

POUR UN USAGE PLUS PRATIQUE DES
RÉSULTATS DE L'ÉVALUATION DE PROGRAMMES D'INTERVENTION:
L'AMPLEUR DE L'EFFET ET L'INTERVALLE DE CONFIANCE

par

Marc Tourigny

Université du Québec à Hull, Département de psychoéducation
C.P. 1250, Succursale B, Hull, CANADA, J8X 3X7
Tel (819) 595-3900 poste 2333; Fax: (819) 595-2384;
E-mail: Marc_Tourigny@uqah.quebec.ca

&

Normand Péladeau

Provalis Research
5000 Adam Street, Montréal, CANADA, H1V 1W5

Communication présentée dans le cadre de la conférence annuelle
de la Société canadienne d'évaluation: "Évaluation en partenariat"

Ottawa

Mai 1997

RÉSUMÉ

L'utilisation des tests d'hypothèses et du calcul des probabilités dans les recherches évaluatives est de plus en plus remise en cause. Plusieurs auteurs proposent comme solution de remplacement, le calcul d'une mesure d'ampleur d'effet, accompagné d'un intervalle de confiance. Cette stratégie permet une meilleure compréhension des effets du programme, fournissant ainsi aux gestionnaires des programmes, des résultats plus appropriés et plus utiles lors de la prise de décisions. Une brève présentation des résultats de l'efficacité d'un programme de traitement pour enfants abusés sexuellement permettra d'illustrer la stratégie proposée.

L'utilisation de tests d'hypothèses et du calcul des probabilités dans le cadre de recherches et tout particulièrement dans le cadre de recherches évaluatives est remise en cause par plusieurs chercheurs en sciences humaines, sociales, médicales et en évaluation de programme (Lipsey, 1988; Morrison et Henker, 1970; Oakes, 1986). En réponse à ce problème, Cohen (1990), de même que Rosnow et Rosenthal (1989) suggèrent le calcul d'une mesure d'ampleur d'effet accompagnée d'un intervalle de confiance. Cette présentation a comme objectif de montrer l'utilité pratique d'utiliser l'ampleur de l'effet et l'intervalle de confiance dans le contexte des évaluations de programme. Dans un premier temps, je situerai brièvement chacun des concepts et les avantages liés à cette stratégie. Par la suite, à l'aide d'un exemple pratique portant sur l'évaluation d'un programme de traitement pour enfants agressés sexuellement, je vais illustrer les avantages liés à l'utilisation de l'ampleur d'effet et de l'intervalle de confiance.

Test d'hypothèses versus l'ampleur d'effet et l'intervalle de confiance

Les études évaluatives mettent souvent l'accent sur la direction ou l'existence d'un effet du programme plutôt que sur l'ampleur de l'effet. Autrement dit, le test d'hypothèse statistique nous permet de dire si ou non nous pouvons rejeter l'hypothèse nulle mais il ne nous donne aucun renseignement sur l'ampleur de l'effet. Ces études vont mentionner par exemple, que les résultats des tests statistiques montrent une amélioration statistiquement significative de l'estime de soi chez le groupe recevant le programme de traitement. Elles présentent donc l'existence d'un effet et sa direction. On se pose alors la question: Est-ce que le programme de traitement améliore la santé psychologique des victimes? Par contre, lorsque l'on présente l'ampleur de l'effet c.-à-d. la mesure de l'association entre deux variables, on répond davantage à la question: De combien le traitement améliore-t-il la santé psychologique des victimes? ou bien, Quel est l'effet du traitement sur la santé psychologique des victimes? Enfin, il faut mentionner qu'un autre désavantage du test d'hypothèses est que sa capacité à fournir de l'information c.-à-d. à trouver une différence significative est très liée au nombre de sujets ce qui n'est pas le cas pour l'ampleur de l'effet.

L'utilisation du test d'hypothèses appauvrit donc la prise de décision. Par exemple, si les amplitudes d'effets ne sont pas rapportés, il devient difficile de comparer les bénéfices de différents programmes, de juger si les bénéfices d'un programme en valent la peine compte tenu des coûts et de comprendre comment les effets d'un programme peuvent varier en fonction de ses caractéristiques et de la nature de sa clientèle. Il est également avantageux d'utiliser l'ampleur de l'effet, même lorsque l'estimé de l'effet n'est pas statistiquement significatif. Il est possible que la puissance statistique soit faible, et que l'estimé de l'ampleur de l'effet puisse nous aider à juger si un programme vaut la peine d'être poursuivi en dépit de résultats non significatifs.

L'ampleur de l'effet (effect size) ne représente toutefois qu'un estimé de l'effet réel d'une variable sur une autre. Pour obtenir une plus grande précision concernant l'effet réel, plusieurs auteurs

proposent l'utilisation de l'ampleur de l'effet et d'un intervalle de confiance. Un intervalle de confiance présente l'étendue à l'intérieur duquel on peut s'attendre à retrouver l'effet réel du programme, et ce, pour un seuil précis de probabilité. Par exemple, si nous établissons le niveau de confiance des intervalles à 95%, nous pouvons être confiant à 95% que l'effet réel du programme se situe à l'intérieur de cet intervalle de confiance. De plus, si cet intervalle ne contient pas la valeur zéro c'est-à-dire la corrélation $r = 0$ qui représente l'hypothèse nulle (à savoir qu'il n'y a pas d'effet du programme) alors il est possible d'affirmer que la probabilité que l'effet dû à l'intervention soit nul est inférieur à 0.05 (1 - 0.95).

