

# La méta-analyse au service de la prévention et de l'intervention : concepts, applications et enjeux

## Using Meta-Analysis in Prevention and Intervention: Concepts, Applications and Issues

Françoise Fortin, Jonathan Lévesque et Frank Vitaro

Volume 36, numéro 1, 2007

URI : <https://id.erudit.org/iderudit/1097201ar>

DOI : <https://doi.org/10.7202/1097201ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Revue de Psychoéducation

ISSN

1713-1782 (imprimé)

2371-6053 (numérique)

[Découvrir la revue](#)

Citer cet article

Fortin, F., Lévesque, J. & Vitaro, F. (2007). La méta-analyse au service de la prévention et de l'intervention : concepts, applications et enjeux. *Revue de psychoéducation*, 36(1), 167–193. <https://doi.org/10.7202/1097201ar>

Résumé de l'article

Ce texte passe en revue les éléments d'ordre méthodologique et les enjeux d'ordre conceptuel sous-jacents à l'utilisation d'une méthode de plus en plus populaire de synthèse d'études corrélationnelles et évaluatives, la méta-analyse. Le contexte pour illustrer ces divers éléments en rapport avec la méta-analyse est celui des programmes de prévention et d'intervention auprès d'enfants et d'adolescents présentant des problèmes d'adaptation ou à risque d'en développer. La méta-analyse est présentée comme un outil de synthèse utile pour dégager les programmes les plus prometteurs ainsi que les conditions de réussite ou d'échec les entourant. Elle comporte cependant des éléments statistiques et des limites qu'il importe de connaître pour mieux en apprécier les résultats. Un tableau-synthèse décrit le contenu et les résultats de 19 méta-analyses réalisées dans le domaine de la prévention ou de l'intervention par rapport à diverses problématiques touchant les jeunes en difficulté.

## Mesure et évaluation

# La méta-analyse au service de la prévention et de l'intervention : Concepts, applications et enjeux

## Using Meta-Analysis in Prevention and Intervention: Concepts, Applications and Issues

F. Fortin<sup>1</sup>

J. Lévesque<sup>1</sup>

F. Vitaro<sup>1</sup>

1. École de psychoéducation,  
Université de Montréal

### Résumé

*Ce texte passe en revue les éléments d'ordre méthodologique et les enjeux d'ordre conceptuel sous-jacents à l'utilisation d'une méthode de plus en plus populaire de synthèse d'études corrélationnelles et évaluatives, la méta-analyse. Le contexte pour illustrer ces divers éléments en rapport avec la méta-analyse est celui des programmes de prévention et d'intervention auprès d'enfants et d'adolescents présentant des problèmes d'adaptation ou à risque d'en développer. La méta-analyse est présentée comme un outil de synthèse utile pour dégager les programmes les plus prometteurs ainsi que les conditions de réussite ou d'échec les entourant. Elle comporte cependant des éléments statistiques et des limites qu'il importe de connaître pour mieux en apprécier les résultats. Un tableau-synthèse décrit le contenu et les résultats de 19 méta-analyses réalisées dans le domaine de la prévention ou de l'intervention par rapport à diverses problématiques touchant les jeunes en difficulté.*

**Mots clés :** méta-analyse, applications, enjeux, prévention, intervention

### Abstract

*The meta-analysis is becoming more and more popular among researchers as a way of resuming results from multiples studies. This text reviews both conceptual and methodological aspects of this technique. The context used to discuss on meta-analysis is the one of prevention and intervention programs for children. The meta-analysis is presented as a useful tool to find the most effective programs and their success' conditions. This technique has statistical considerations and weaknesses that the reader has to be aware of in order to appreciate its results. A table resume the content and the results of 19 meta-analysis of prevention and intervention studies on externalized and internalized behaviour problems.*

**Keywords:** meta-analysis, statistical consideration, prevention, intervention

### Correspondance :

Françoise Fortin  
Université de Montréal  
École de psychoéducation  
C.P. 6128, succ. Centre-ville  
Montréal (Québec) H3C 3J7  
francoise.fortin@umontreal.ca

L'évaluation au cours des 25 dernières années d'un grand nombre de programmes de prévention et d'intervention pour les jeunes en difficulté a produit un nombre exponentiel de résultats souvent difficiles à synthétiser. Cette difficulté a été amplifiée par la grande variabilité de ces programmes au chapitre de leur contenu et de leurs effets (Schmidt & Hunter, 2003). Diverses stratégies ont été proposées afin de donner du sens à cette masse critique d'informations et ainsi en dégager des leçons utiles pour l'intervention préventive et curative.

Un des moyens fréquemment utilisés pour résumer les résultats de différentes études, particulièrement en sciences sociales, est la recension des écrits. Cette méthode de synthèse de type narratif comporte certains avantages, notamment au niveau de l'analyse détaillée des études primaires qu'elle permet. En revanche, il devient difficile, voire impossible, de garder la trace de tous les résultats obtenus et des paramètres pouvant expliquer la variabilité entre les études (c.-à-d. les variables modératrices) lorsque le nombre de recherches évaluatives sur une problématique donnée devient important (Shadish, Cook, & Campbell, 2002). En outre, la recension des écrits de type narratif, de par sa nature interprétative, laisse place à des biais possibles de la part de son auteur, ce qui par conséquent, risque de compromettre la validité et la généralisation des conclusions qui s'en dégagent.

Afin de pallier ces limites, les chercheurs se sont tournés vers des méthodes quantitatives de synthèse des connaissances. Ce désir de synthétiser quantitativement des résultats de recherche n'est pas nouveau. Déjà en 1904, Karl Pearson a regroupé 11 études afin d'établir l'efficacité du vaccin contre la typhoïde (Chalmers, Hedges, & Cooper, 2002). Pour ce faire, il a calculé des coefficients de corrélation pour chaque étude et en a fait la moyenne, pavant ainsi la voie aux synthèses de nature quantitative.

