. Assessing the feasibility of using actuarial risk assessment tool to identify risks in child protection cases [thèse]. Bournemouth, Royaume-Uni : Bournemouth University; 2011. Disponible à : http://eprints.bournemouth.ac.uk/19387/1/Wood,Heather_Ph.D._Nov.2011.pdf.