L'avantage d'un tel intervalle de confiance est qu'il nous donne l'étendue des estimés possibles de l'effet réel d'un programme alors que le test d'hypothèses nous présente uniquement si l'estimé de l'effet réel est différent de zéro ou non.

Un autre avantage de présenter les résultats sous forme d'intervalle de confiance est de permettre de documenter le degré d'incertitude des résultats beaucoup mieux que ne le ferait une présentation de tests d'hypothèses. La largeur de l'intervalle de confiance constitue un indicateur direct de la puissance statistique de la recherche ou plus justement de la précision statistique pouvant être atteinte (Reichardt et Gollob, 1987). Plus l'intervalle de confiance est grand et plus la précision statistique est faible. De même, plus le nombre de sujets augmente et plus l'intervalle de confiance diminue.

DÉMONSTRATION DE L'UTILISATION DE L'AMPLEUR DE L'EFFET ET DE L'INTERVALLE DE CONFIANCE

Les résultats présentés dans le cadre de cette démonstration portent sur l'évaluation d'un programme de traitement d'enfants abusés sexuellement (PTEAS) mis sur pied en 1982 au Québec (Lebeau et Benoit, 1984). Dans un premier temps, une brève description du programme de traitement est faite, suivie d'une présentation de la méthodologie utilisée pour l'évaluation du programme. Enfin, les résultats de l'évaluation sont présentés en termes d'ampleur de l'effet et d'intervalle de confiance.

1. DESCRIPTION DU PROGRAMME DE TRAITEMENT

Le programme se présente comme une adaptation de modèles développés aux États-Unis, dont ceux de Giarretto (1981) et de Sgroi (1986). Les buts premiers du programme sont: 1^o de prévenir la récurrence c.-à-d. de protéger la victime en mettant fin aux comportements d'abus sexuel; 2^o de réparer les dommages psychologiques subits par les différents acteurs (victime, agresseur, conjointe).

Le P.T.E.A.S. se compose de cinq modules constitués en éléments ou sous-éléments. Le Tableau I présente la structure du programme. Le premier module concerne l'évaluation du signalement pour abus sexuel, il comprend des rencontres avec les différents acteurs impliqués soit l'agresseur, sa conjointe et la victime. C'est à ce niveau que les éléments placement de l'enfant et/ou

tribunal de la jeunesse peuvent être introduits. La durée prévue de cette première étape est d'environ une semaine. Le deuxième module a trait à la thérapie individuelle offerte aux différents acteurs impliqués (victime, agresseur, mère et/ou conjointe et au besoin à la fratrie); sa durée peut varier de 3 à 6 mois à raison d'une heure/semaine/personne (12 à 24 rencontres). Le troisième module est celui de la thérapie conjointe (dans l'ordre mère/fille, père/mère et père/fille). Ce module dure normalement de 3 à 6 mois à raison d'une heure/semaine/dyade (12 et 24 rencontres). Le quatrième module concerne la thérapie familiale et sa durée minimum est établie à 3 mois (12 rencontres). Enfin, le module entraide est constitué d'éléments de thérapie de groupe pour les victimes, les agresseurs, les conjointes et les couples. Pour l'ensemble du traitement thérapeutique, le temps moyen prévu pour les victimes est de 15 à 27 mois (soit de 60 à 96 heures) et de 18 à 27 mois (soit de 72 à 108 heures) pour les adultes. Pour les enfants de moins de 12 ans la durée est toutefois moins importante, la thérapie de groupe étant plus courte.

1.1 L'ÉVALUATION DU P.T.E.A.S.

Le but de la recherche était d'évaluer l'efficacité du P.T.E.A.S. auprès des victimes principalement et également auprès des agresseurs et de leurs conjointes. Dans ce cadre, le principal objectif concernant les victimes, consistait à évaluer les effets à court terme (16 mois) de l'intervention sur les principaux aspects psychologiques que le programme vise à modifier chez la victime, et ce, en distinguant deux volets de l'intervention.

2. MÉTHODOLOGIE

2.1 Devis, variable indépendante et stratégies d'analyse

Devis et variable indépendante

Il s'agit d'un devis de recherche comprenant des mesures pré-traitement et post-traitement (i.e. 16 mois plus tard). Contrairement à une pratique courante en évaluation de programme, la variable indépendante (i.e. l'intervention) a été traitée sous forme de variable continue plutôt que dichotomique (par exemple avec l'inclusion d'un groupe contrôle). Cette méthode a l'avantage de pouvoir considérer toute la variabilité retrouvée généralement au niveau de l'application d'un traitement. La méthode améliore également la validité de la mesure traitement et augmente considérablement la puissance statistique ou le niveau de précision du devis de recherche (Cook et Poole, 1982). Le principal inconvénient de ce type de devis se situe au niveau du biais de sélection, dans la mesure où la décision de participer ou non au programme d'intervention n'est généralement pas le fruit du hasard mais peut dépendre de différentes variables telles la sévérité du problème, la personnalité du participant, etc. Un effort particulier a été déployé afin d'identifier la présence d'un tel biais de sélection et de le contrôler.