Il existe trois principales méthodes quantitatives de synthèse soit le décompte des votes (Bushman, 1994), la combinaison des degrés de signification (Becker, 1994) et la méta-analyse. Sans être exemptes de limites (qui seront discutées plus loin), cette dernière permet entre autres, contrairement aux autres méthodes, de tenir compte de la taille des échantillons ou des lacunes méthodologiques des études primaires et de tester la présence de possibles effets modérateurs. Ce dernier élément est particulièrement important, car une variable modératrice affecte l'ampleur et/ou la direction de la relation entre le programme et les effets observés (Baron & Kenny, 1986). Elle nous renseigne ainsi sur les clientèles et les milieux les plus susceptibles de bénéficier des interventions.

La méta-analyse doit son nom à Glass (1976) qui, le premier, propose ce terme pour désigner « l'analyse statistique d'un nombre important de résultats d'études individuelles avec l'objectif d'en intégrer les résultats (p.752) ». L'agrégation des multiples résultats se fait sur la base d'un indice commun : la taille de l'effet (TE ; effect-size). Le calcul d'un tel indice pour chaque étude pertinente rend possible la réalisation d'analyses comparables à celles effectuées sur les sujets dans une étude primaire (Lipsey, 1995), à la différence que les études sont considérées alors comme les sujets. Malgré de virulentes critiques quant à la

pertinence d'amalgamer de façon statistique des études aux différents devis, échantillons et variables dépendantes (voir entre autres, Eysenck, 1978; Mann, 1990, 1994), cette méthode s'est raffinée et est devenue un outil incontournable dans le domaine de la prévention et de l'intervention, son utilisation ayant quintuplé au cours des vingt dernières années (Field, 2003).

Dans le contexte de cet article, trois objectifs spécifiques sont abordés : 1) situer la contribution actuelle de la méta-analyse dans le domaine de la prévention et de l'intervention en fonction de ses forces, mais aussi de ses limites; 2) introduire, à l'aide d'exemples, les principaux concepts sous-jacents à cette méthode et 3) résumer les principaux enjeux méthodologiques, théoriques et cliniques y étant associés. Enfin, des pistes sont suggérées quant aux perspectives futures de cette méthode dans le domaine de la prévention et de l'intervention. Il est à noter que ce texte ne vise pas à former des experts de la méta-analyse, mais plutôt à introduire le lecteur novice aux rouages de base de cette procédure afin qu'il en devienne un consommateur averti.

### **Contributions de la méta-analyse**

Les premières méta-analyses, comme celle de Smith et Glass (1977), étaient davantage de nature descriptive. Elles tentaient de déceler un effet principal général afin de répondre aux questions des décideurs, à savoir quels programmes, thérapies ou interventions devraient être financés. Sous sa forme descriptive, la méta-analyse était considérée comme purement empirique et athéorique (Eysenck, 1978)

Depuis la méta-analyse de Smith et Glass (1977) sur l'efficacité des psychothérapies, le contexte a évolué. Outre l'expérimentation croissante d'interventions s'adressant à diverses problématiques, c'est le développement d'une « science de la prévention » au début des années 1990 (Coie, Watt, West, Hawkins, Asarnow, Markman, Ramey, Shure, & Long 1993) qui a marqué une nouvelle ère dans la recherche appliquée en sciences sociales. Se situant dans une perspective de psychopathologie développementale, cette nouvelle discipline cible les facteurs de risque et de protection afin de prévenir ou, à tout le moins, diminuer l'impact de diverses problématiques. Dans cet enthousiasme, une multitude de programmes ont été développés, implantés et évalués. Certains se sont révélés fort efficaces, d'autres moins ; d'autres encore ont produit des effets iatrogènes (c.-à-d. le programme a nuit aux participants ; ex. Dishion, McCord, & Poulin, 1999). Tout comme Smith et Glass (1977) l'avaient pressenti 30 ans auparavant, le besoin de dégager des constats à partir de ces nombreux essais préventifs est devenu urgent. Pour ce faire, l'utilisation d'une méthode de synthèse des connaissances, la méta-analyse, s'est imposée.

Cette technique permet de dégager les stratégies les plus efficaces en fonction d'une problématique (Weissberg, Kumpfer, & Seligman, 2003). Elle permet également de connaître l'impact de la qualité de la mise en œuvre et des caractéristiques des programmes ainsi que les clientèles et les milieux les plus susceptibles de bénéficier de l'intervention (Coie et al., 1993 ; Lochman, 2001). Le

tableau 1 présente une liste non exhaustive des résultats de méta-analyses selon différentes problématiques. Par ailleurs, la méta-analyse contribue à identifier les facteurs de risque ou de protection à cibler ainsi que les périodes optimales d'intervention. De plus, elle rend possible la validation des modèles théoriques qui ont présidé à la conception initiale des programmes de prévention ou d'intervention. La section suivante explore chacune de ces possibilités à tour de rôle en plus d'initier le lecteur aux éléments d'ordre conceptuel et statistique dont la compréhension est nécessaire afin d'apprécier les résultats d'une méta-analyse.

### **Évaluation des effets principaux ou sommes-nous efficaces?**

Après l'examen de plus de 300 méta-analyses, Lipsey et Wilson (1993) concluent qu'en général, les interventions éducationnelles, psychologiques et comportementales sont efficaces, mais que leur efficacité varie en fonction des clientèles ciblées, des types d'intervention ou du milieu dans lequel celles-ci sont implantées. Il est à noter qu'aucune des études primaires comprises dans ces 300 méta-analyses n'utilisait la même méthode, les mêmes programmes et exactement les mêmes construits. Le lecteur avisé doit maintenant se demander sur quelle base statistique la méta-analyse rend les diverses études primaires comparables.