Dans l'analyse de l'effet du programme de traitement, nous avons également crû bon de distinguer les deux grands volets d'intervention du programme de traitement, soit 1^o le nombre de rencontres individuelles, dyadiques et familiales auxquelles la victime a participé (volet I.D.F.); 2^o le nombre de rencontres de groupe offertes dans le cadre de l'Association Jeunes Unis (volet groupe). Les analyses effectuées ont permis d'établir l'effet du niveau de participation de chacun des deux volets en contrôlant pour le niveau de participation des sujets à l'autre volet d'intervention. Par exemple, l'effet mesuré du volet I.D.F. tient compte du niveau de participation des sujets au volet groupe et par conséquent, ne peut être attribuable à la participation des sujets à ce deuxième volet.

Stratégies d'analyses

La principale stratégie d'analyse statistique a consisté à utiliser des analyses de régression multiple par bloc hiérarchique (Cohen & Cohen, 1977). Cette méthode permet de spécifier l'ordre dans lequel les variables sont entrées dans l'équation de régression. Cet ordre d'entrée étant déterminée selon des considérations logiques telles l'importance théorique des variables, la préséance causale ou encore la nécessité de contrôler certaines variables. Le modèle typique d'analyse hiérarchique a consisté à entrer dans une première étape le score de la variable dépendante obtenu au Temps 1, soit avant le début de l'intervention. Par la suite, les deux variables représentant le niveau de participation à chacun des volets d'intervention sont entrées simultanément. Le coefficient de corrélation partielle de Pearson associé à chacune de ces variables est utilisé comme la mesure de l'effet spécifique du volet sur la variable dépendante. La première étape qui consiste à entrer la mesure de la variable dépendante obtenu avant le début de l'intervention est mathématiquement équivalente à une stratégie d'analyse de covariance avec la mesure au pré-test comme co-variable. Elle a pour avantage d'augmenter la puissance statistique du devis et la précision des estimations d'ampleur d'effet en réduisant la variation inter-sujets (Lipsey, 1990). Cette stratégie permet également une interprétation des résultats en terme de changement.

La mesure d'effet choisie pour les besoins de la présente étude est le coefficient de corrélation partielle de Pearson qui exprime l'ampleur de la relation existant entre le niveau d'implantation et les changements observées entre le pré-test et le post-test sur les variables dépendantes. Pour faciliter l'interprétation des résultats, ces coefficients de corrélation peuvent être classifiés en trois catégories: les corrélations dont la valeur absolue se situe entre 0.1 et 0.3 sont considérées comme représentant un effet léger; celles dont la valeur se situe entre 0.3 et 0.5 expriment un effet moyen, tandis que celles dont la valeur est supérieure à 0.5 seront considérées comme fortes. Les autres corrélations dont la valeur se situe entre -0.1 et 0.1 seront considérées comme l'indice d'un effet nul ou négligeable. Pour donner une indication de l'ampleur d'effet auquel on est en droit de s'attendre, notons que l'effet généralement observé des programmes d'intervention en psychologie se situe entre 0.25 et 0.30. Contrairement à l'usage courant, nous n'avons pas utilisé le test d'hypothèse pour déterminer la

présence ou non d'un effet. Nous avons plutôt identifié un intervalle de confiance pour chaque variable de corrélation partielle.

Intervalle de confiance

Compte tenu de la faible taille de notre échantillon, nous avons établi le niveau de confiance des intervalles à 80% ce qui veut dire qu'il y a 80% de vraisemblance que l'effet réel du programme se situe à l'intérieur de cet intervalle. Par exemple, si cet intervalle ne contient pas la valeur zéro c'est-à-dire la corrélation $r = 0$ qui représente l'hypothèse nulle (à savoir qu'il n'y a pas d'effet du programme) alors il est possible d'affirmer que la probabilité que l'effet dû à l'intervention soit nul est inférieur à 0.20 (1 - 0.80). Salsburg (1985) considère qu'un tel intervalle de 80% fournit des informations que l'on peut considérer comme raisonnablement sûres et par conséquent, amplement suffisantes pour guider la prise de décision. Malgré la faible taille de l'échantillon et le grand nombre de mesures dépendantes, nous avons décidé de ne pas procéder à une diminution du nombre de variables que ce soit par une réduction logique ou statistique. Par conséquent, on doit s'attendre à ce qu'un intervalle de confiance sur cinq indique faussement un effet réel du traitement. Il nous semble donc périlleux de porter un jugement précis s'appuyant sur un seul intervalle de confiance. La discussion et l'interprétation des résultats devraient plutôt reposer sur l'identification d'un patron clair de changement sur un ensemble d'indices. Ainsi, on pourra conclure à un changement dans la mesure ou l'on observe des changements dans la même direction sur plusieurs mesures apparentées. Par ailleurs, l'observation d'un effet sur un seul des indices, même s'il est important, pourra être considéré comme vraisemblablement dû au hasard.

2.2 Recrutement et sujets

Le recrutement des familles s'est effectuée entre le 1^{er} avril 1988 et le 31 août 1990. Toutes les familles répondant aux critères suivants ont été sollicitées pour participer à la recherche: 1^o au moins un enfant de la famille avait été signalé pour abus sexuel et l'abus sexuel avait été jugé fondé; 2^o l'abus sexuel avait été commis par un membre de la famille nucléaire ou de la famille élargie (parenté) de l'enfant; 3^o l'enfant victime d'abus sexuels devait avoir au moins 6 ans lors la première entrevue. Seuls les abus sexuels intrafamiliaux ont été considérés afin de diminuer l'hétérogénéité de l'échantillon. Au total, 71 familles ont été référées par le P.T.E.A.S. et étaient éligibles à participer à la recherche, de ce nombre 50 ont acceptées d'y participer soit un taux d'acceptation de 70% (50/71). Les principales raisons justifiant la non-participation à la recherche ont été un manque de disponibilité des parents. Dans d'autre cas, certains membres de la famille voulaient oublier l'abus ou encore l'agresseur ne reconnaissait pas l'abus sexuel (des démarches en Cour étaient amorcées et l'agresseur n'avait pas encore été reconnu coupable). Le Tableau II présente les caractéristiques des enfants ayant participé à la recherche. Au total 41 enfants ont été rencontrés à leur entrée dans le programme et 16 mois plus tard afin de remplir une batterie de tests. Près de trois victimes sur quatre (73%) sont de sexe féminin

et la moyenne d'âge au moment du signalement est d'un peu moins de 12 ans. L'examen de la structure familiale dans laquelle vivait l'enfant lors de l'abus sexuel montre qu'il y a autant de familles intactes (34%), que recomposées (31%) ou monoparentales (34%) ce qui vu sous l'angle de la population totale, indique une surreprésentation des deux derniers types de famille. Le revenu de ces familles est peu élevé puisque 44% d'entre elles ont un revenu inférieur à \$20,000 les situant sous le seuil de pauvreté (Conseil national du bien-être social, 1992).

2.3 Variables dépendantes et instruments de mesures

Plusieurs variables dépendantes ont été mesurées à l'aide d'instruments standardisés afin de déterminer les effets de la participation au programme. Des mesures de la dépression (Children's Depression Inventory de Kovac, 1983), du sentiment de compétence (The Pictorial Scale of Perceived Competence and Acceptance for Young Children (Harter, 1985) et une version pour adolescent The Self-Perception Profile for Adolescents (Harter, 1988)), du lieu de contrôle (Internal-External Scale for Children, Nowicki et Strickland, 1973), d'anxiété (Revised Children's Manifest Anxiety Scale de Reynolds et Richmond, 1985), d'affirmation de soi (Children's Action Tendency Scale de Deluty, 1979) et du réseau social de l'enfant (instrument-maison) ont été prises auprès de l'enfant.

3. PRÉSENTATION DES RÉSULTATS

3.1 Description de l'implantation du programme

Le Tableau III montre que dans l'ensemble, les victimes ont peu participé aux différents volets du programme. Bien que la durée moyenne de prise en charge par les services sociaux soit de 15 mois, les enfants ont participé en moyenne à 14 rencontres de thérapie individuelle, dyadique et familiale. La thérapie individuelle est la forme de thérapie la plus utilisée par les intervenants sociaux: 83% des victimes en ont reçu et le nombre de rencontres thérapeutiques moyen était de 12 rencontres par enfant. À l'inverse, les thérapies dyadique et familiale ont été peu utilisées auprès des enfants: un enfant sur deux (51%) n'a reçu aucune thérapie dyadique et 85% n'ont eu aucune thérapie familiale. Enfin, concernant le volet groupe, les enfants ont participé à 7 rencontres en moyenne (Tableau III). Plus de la moitié des enfants (56%) n'ont assisté à aucune rencontre de groupe.

3.2 Effets du programme

Les Tableaux IV, V et VI présentent l'ensemble des résultats concernant les effets de chacun des volets en fonction des variables mesurées. Pour chaque échelle ou sous-échelle, on retrouve à la droite du tableau, la moyenne, l'écart type et le nombre d'enfants pour lesquels l'information a été obtenue et ce, pour chacun des deux (2) temps de mesure. À la gauche du tableau, on retrouve les effets de chacun des volets exprimés par la corrélation partielle de Pearson, l'intervalle de confiance et l'effet thérapeutique (en termes d'ampleur et de direction). Par exemple, concernant la dépression, le Tableau

IV montre que la participation au volet I.D.F. est associée négativement à la dépression ($r = -.30$), c'est à dire que plus l'enfant participe et plus la dépression diminue. Il s'agit d'un effet léger et l'intervalle de confiance qui accompagne cette mesure situe l'effet réel entre $-.48$ et $-.08$. Autrement dit, on peut être sûr à 80% de la vraisemblance que l'effet réel se situerait dans cet intervalle.