L'indice de taille de l'effet (indice TE) est le dénominateur commun qui permet de comparer les résultats d'études primaires entre elles. Contrairement au degré de signification, l'indice TE tient compte de la taille de l'échantillon, ce qui rend plus facile la comparaison des études. Dans sa démarche analytique, le méta-analyste doit calculer l'indice TE moyen pour l'ensemble des études, de même que pour chacun des construits de la variable dépendante (effets du programme). Par exemple, il est possible qu'une méta-analyse regroupe des mesures sur la récidive, les difficultés de comportement et les connaissances sur la résolution de conflits, il faut donc calculer un indice TE pour chacune de ces mesures ainsi qu'un indice TE représentant l'ensemble de ces mesures.

Il existe plusieurs formules pour calculer les TE. Dans le domaine de la prévention et de l'intervention, trois principaux types cohabitent : les indices de la famille des  $d$ , qui impliquent une comparaison entre deux groupes, les rapports de risque, lorsque la variable dépendante est dichotomique et ceux de la famille des  $r$ , qui se rapportent aux relations entre deux variables (Durlack, 1995 ; Lipsey & Wilson, 2001a ; McCartney & Rosenthal, 2000 ; Rosenthal, 1995 ; Shadish et al., 2002).

*La famille des  $d$ .* L'indice  $d$  repose sur la comparaison entre deux groupes au niveau d'une mesure dépendante continue. Il est possible de calculer un indice standardisé et un indice non standardisé, ce dernier requérant que la mesure des effets du programme soit la même pour toutes les études primaires concernées. Pour obtenir une TE de la famille des  $d$ , il suffit de calculer la différence entre la moyenne du groupe expérimental et celle du groupe contrôle. Afin de standardiser cet indice, il faut ensuite diviser cette différence par l'écart-type regroupé (c.-à-d. commun aux deux groupes) ou celui du groupe de contrôle (Lipsey & Wilson, 2001a). Le choix de l'écart-type à utiliser est fonction du type d'effet escompté du

programme. Ainsi, lorsque l'intervention vise à réduire l'écart entre les membres d'un groupe donné (ex. : que les enfants les moins performants réussissent davantage à un test de mathématiques), il est préférable d'utiliser l'écart-type du groupe contrôle, car la variance (variabilité) du groupe expérimental devrait être réduite par le programme. Parmi les méta-analyses recensées au tableau 1, seulement trois d'entre elles (Grossman & Hugues, 1992 ; Serketich & Dumas, 1996 ; Weisz, Weiss, Han, Granger, & Morton, 1995) ont utilisé l'écart-type du groupe contrôle, les autres ayant préféré l'écart-type regroupé. Les *d* sont la famille d'indices la plus souvent utilisée dans le domaine de la prévention et de l'intervention. D'ailleurs, toutes les méta-analyses du tableau 1, à l'exception de celle de Wilson, Lipsey, & Derzon, (2003) se basent sur cette famille d'indices.

*Le rapport de risque.* Dans le cas où la variable dépendante est dichotomique, le calcul du rapport de risque (odds ratio) s'avère préférable. Cet indice réfère à la probabilité qu'un événement particulier survienne davantage dans un groupe que dans un autre, tout en tenant compte du taux de base de la survenue de cet événement (par exemple, le taux de récurrence chez deux groupes de jeunes délinquants, l'un ayant bénéficié d'une intervention, l'autre pas).

*La famille des *r*.* Les indices relevant de la famille des *r* sont utiles lorsque l'objectif consiste à vérifier l'association entre deux variables. Ainsi, un chercheur pourrait être intéressé à savoir si la qualité de l'attachement parent-enfant est liée à la sensibilité parentale. Le coefficient de corrélation de Pearson, indice bien connu, est le représentant type de cette famille. Toutefois, il ne s'utilise qu'avec des variables continues. Lorsqu'une des variables est continue et l'autre dichotomique, la corrélation au point biserialle doit être utilisée.

Fait à noter, il est possible de calculer d'autres indices TE. Par exemple, on peut contraster les moyennes pré et post-test chez un seul groupe de participants dans le cas de devis sans groupe de comparaison comme l'ont fait Wilson et al. (2003) ou calculer des proportions lorsqu'il n'y a qu'une seule variable. Ces calculs sont cependant peu utilisés dans le domaine des sciences sociales et présentent un intérêt secondaire. Enfin, il est possible que les différentes études qui sont incluses dans la méta-analyse ne se basent pas sur les mêmes familles d'indices, certaines transformations doivent alors être effectuées afin de reporter toutes les études sur un indice commun (généralement le *d*). Cependant, ce cas de figure est peu fréquent dans le domaine de l'intervention. Par exemple, parmi les études du tableau 1, seulement 4 d'entre elles ont dû effectuer ces transformations (Farrington & Welsh, 2003; Ripens, Aleman, & Goudena, 1997; Tobler, Roona, Ochshorn et al., 2000 ; Wilson, Gottfredson, & Najaka, 2001). Le lecteur intéressé par une discussion plus approfondie de ces méthodes est référé à Wilson et Lipsey (2001a).

Par ailleurs, Cohen (1977) offre des balises permettant d'interpréter l'importance des indices TE. Cet auteur propose des valeurs repères pour déterminer si l'indice représente un petit, un moyen ou un grand effet. Même si cette méthode comporte des limites en raison de la relativité des valeurs repères selon le

domaine d'étude, sa simplicité et son caractère pratique en ont popularisé l'utilisation.

Pour les indices de la famille des  $d$ , une valeur TE de .20 est considérée comme petite, .50 moyenne et .80 grande. Pour illustrer son propos, Cohen avance qu'un  $d$  de .50 serait assez grand pour que la différence entre deux groupes soit « visible à l'œil nu », par exemple la différence de connaissances entre un diplômé du collégial et un diplômé de l'école secondaire. Un  $d$  de .80, quant à lui, pourrait s'observer au niveau de la différence de connaissances entre un doctorant et un étudiant au collège.