3.3.1 Effets associés au volet I.D.F.

La participation aux rencontres offertes dans le cadre du volet I.D.F. est associée à plusieurs effets positifs sur l'état psychologique de l'enfant. Les effets moyens sont: 1^o une amélioration du sentiment de compétence global ($r = .32$) avec une influence particulière de la sous-échelle de compétence athlétique ($r = .30$); 2^o une diminution du lieu de contrôle (externalisation) ($r = -.39$); 3^o une diminution de l'insatisfaction dans les relations avec les personnes significatives du réseau social ($r = -.34$). Les effets légers sont: 1^o une diminution de la dépression ($r = -.30$); 2^o une diminution des sous-échelles d'anxiété physiologique ($r = -.18$) et sociale ($r = -.18$); 3^o une augmentation de trois sous-échelles du sentiment de compétence soit la compétence cognitive ($r = .25$), la perception de l'apparence physique ($r = .20$) et celle d'acceptation sociale ($r = .21$). La participation au volet est toutefois reliée à des effets négatifs¹ sur l'enfant, plus l'enfant y participe et moins le volet augmente (ou plus il détériore) sa satisfaction face aux relations avec les personnes du réseau social ($r = -.35$), le nombre de personnes dans son réseau social ($r = -.25$) et plus il augmente la soumission de l'enfant ($r = .13$). Enfin, la participation au volet I.D.F. n'a aucun effet sur l'assertion ($r = -.06$), l'agression ($r = -.06$), le score global d'anxiété ($r = -.04$) et la sous-échelle préoccupation ($r = .079$), sur la compétence sociale ($r = -.01$) et la perception de sa valeur personnelle ($r = .05$).

3.3.2 Effets associés au volet de groupe

Les Tableaux IV, V et VI montrent que la participation aux rencontres de groupe n'a aucune effet sur la dépression ($r = .04$), l'anxiété ($r = -.04$), les réactions de l'enfant à savoir l'agression ($r = .02$), la soumission ($r = -.05$) et l'assertion ($r = -.02$), le nombre de personnes significatives du réseau social ($r = -.08$) et l'insatisfaction dans les relations avec les personnes du réseau social ($r = -.07$) et la compétence sociale ($r = -.02$). La participation aux rencontres de groupe présente toutefois des effets négatifs sur le sentiment de compétence, le lieu de contrôle et la satisfaction face aux relations avec les personnes significatives du réseau social. En termes d'effets légers, les résultats montrent que plus les enfants participent aux rencontres de groupe et moins ils s'améliorent (ou plus ils se détériorent) au niveau du sentiment de compétence global ($r = -.26$), des sous-échelles d'apparence physique ($r = -.27$) et de perception de sa valeur personnelle ($r = -.26$). Les effets négatifs légers des rencontres de

¹ L'effet négatif se définit comme une détérioration d'un aspect que le programme cherche à améliorer ou une amélioration d'un aspect que le programme cherche à améliorer mais une amélioration moins importante que celle obtenu pour les enfants qui ne participent pas ou très peu. Par exemple, l'association négative entre la participation aux rencontres de groupe et le sentiment de compétence peut signifier que plus l'enfant participe au groupe et plus son sentiment de compétence diminue, cette association peut également signifier que plus l'enfant participe au groupe et moins son sentiment de compétence augmente.

groupe concernent également le lieu de contrôle (externalisation) ($r = .21$), la satisfaction face aux personnes significatives du réseau social ($r = -.10$) et trois sous-échelles du sentiment de compétence soit la compétence cognitive ($r = -.10$), la compétence athlétique ($r = -.17$) et la perception de sa valeur personnelle ($r = -.27$). Notons qu'aucun effet positif des rencontres de groupe n'a été décelé.

Dans l'ensemble, les résultats montrent deux patterns distincts des effets de la participation au programme. La participation au volet I.D.F. est associée un effet positif léger sur la majorité des aspects évalués, alors que la participation aux rencontres de groupe n'a aucun effet ou a des effets négatifs légers sur les variables examinées.

CONCLUSION

L'exemple qui précède illustre bien l'apport d'informations supplémentaires que peut fournir l'utilisation de l'ampleur de l'effet et de l'intervalle de confiance dans le contexte de l'évaluation de programme. Cette stratégie est de plus en plus reconnue comme une stratégie pouvant améliorer l'analyse statistique des données scientifiques. Récemment, un groupe de travail sur l'inférence statistique mis sur pied par l'American Psychological Association, recommandait certaines approches afin d'augmenter la qualité de l'utilisation des données et afin de se protéger contre certaines erreurs d'interprétations des résultats quantitatifs. Parmi ces approches, le groupe (Task Force on Statistical Inference, 1997) recommandait:

(b) enhanced characterization of the results of analyses (beyond simple p value statements) to include both direction and size of effect (e.g., mean difference, regression and correlation coefficients, odds-ratios, more complex effect size indicators) and their confidence intervals should be provided routinely as part of the presentation. These characterizations should be reported in the most interpretable metric (e.g., the expected unit change in the criterion for a unit change in the predictor, Cohen's d). (p. 2)