Pour les indices de la famille des  $r$ , Cohen (1977) propose de considérer un  $r$  de .10 comme petit, un  $r$  de .30 comme moyen et un  $r$  de .50 comme grand. Une fois encore, l'auteur appelle à la prudence. En effet, les indices TE doivent être interprétés en fonction du domaine d'intérêt. Prentice et Miller (1992) rapportent qu'un petit indice TE peut laisser entrevoir des effets impressionnants. C'est le cas, par exemple, lorsque la variable dépendante est influencée par une variable indépendante pour laquelle on prédisait a priori une relation nulle. Afin d'illustrer leur affirmation, ces auteurs citent une étude qui a démontré que la beauté physique influençait positivement les décisions judiciaires. Dans cet exemple, même si l'indice TE est considéré comme petit, il n'en demeure pas moins que ce résultat comporte des conséquences réelles (des gens moins beaux courant plus de risque d'être condamnés) et que ce résultat demeure impressionnant considérant l'a priori d'objectivité de la justice (hypothèse d'une relation nulle). Suite à ces considérations conceptuelles et statistiques, l'examen du Tableau 1 nous permet de constater que les interventions auprès de certaines populations ciblées semblent en effet donner des résultats encourageants. Par exemple, les enfants et adolescents présentant des comportements antisociaux (TE = .62 et .86; Ang & Hughes, 2002; Serketich & Dumas, 1996), ou un trouble de déficit d'attention avec hyperactivité (TE = .45, DuPaul & Eckert, 1997) sont parmi les clientèles bénéficiant le plus des programmes qui leur sont offerts. À l'inverse, la prévention et l'intervention auprès des adolescentes à risque de grossesse précoce (TE = .15; Corcoran, O'Dell-Miller, & Bultman, 1997) et auprès des jeunes contrevenants récidivistes (TE = .12; Lipsey & Wilson, 1998) semblent peu efficaces et appellent au développement de nouveaux programmes ou stratégies.

Au niveau des types d'intervention, l'indice TE moyen obtenu par Grossman et Hughes (1992) suggère, par exemple, que les programmes basés sur l'acquisition d'habiletés d'autocontrôle chez des jeunes manifestant des troubles internalisés sévères présentent une bonne efficacité (TE = .74). Des conclusions similaires peuvent être tirées des programmes d'habiletés parentales de type comportemental auprès de familles dont un des enfants présente des comportements antisociaux (TE = .86; Serketich & Dumas, 1996). Toutefois, les programmes d'acquisition d'habiletés sociales semblent donner des résultats plutôt modestes (i.e. TE = .20; Quinn, Kavale, Mathur, Rutheford, & Forness 1999).

Tableau 1. Tableau-synthèse de 19 méta-analyses du domaine de la prévention

Auteurs	Domaine	Nombre d'études incluses	Population ciblée	Contexte d'intervention	Construits mesurés	Modérateurs étudiés	Modèle utilisé	Étendue et TE moyen (famille des d)	Qualité méthodologique des études primaires	Importance pratique
Ang & Hughes (2002)	Problèmes de comportement	38	Enfants et adolescents antisociaux de 6-18 ans	n.d.	Ajustement social et comportemental	Composition des groupes, type de VD, type de traitement, type de groupe de comparaison, expérience du thérapeute, type de population	fixe	- .83 à 2.51 (.62)	4**	Taux de succès de 73% supérieur pour le groupe expérimental au niveau de l'ajustement social
Bakermans-Kranenburg, van Ijzendoorn, & Juffer (2003)	Attachement	70	Enfants, adolescents et/ou parents	Varié	Attachement, sensibilité maternelle	Devis, type de VD	aléatoire	n.d. (.33)	3***	n.d.
Corcoran, O'Dell-Miller, & Bultman (1997)	Prévention des grossesses à l'adolescence	32	Adolescentes entre 11 et 20 ans	Scolaire et communautaire	Activité sexuelle (AS), utilisation de la contraception (C) et taux de grossesses (G)	Devis, type d'intervention, type de programme et caractéristiques des participants	fixe	AS= -.424 à .571 (.011) C= -.2945 à 1.151 (.27) G= -.2517 à .4735 (.153)	4**	n.d.
Dew, Bromet, Brent, & Greenhouse (1987)	Suicide	5	Personnes suicidaires	Centres de prévention du suicide	Taux de suicide	Aucun	fixe	-1.186 à 1.129 (-.004)	4**	n.d.
DuPaul & Eckert (1997)	TDAH	63	Enfants TDAH de 5 à 15 ans	Scolaire	Problèmes de comportements (C), performance académique (A)	Type d'intervention, nature de l'école, placement éducatif et statut de publication	n.d.	C= n.d. (.45) A= n.d. (n.d.)	3***	n.d.
Durlak & Wells (1997)	Générique	177	Jeunes de 18 ans et moins	Multiple	Symptômes externalisés et internalisés, performance académique, statut sociométrique, processus cognitifs, évaluations physiologiques	Type de programme, niveau développemental des participants, qualité méthodologique des études primaires	fixe	-.45 à 2.36 (.34)	4**	Taux de succès de 8 à 46% supérieur (selon les VD) pour le groupe expérimental



Grossman & Hughes (1992)	Problèmes internalisés	22	Enfants et adolescents de moins de 18 ans ayant un trouble internalisé sévère	Scolaire et clinique (auto-contrôle)	Comportement de l'enfant (anxiété, dépression)	Type de traitement (individuel ou en groupe), problématique de référence, nombre de thérapies, âge des sujets	fixe	.25 à 1.54 (.74)	4**	n.d.
Farrington & Welsh (2003)	Délinquance	40	Enfants, adolescents et leur famille	Familial	Comportements antisociaux court et long terme (CAC et CAI), délinquance court et long terme (Dc et DI)	Type de programme	fixe	CAC= .157 CAI= .016 Dc= .132 DI= .281	5**	Réduction entre 10 et 15% de la récidive pour le groupe expérimental
Lewis & Vosburgh (1988)	Intervention précoce	65	Enfants de maternelle	Scolaire (maternelle)	Intelligence, performance académique, lecture, math, langage, habiletés motrices et perceptuelles, perceptions auditives et visuelles, mémoire	Implication des parents, niveau de structure du programme	fixe	n.d. (.413)	4**	n.d.
Lipsey & Wilson (1998)	Délinquance	200	Délinquants juvéniles 14-18 récidivistes	Institution, milieu scolaire et familial	Récidive	Devis, institutionalisation, type de traitement, dosage, caractéristiques du programme, type d'offense précédente, rôle du chercheur	fixe	n.d. (.12)	4**	Réduction de la récidive de 12% pour le groupe expérimental
Quinn et al. (1999)	Problèmes affectifs et de comportement	35	Enfants et adolescents avec problèmes affectifs et de comportement	Varié (habiletés sociales)	Prosocialité, problèmes de comportements, comportements spécifiques	Validité théorique du programme, qualité de l'implantation, devis, programme démonstration ou de routine, durée de l'intervention, source d'évaluation, type de VD, instruments, âge	fixe	-1.32 à 2.136 (0,199)	3***	n.d.

Rispens, Aleman, & Goudena (1997)	Abus sexuels	16	Enfants 5-12 ans	Scolaire	Connaissances des concepts de l'abus sexuel, habiletés d'auto-protection	Qualité méthodologique, caractéristiques des enfants, caractéristiques des programmes	fixe	.14 à 1.85 (.71)	4**	n.d.
Serketich & Dumas (1996)	Problèmes de comportement	26	Enfants et adolescents antisociaux	Famille habiletés parentales de type comportemental	Ajustement parental (AP), comportement externalisé chez l'enfant (C)	Devis, programme original versus version améliorée	fixe	AP= n.d. (.44) C= n.d. (.86)	4**	Taux de succès de 80% pour le groupe expérimental au niveau de l'ajustement parental
Sweet & Appelbaum (2004)	Visite à domicile	60	Enfants et leurs parents	Familial	Enfants : dév. cognitif (C), dév. sociaffectif (S), abus (A) abus potentiel (AP), stress parental (S) Parents : comportements parentaux (CP), attitudes parentales (AP), éducation de la mère (EM), emploi (E), assistance publique (Ass)	Devis, population ciblée, objectifs visés par le programme	aléatoire	C= .184 S= .096 A= .318 AP= .239 S= .21 CP= .139 AP= .110 EM= .134 E= .017 Ass= -.043	4**	n.d.
Tobler et al. (2000)	Consommation de drogues et d'alcool	207	Adolescents au secondaire	Scolaire	Consommation de drogue	Type de programme, taille de l'échantillon, type d'intervenant, type de groupe contrôle, attrition, caractéristiques des sujets, intensité du programme, drogue ciblée	fixe	-.71 à 1.46 (.18)	5* et 4**	n.d.
Weisz et al. (1995)	Psychothérapie	150	Enfants et adolescents suivant une psychothérapie	Clinique	Multiple	Type d'intervention, type de problème, caractéristiques des sujets, qualification du thérapeute	fixe	n.d. (.54)	4**	n.d.

Wilson, Gottfredson, & Najaka (2001)	Problèmes de comportement	165	Enfants et adolescents 5-18 ans	Scolaire	Délinquance (D), consommation de drogue (C), décrochage scolaire (DS), antisocialité (A)	Type de programme, centration sur l'environnement ou l'individu, devis, type de VD, âge des sujets et niveau de risque	aléatoire	D= -.17 à .34 (n.d.) C= -.14 à .10 (n.d.) DS= -.20 à .29 (n.d.) A= .03 à .32 (n.d.)	4**	n.d.
Wilson & Lipsey (2000)	Délinquance	28	Adolescents 10-18 ans	Activité extérieure	Délinquance (D), habileté sociale (HS), locus de contrôle (LC), estime de soi (ES), ajustement psychologique (AP), ajustement scolaire (AS)	Répartition des sujets, types de VD, équivalence des groupes, niveau de risque de délinquance, âge, durée du traitement, intensité, thérapie complémentaire	fixe	D= .18 HS= .28 LC= .10 ES= .31 AP= .25 AS= .30	4**	Réduction de la récurrence de 8% pour le groupe expérimental
Wilson, Lipsey, & Derzon (2003)	Comportements agressifs	221	Enfants et adolescents 4-18 ans	Scolaire	Agressivité	Type d'intervention, devis, âge, type de programme, caractéristiques des sujets	mixte	-.07 à .74 (.25)	3***	Réduction du nombre de bataille de 7% dans le groupe expérimental

- \* méta-analyse composée d'études primaires ayant des devis expérimentaux uniquement  
\*\* méta-analyse composée d'études primaires ayant des devis expérimentaux et quasi-expérimentaux  
\*\*\* méta-analyse composée d'études primaires ayant des devis expérimentaux, quasi-expérimentaux et pré-expérimentaux  
\*\*\*\* méta-analyse composée d'études primaires ayant des devis quasi-expérimentaux et pré-expérimentaux  
\*\*\*\*\* méta-analyse composée d'études primaires ayant des devis pré-expérimentaux uniquement

Une autre information intéressante obtenue par les méta-analyses concerne l'efficacité des programmes de prévention universelle. Puisque ceux-ci s'adressent à toute une population, autant les individus à risque que ceux qui ne développeront jamais de problèmes, il peut paraître difficile d'obtenir des résultats probants, car le degré d'amélioration possible est faible pour la majorité de la population. Néanmoins, la méta-analyse de Durlack et Wells (1997) sur l'efficacité des programmes de prévention auprès des jeunes de moins de 18 ans et celle de Lewis et Vosburgh (1988) concernant les programmes offerts aux enfants de maternelle suggèrent que les programmes de prévention universelle ont un impact modéré, mais réel (des indices TE de .34 et .41 respectivement).