BIBLIOGRAPHIE

- Cohen, J. (1990). Things i have learned (so far). American Psychologist, 45 (12), 1304-1312.
- Cohen, J. & Cohen, R. (1977). Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, NJ: Earlbaum.
- Conseil national du bien-être social (1992). Profil de la pauvreté, 1980 à 1990. Ottawa: Ministère des Approvisionnements et Services Canada.
- Cook, T.J., & Poole, W.K. (1982). Treatment implementation and statistical power. Evaluation Review, 6, 425-430.
- Deluty, R.H. (1979). Children's Action Tendency Scale: A self-report measure of aggressiveness, assertiveness, and submissiveness in children. Journal of Consulting and Clinical Psychology, 47 (6), 1061-1071.
- Giarretto, H. (1981). A comprehensive child sexual abuse treatment program. In In P.B. Mrazek & C.H. Kempe (Eds.) "Sexually abused children and their families" (pp. 179-198). Oxford, England: Pergamon Press: .
- Harter, S. (1988). Manual for the Self-Perception profile for Adolescents. Denver: University of Denver.
- Harter, S., & Pike, R.G. (1980). The Pictorial Scale of Perceived Competence and Acceptance for Young Children. Denver: University of Denver.
- Kovacs, M. (1983). The Children's Depression Inventory: A self-rated Depression scale for school-aged youngsters. Unpublished manuscript, University of Pittsburg.
- Lebeau, T.M., & Benoît, D. (1984). Rapport sur le programme de traitement des enfants abusés sexuellement. Repentigny: Centre des Services Sociaux de Laurentides-Lanaudière.
- Lipsey, M.W. (1988). Practice and malpractice in evaluation research. Evaluation Practice, 9 (4), 5-24.
- Lipsey, M.W. (1990). Design sensitivity: Statistical power for experimental research. Beverly Hills, CA: Sage Publication.
- Morrison, D.E., & Henker, R.E. (1970). The significance test controversy: A reader. Chicago: Adline Publishing Company.
- Nowicki, S., & Strickland, B.R. (1973). A locus of control scale for children. Journal of Consulting and Clinical Psychology, 40, 148-155.
- Oakes, M. (1986). Statistical inference: A commentary for social and behavioral sciences. New York: Wiley.
- Reichardt, C.S., & Gollob, H.F. (1987). Taking uncertainty into account when estimating effects. In in M.M. Mark, & R.L. Shotland (Eds.) "Multiple methods in program evaluation" New directions for program evaluation, no 35. (pp. 7-22). San Francisco: Jossey-Bass.

Reynolds, C.R., & Richmond, B.O. (1985). Revised Children's Manifest Anxiety Scale (R.C.M.A.S.). , 102 pages.

Rosnow, R.L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. Psychological Bulletin, 44(10), 1276-1284.

Salsburg, D.S. (1985). The religion of statistics as practiced in medical journals. American Statistician, 39, 220-223.

Sgoui, M.F. (1986). L'agression sexuelle et l'enfant: approche et thérapie. St-Laurent: Éditions du Trécarré.

Task Force on statistical inference (1997). Initial report. Washington, DC: APA.

Tableau I
STRUCTURE DU PROGRAMME

MODULES ÉLÉMENTS	DURÉE PRÉVUE DU TRAITEMENT (ENFANTS)
MODULE ÉVALUATION	1 SEMAINE
1.1 AUPRÈS DE LA VICTIME	
1.2 AUPRÈS DE LA MÈRE	
1.3 AUPRÈS DE L'ABUSEUR	
MODULE THÉRAPIE INDIVIDUELLE	12 à 24 rencontres
2.1 THÉRAPIE INDIVIDUELLE AVEC LA VICTIME	
2.2 THÉRAPIE INDIVIDUELLE AVEC LA MÈRE	
2.3 THÉRAPIE INDIVIDUELLE AVEC L'ABUSEUR	
MODULE THÉRAPIE DYADIQUE	12 à 24 rencontres
3.1 THÉRAPIE DYADIQUE MÈRE/VICTIME	
3.2 THÉRAPIE DYADIQUE CONJUGALE	
3.3 THÉRAPIE DYADIQUE ABUSEUR/VICTIME	
MODULE THÉRAPIE FAMILIALE	12 rencontres min.
MODULE ENTRAIDE	24 à 36 rencontres
5.1 Thérapie de groupe	5.2 Auto-croissance
5.1.1 Groupe victimes	5.2.1 Info./référence
5.1.2 Groupe abuseurs	5.2.2 Parrainage
5.1.3 Groupe mères	
5.1.4 Groupe couples	
DURÉE TOTALE DU PROGRAMME	60 à 96 rencontres

Tableau II
Caractéristiques des enfants abusés sexuellement et
des familles où ils vivaient lors de l'abus sexuel

Caractéristiques reliées aux victimes d'abus sexuels	N	%
SEXE DE LA VICTIME		
Masculin	11	26.8
Féminin	30	73.2
Nombre total d'enfants	41	100.0
AGE DE LA VICTIME AU T1 (Age moyen de 11.6 ans)		
6 à 8 ans	9	22.0
9 à 11 ans	12	29.3
12 à 14 ans	12	29.3
15 à 17 ans	8	19.5
Nombre total d'enfants	41	100.0
STRUCTURE FAMILIALE		
Famille intacte	11	34.4
Famille recomposée	10	31.3
Famille monoparentale	11	34.4
Nombre total de familles	32	100.0
REVENU FAMILIAL		
Moins de \$10,000	2	6.3
\$10,000 à \$19,999	12	37.5
\$20,000 à \$29,999	5	15.6
\$30,000 à \$39,999	5	15.6
\$40,000 et plus	6	18.8
Données manquantes	2	6.3
Nombre total de familles	32	100.0

Tableau III
Caractéristiques reliées à l'implantation du programme (PTEAS)
auprès des enfants