### **Les modérateurs ou l'étude des conditions d'implantation**

Outre l'indice TE moyen, le méta-analyste est aussi intéressé aux éléments susceptibles d'augmenter ou de diminuer l'efficacité des interventions que ce soit des caractéristiques des participants ou des conditions d'implantation. Au niveau statistique, certains auteurs (Hedges & Olkin, 1985; Lipsey & Wilson, 2001a) proposent d'utiliser un test d'homogénéité (test Q) afin de savoir si certaines variables influencent les effets observés. Le test Q nous renseigne sur l'hétérogénéité des indices TE pour l'ensemble des études recensées. Concrètement, il indique si toutes les études semblent provenir d'une même population ce qui sera indiqué par un niveau élevé d'homogénéité. Dans un tel cas, il est peu probable que des variables extérieures modèrent les résultats obtenus. À l'inverse, la présence d'hétérogénéité suggère que les études ne proviennent pas de la même population. Cette variabilité observée peut être expliquée par l'effet de variables modératrices connues ou inconnues. Le résultat au test Q permet donc au chercheur de sélectionner le modèle d'analyse qui correspond le mieux à ses données selon la présence ou non de variabilité entre les études et la connaissance ou l'ignorance des variables pouvant expliquer cette dernière. Trois modèles principaux d'analyse ont été développés soit les modèles à effets fixes, à effets aléatoires et à effets mixtes.

Le modèle à effets fixes suppose que toute la variabilité observée dans les études se limite à la variabilité entre les sujets de chaque étude (c.-à-d. à la variabilité intra-étude), à l'erreur d'échantillonnage. Le chercheur postule alors qu'il connaît toutes les variables qui peuvent rendre compte de l'hétérogénéité des indices TE (Hedges & Vevea, 1998; Raudenbush, 1994). Le modèle à effets aléatoires, quant à lui, suppose deux niveaux de variabilité : intra-étude (entre les sujets) et inter-études (entre les études). Ces variabilités inter-études sont présumées inconnues et aléatoires. Enfin, le modèle mixte postule qu'une partie de la variabilité entre les études est inconnue et aléatoire alors qu'une autre partie peut être expliquée par l'intermédiaire de variables modératrices connues.

Lorsque le chercheur postule qu'il connaît certaines ou toutes les variables pouvant influencer les effets des programmes, comme c'est le cas dans les modèles fixes et mixtes, il a recourt à des tests tels que l'ANOVA, la régression analogue (semblables à la régression linéaire multiple) ou la régression des moindres carrés

afin d'évaluer l'effet de ces variables (Lipsey & Wilson, 2001a). Dans le domaine de la prévention, la majorité des méta-analyses utilisent des ANOVA, c'est le cas de Lewis et Vosburg (1988), Serketich et Dumas (1996), Sweet et Appelbaum (2004) et Tobler et al. (2000) parmi les études du tableau 1. Par ailleurs, il est à noter qu'un terme d'erreur supplémentaire est ajouté à aux tests statistiques afin de tenir compte du type de sujets qui constituent la méta-analyse (c.-à-d. les études primaires). Cependant, tous les postulats de base associés à chacune des techniques statistiques précédentes s'appliquent.

Dans la pratique, l'identification de ces variables modératrices est importante, car les décideurs veulent connaître, outre l'efficacité générale des programmes, quelles sont les stratégies les plus prometteuses, auprès de qui, de quelle manière et dans quelles conditions les interventions sont les plus efficaces. Par exemple, Tobler et al. (2000) constatent, pour la prévention de la toxicomanie, que les interventions les plus efficaces sont interactives (i.e. favorisent la participation des sujets) et visent le développement des habiletés interpersonnelles. À l'opposé, se trouvent les interventions orientées vers le simple développement de connaissances à propos de la drogue, dont les effets sont généralement nuls ou négatifs. Pour les psychothérapies auprès des enfants et des adolescents, les thérapies de nature comportementale ciblant directement le problème de référence sont les plus efficaces (Weisz et al., 1995). Enfin, les interventions auprès d'une population de jeunes délinquants institutionnalisés ayant les indices TE les plus élevées sont les programmes de type comportemental et ceux visant le développement des habiletés interpersonnelles (Lipsey & Wilson, 1998).

Une autre question régulièrement soulevée est celle de l'efficacité des interventions qui sont habituellement appliquées par les intervenants des milieux scolaires (c.-à-d. les programmes de routine). Une méta-analyse a démontré que les programmes dits de démonstration, c'est-à-dire implantés dans le cadre d'une évaluation rigoureuse, étaient plus efficaces que les programmes de routine (Lipsey, 1995; Wilson et al., 2003). Ce résultat devrait amener les décideurs à développer des systèmes d'évaluation, de soutien et de monitoring des différentes interventions mises en place au sein de leurs organisations afin de s'assurer de l'efficacité de celles-ci et de la rigueur de leur mise en oeuvre.

Dans le domaine de la prévention de la délinquance, certaines études primaires (ex. Dishion et al., 1999) concluent que l'intervention mise en place a produit des effets iatrogènes. Une des hypothèses soulevées par ces études est que la composition du groupe (uniquement des jeunes présentant des problèmes de comportement) aurait pu favoriser un entraînement à la déviance. Ang et Hugues (2002) ont effectué une méta-analyse afin d'étudier systématiquement cette hypothèse. Ils en arrivent à la conclusion que les interventions comprenant des groupes mixtes (incluant des jeunes prosociaux) ainsi que les traitements individuels obtenaient des indices TE plus élevés que les groupes homogènes composés uniquement de jeunes ayant des difficultés de comportement. Ces résultats mettent l'accent sur l'importance d'inclure des enfants prosociaux lors de la composition des groupes en prévention de la délinquance.

Les interventions précoces, particulièrement au préscolaire, sont très populaires, mais qu'en est-il de leurs conditions gagnantes ? A ce sujet, Lewis et Vosburgh (1988) sont arrivés à la conclusion que les programmes davantage structurés ainsi que ceux qui impliquent les parents obtiennent des indices TE plus élevés que les autres programmes. Ces résultats nous renseignent sur les éléments à inclure lors du développement d'une intervention au préscolaire.

Enfin, un débat régulièrement soulevé est celui du choix de la cible d'intervention : l'individu ou l'environnement. À ce propos, Wilson et al. (2001), dans une méta-analyse sur la prévention des problèmes de comportement, ont relevé que les interventions centrées sur l'environnement sont particulièrement efficaces pour réduire la délinquance et la consommation de drogues et d'alcool. Ce résultat nous éclaire sur la cible à privilégier en fonction de la problématique à prévenir.

### **La validation de modèles théoriques**

Depuis quelques années, de plus en plus de chercheurs utilisent la méta-analyse afin de valider des modèles théoriques d'intervention ou de prédiction. Ce qui fait de la méta-analyse un outil puissant pour répondre au critère de réfutabilité de Popper ou de falsifiabilité de Kuhn (Cook, Cooper, Cordray, Hartmann, Hedges, Light, Louis, & Mosteller, 1992; Price, 1997), c'est qu'elle regroupe les données de plusieurs études qui peuvent être considérées comme autant de répliquations d'une étude originale tentant de démontrer que l'hypothèse nulle est fausse. Si l'hypothèse nulle résiste aux différents tests, la crédibilité de l'hypothèse alternative s'en trouve rehaussée. Dans le domaine de la prédiction, la méta-analyse permet également de mieux cerner les facteurs de risque et de protection associés au développement ou à la persistance de problèmes d'adaptation, autant en ce qui concerne les processus qui leur sont sous-jacents (médiateurs) ou les conditions (modérateurs) qui en diminuent ou augmentent les impacts (Cook et al., 1992).

Un moyen utilisé pour la validation est l'analyse de l'homogénéité des catégories des variables du modèle. Ce dernier est validé lorsque toutes les catégories proposées sont considérées homogènes (Price, 1997). À l'aide de cette technique, Durlak et Wells (1997) concluent que la typologie des programmes de prévention proposée par Coie et al. (1993) est généralement valide. Pour ce faire, les auteurs ont séparé les études recensées d'abord selon leur centration (personnes, moments de transition ou environnement), puis selon leur stratégie générale (éducation affective, résolution de problèmes interpersonnels, etc.). Par la suite, ils ont effectué des tests d'homogénéité pour chacune des catégories ainsi construites. Les résultats de la très grande majorité de ces tests étant non significatifs, cela leur a permis de conclure que les catégories proposées étaient homogènes, donc que le modèle de Coie et al. (1993) est valide.

### **Avancement des débats**

Finalement, la méta-analyse favorise l'avancement de certains débats. Par exemple, le domaine de la prévention de l'abus sexuel chez les enfants fait l'objet de

controverse, certains auteurs prétendant que ce type de prévention n'est pas efficace car elle ne cible pas la bonne clientèle et que des effets pervers ou iatrogènes peuvent y être associés tels que de faux dévoilements ou une diminution de la confiance envers l'adulte (Van Gijsegem, 1999). Cependant, Rispens et al. (1997) ont obtenu un indice TE moyen de .71 après l'analyse de 16 études dont les variables dépendantes étudiées étaient les connaissances et les habilités d'autoprotection. Malgré la taille restreinte de l'échantillon, ce résultat indique qu'il y a peut-être de la place pour de tels programmes et que d'autres études évaluatives sont nécessaires afin de confirmer la présence de potentiels effets iatrogènes.

### **Enjeux théoriques, méthodologiques et cliniques**

Les enjeux présentés dans la prochaine section donnent un aperçu de certaines critiques portées envers la méthode ainsi que les différents choix offerts aux concepteurs d'une méta-analyse afin d'y répondre.

#### **Biais de publication**

La sélection des études à inclure dans une méta-analyse est une étape cruciale directement associée à sa validité externe. Certains méta-analystes ne recensent que les études publiées, or, celles-ci présentent davantage de résultats significatifs que les études non publiées, les journaux scientifiques étant généralement peu enclins à publier des études aux résultats nuls (Begg, 1994). Ce biais de sélection, qui découle du biais de publication, influence les possibilités de généralisation des résultats de la méta-analyse, car l'indice TE risque alors d'être surestimé. Ce problème peut aussi se poser lorsque le chercheur ne réussit pas à repérer ou à mettre la main sur l'ensemble des études existantes en lien avec le phénomène étudié (« file drawer problem », Rosenthal, 1979). Rosenthal (1991) propose une façon d'estimer les dommages que pourrait causer ce biais de publication ou « file drawer problem ». Il s'agit de calculer le nombre d'études aux résultats non significatifs auxquelles le chercheur n'aurait pas eu accès et qui viendraient invalider ses conclusions. Comme seuil de tolérance, Rosenthal (1991) propose le critère arbitraire de  $5k + 10$  ( $k$  étant le nombre d'études en possession du méta-analyste). Il lui apparaît raisonnable de penser que le nombre d'études non recensées ne devrait pas dépasser cinq fois le nombre d'études récupérées (dans le cas d'une recension sérieuse); par conséquent, un résultat en deçà de  $5k + 10$  ne représente pas une menace sérieuse à la validité externe de la méta-analyse.

Peu de méta-analystes semblent se préoccuper d'un potentiel biais de publication. La lecture d'un échantillonnage de méta-analyses a démontré que leurs auteurs prennent pour acquis que leur stratégie de collecte des études primaires leur a permis de mettre la main sur la quasi-totalité des études existantes. Dans leur méta-analyse sur la prévention de l'abus sexuel, Rispens, et al. (1997) mentionnent avoir utilisé la méthode de Rosenthal pour vérifier l'impact du « file drawer problem » sans toutefois mentionner le résultat.