Caractéristiques du traitement reçu	Moyenne par enfant	N = 41	%
DURÉE de la prise en charge (mois)	14.6		
RENCONTRES INDIVIDUELLES	11.9		
Aucune rencontre		7	17.1
1 à 9 rencontres		18	43.9
10 à 18 rencontres		6	14.6
19 à 27 rencontres		4	9.8
28 à 44 rencontres		6	14.6
RENCONTRES DYADIQUES	1.5		
Aucune rencontre		21	51.2
1 à 9 rencontres		18	43.9
10 à 18 rencontres		2	4.9
RENCONTRES FAMILIALES	0.7		
Aucune rencontre		35	85.4
1 à 11 rencontres		6	14.6
TOTAL RENCONTRES (VOLET I.D.F.)	14.1		
Aucune rencontre		6	14.6
1 à 9 rencontres		18	43.9
10 à 18 rencontres		5	12.2
19 rencontres et plus		12	29.3
RENCONTRES DU VOLET GROUPE	7.0		
Aucune rencontre		23	56.1
1 à 9 rencontres		8	19.5
10 à 18 rencontres		4	9.8
19 à 51 rencontres		6	14.6

TABLEAU IV

Impact du volet thérapie individuelle, dyadique et familiale (I.D.F.) et du volet de groupe (J.U.) sur le réseau social de l'enfant¹, la dépression² et le lieu de contrôle³

Temps	Données descriptives			Volet	Effets des volets			Effet thérapeuti. ⁴
	Moyenne	σ	N		R partiel	Interval de conf. 80%		
1. RÉSEAU SOCIAL DE L'ENFANT								
1.1 Nombre de personnes dans le réseau social								
0 mois	12.26	8.38	39	I.D.F.	-.246	[-.437 à -.034]		Léger et négatif
16 mois	15.26	9.29	39	Groupe	-.076	[-.285 à .141]		Aucun effet
1.2. Satisfaction face aux personnes du réseau social								
0 mois	22.76	17.54	37	I.D.F.	-.345	[-.525 à -.135]		Moyen et négatif
16 mois	24.27	14.69	37	Groupe	-.103	[-.316 à .120]		Léger et négatif
1.3. Insatisfaction face aux personnes du réseau social								
0 mois	0.78	1.51	37	I.D.F.	-.341	[-.522 à -.130]		Moyen et positif
16 mois	0.27	0.84	37	Groupe	-.070	[-.286 à .153]		Aucun effet
2. DÉPRESSION								
0 mois	10.89	6.85	38	I.D.F.	-.295	[-.481 à -.083]		Léger et positif
16 mois	8.29	6.08	38	Groupe	.038	[-.180 à .253]		Aucun effet
3. LIEU DE CONTRÔLE (EXTERNALISATION)								
0 mois	12.30	5.27	30	I.D.F.	-.390	[-.581 à -.158]		Moyen et positif
16 mois	9.70	4.67	30	Groupe	.174	[-.076 à .404]		Léger et négatif

- (1) Cette variable a été mesurée à l'aide d'un questionnaire maison, elle regroupe trois aspects soit: un nombre de personnes significatives dans le réseau social de l'enfant, un score de satisfaction et un d'insatisfaction vis-à-vis des relations avec les personnes du réseau social. Plus le score est élevé et plus il y a de personnes, de satisfaction ou d'insatisfaction.
- (2) Cette variable a été mesurée à partir d'une traduction du questionnaire "Children's Depression Inventory" de Kovacs (1983).
Il s'agit d'un score global de dépression. Plus le score augmente et plus l'enfant est dépressif.
- (3) Cette variable est mesurée à l'aide du questionnaire de Nowicki et Strickland (1973) et montre que plus le score est élevé plus l'enfant se perçoit comme ayant du contrôle sur ce qu'il lui arrive. Cette variable est souvent corrélée à l'estime de soi, la dépression et a été choisie parce qu'elle se rapproche du concept de victimisation.
- (4) Le premier terme désigne l'ampleur de la corrélation observée i.e. de l'effet du traitement (soit léger, moyen et fort) alors que le second terme désigne le fait que la participation au volet a un effet positif sur l'enfant ou bien un effet négatif. L'effet négatif est défini comme le fait qu'une participation au volet est associée à une moins grande amélioration (comparativement à ceux qui participent moins) ou bien carrément à une détérioration de la variable concernée.

TABLEAU V

Impact du volet thérapie individuelle, dyadique et familiale (I.D.F.) et du volet de groupe (J.U.) sur l'anxiété¹ et les réactions² de l'enfant

Données descriptives				Effets des volets			
Temps	Moyenne	σ	N	Volet	R partiel	Interval de conf. 80%	Effet thérapeuti.
1. ANXIÉTÉ TOTALE							
0 mois	12.10	6.66	39	I.D.F.	-.041	[-.253 à .174]	Aucun effet
16 mois	8.82	6.14	39	Groupe	-.035	[-.247 à .181]	Aucun effet
1.1 Anxiété physiologique (sous-échelle)							
0 mois	3.56	2.36	39	I.D.F.	-.175	[-.375 à .041]	Léger et positif
16 mois	2.85	2.18	39	Groupe	.073	[-.143 à .282]	Aucun effet
1.2 Préoccupations (sous-échelle)							
0 mois	5.71	3.26	39	I.D.F.	.079	[-.137 à .288]	Aucun effet
16 mois	4.21	3.38	39	Groupe	-.034	[-.247 à .181]	Aucun effet
1.3 Anxiété sociale (sous-échelle)							
0 mois	2.82	1.92	39	I.D.F.	-.175	[-.375 à .040]	Léger et positif
16 mois	1.77	1.51	39	Groupe	-.030	[-.242 à .185]	Aucun effet
2 RÉACTIONS DE L'ENFANT							
2.1 Assertion (sous-échelle)							
0 mois	20.03	2.99	39	I.D.F.	-.058	[-.269 à .158]	Aucun effet
16 mois	20.51	3.14	39	Groupe	-.015	[-.229 à .199]	Aucun effet
2.2 Agression (sous-échelle)							
0 mois	6.08	4.16	39	I.D.F.	-.063	[-.273 à .153]	Aucun effet
16 mois	6.97	4.21	39	Groupe	.024	[-.191 à .237]	Aucun effet
2.3 Soumission (sous-échelle)							
0 mois	12.87	3.68	39	I.D.F.	.127	[-.089 à .332]	Léger et négatif
16 mois	11.49	3.99	39	Groupe	-.045	[-.256 à .171]	Aucun effet

- (1) Cette variable a été mesurée à partir d'une traduction du questionnaire "Revised Children's Manifest Anxiety Scale" de Reynolds et Richmond (1985). On retrouve un score globale d'anxiété et trois sous-échelles soit une d'anxiété sociale, de préoccupations et d'anxiété physique. Plus le score augmente et plus l'enfant est anxieux.
- (2) Le Children's Action Tendency Scale (Deluty, 1979) permet d'établir dans quelle mesure l'enfant qui se retrouve dans une situation où ses droits ne sont pas respectés aura tendance à réagir de façon agressive, assertive ou au contraire à adopter une position de soumission. Les scores à chacune de ces 3 sous échelles sont interreliés de telle sorte qu'une augmentation sur une des sous-échelles entraînent nécessairement une diminution sur une des deux autres sous-échelles. L'assertion est considérée comme la meilleure réaction.

TABLEAU VI

Impact du volet thérapie individuelle, dyadique et familiale (I.D.F.) et du volet de groupe (J.U.) sur le sentiment de compétence¹ de l'enfant

Données descriptives				Effets des volets			
Temps	Moyenne	σ	N	Volet	R partiel	Interval de conf. 80%	Effet thérapeuti.
1. SENTIMENT DE COMPÉTENCE							
0 mois	2.83	0.57	39	I.D.F.	.321	[.114 à .500]	Moyen et positif
16 mois	2.99	0.43	39	Groupe	-.257	[-.447 à -.122]	Léger et négatif
1.1 Compétence cognitive (sous-échelle)							
0 mois	2.75	0.83	39	I.D.F.	.250	[.038 à .440]	Léger et positif
16 mois	3.09	0.79	39	Groupe	-.102	[-.309 à .115]	Léger et négatif
1.2 Compétence sociale (sous-échelle)							
0 mois	2.85	0.82	39	I.D.F.	-.005	[-.219 à .209]	Aucun effet
16 mois	3.12	0.70	39	Groupe	-.022	[-.235 à .193]	Aucun effet
1.3 Compétence athlétique (sous-échelle)							
0 mois	2.49	0.78	32	I.D.F.	.300	[.062 à .506]	Moyen et positif
16 mois	2.54	0.83	32	Groupe	-.138	[-.368 à .109]	Léger et négatif
1.4 Perception de son apparence physique (sous-échelle)							
0 mois	3.03	0.78	39	I.D.F.	.202	[-.013 à .399]	Léger et positif
16 mois	3.05	0.70	39	Groupe	-.273	[-.460 à -.063]	Léger et négatif
1.5 Acceptation sociale (sous-échelle)							
0 mois	2.78	0.74	32	I.D.F.	.205	[-.039 à .427]	Léger et positif
16 mois	2.91	0.61	32	Groupe	-.159	[-.386 à .087]	Léger et négatif
1.6 Perception de sa valeur personnelle (sous-échelle)							
0 mois	2.87	0.82	32	I.D.F.	.048	[-.197 à .287]	Aucun effet
16 mois	3.16	0.71	32	Groupe	-.263	[-.475 à -.021]	Léger et négatif

- (1) Une échelle développée par Harter (1980, 1988) permet de mesurer différents aspects du sentiment de compétence de l'enfant. Deux versions différentes ont été utilisées en fonction de l'âge de l'enfant. La première version, un questionnaire de 24 items (Pictorial Scale of Perceived Competence and Social Acceptance for Young Children de Harter, 1983), s'adresse aux enfants de 5 à 8 ans et comprend quatre sous-échelles touchant la compétence cognitive, la compétence athlétique, l'acceptation sociale et l'acceptation par la mère. La seconde version, une version abrégée du Self-Perception Profile for Adolescents de Harter (1988), compte 36 items et couvre les sous-échelles suivantes: la compétence cognitive, la compétence athlétique, l'acceptation sociale, la perception de son apparence physique, la conduite comportementale et la perception de sa valeur personnelle (self-worth). Les deux versions différentes expliquent que le nombre de sujets varie selon les sous-échelles.