Par ailleurs, certains méta-analystes décident consciemment de ne retenir que les études soutenues par un devis méthodologique de qualité (c.-à-d. de nature expérimentale ou quasi expérimentale), laissant de côté les résultats obtenus avec des devis sans groupe de comparaison (Cook, 1991), augmentant le risque de retrouver un biais d'échantillonnage plus ou moins important. Certains auteurs jugent en effet indigne qu'une activité scientifique sérieuse tire des conclusions à partir d'études ayant des devis qui n'offrent pas suffisamment de protection au plan de la validité interne (ex. Eysenck, 1978). Par exemple, Farrington et Welsh (2003) ont décidé de retenir, pour leur méta-analyse sur l'efficacité des programmes centrés sur la famille en prévention de la criminalité, que les études comprenant une distribution aléatoire ou un pairage des sujets entre les conditions expérimentales et témoin. Rosenthal (1995, 1996) et Cook (1991) adoptent toutefois une position différente et arguent qu'une méta-analyse bien effectuée devrait permettre de tenir compte des différences méthodologiques initiales entre les études afin d'en évaluer l'influence possible sur les indices TE. Ces auteurs proposent alors de pondérer les études selon la qualité de leur méthodologie plutôt que de les écarter d'office puisque, de toute manière, il n'existe pas d'études parfaites (Schmidt & Hunter, 2003).

### **Les modérateurs**

La plupart des auteurs s'entendent maintenant pour dire qu'il est essentiel d'évaluer l'impact de la qualité de la méthodologie sur l'efficacité perçue des programmes (Lipsey & Wilson, 2001a; Durlak, 1995; Shadish et al., 2002; Lipsey, 1997; Hall & Rosenthal, 1991; Banger-Drown et al., 1997). A ce chapitre, la méthodologie des études primaires semble davantage responsable de la variation entre les indices TE provenant d'études différentes que les caractéristiques reliées à l'intervention elle-même ou aux conditions de mise en oeuvre (Lipsey, 1997, 2002). Dans le domaine de la prévention, la très grande majorité des études évaluent la qualité de la méthodologie comme variable modératrice et rapportent, à l'instar de Lipsey (1997, 2001) que cette dernière explique une partie importante de la variabilité observée entre les études. Par ailleurs, certains auteurs (ex. : Mann, 1994) considèrent que puisque les méta-analystes sélectionnent eux-mêmes les éléments de la qualité de la méthodologie qu'ils vont codifier, les analyses seraient donc biaisées dès le départ au même titre que les recensions des écrits. Ainsi, alors que des études telles que celles de Franklin et al. (1997) ou de Wilson et al. (2001) codifient un grand nombre d'éléments allant du type de devis, à la qualité des instruments en passant par les techniques d'échantillonnage et la rigueur des passations ; d'autres études (ex. : Dew, Bromet, & Brent, 1987; Lewis & Vosburg, 1988) n'évaluent pas la qualité de la méthodologie ou se contentent de codifier que le type de devis (ex. : Serketich & Dumas, 1996).

Plusieurs aspects autres que les éléments d'ordre méthodologique peuvent également exercer un effet modérateur sur les indices TE découlant de programmes de prévention ou d'intervention. Cependant, de nombreuses contraintes statistiques, méthodologiques et conceptuelles sont associées à l'étude des modérateurs potentiels. Ainsi, avec l'ajout d'un seul modérateur, le nombre minimal



d'études nécessaires pour mener à bien une méta-analyse peut aller jusqu'à doubler (Hedges & Pigott, 2004), les postulats associés aux diverses analyses statistiques devant être respectés. Par ailleurs, plusieurs études primaires ne rapportent pas d'informations au sujet des modérateurs susceptibles d'intéresser le chercheur. Celui-ci doit donc se contenter de données incomplètes, ce qui limite la généralisation des résultats des tests de modération effectués.

Il peut être difficile, voire impossible, d'attribuer un rôle causal à certains modérateurs puisque leur analyse découle souvent d'analyses post-hoc dans le cadre d'études corrélationnelles (c.-à-d. sans aucune manipulation expérimentale) (Hall & Rosenthal, 1991; Lipsey, 2003). De plus, plusieurs modérateurs pouvant influencer nos résultats demeurent inconnus. Enfin, il existe souvent des interrelations entre les différents modérateurs étudiés (Lipsey, 2003). Il est toutefois possible, afin de mieux contrôler pour la multicollinéarité potentielle entre les modérateurs, d'avoir recours à des techniques statistiques multivariées, lorsque la taille de l'échantillon le permet. Cependant, l'utilisation de ces techniques est encore très rare dans le domaine de la prévention et de l'intervention.

### **Pondération des études primaires**

Plusieurs auteurs (Hedges & Olkin, 1985; Hunter & Schmidt, 1990, Schmidt & Hunter, 2003; Cook et al., 1992) favorisent l'utilisation de pondérations afin de réduire les disparités initiales au niveau de certaines sources artificielles de variation, telle la taille de l'échantillon ou l'erreur de mesure. En effet, afin que les études soient vraiment comparables entre elles, il est important de les rendre les plus équivalentes possible. C'est ainsi que plusieurs auteurs utilisent le poids de la variance inversée comme méthode d'ajustement (Hedges & Olkin, 1985; Hunter & Schmidt, 1990). Cette technique permet de minimiser l'influence des différences de variabilité (variance) entre les études. Lorsqu'une étude comporte moins de quinze participants, Hunter et Schmidt (1990) proposent un ajustement qui diminue le poids de celle-ci dans l'échantillon total. Plusieurs autres types de pondérations existent, notamment afin de tenir compte de l'erreur de mesure ou des différentes sources d'invalidité interne. Certains de ces ajustements sont toutefois contestés, des analyses de modélisation ayant démontré la présence de certaines lacunes associées aux estimations qui en découlent (Wan den Noorthgate & Onghena, 2003). Dans le domaine de la prévention, le poids de la variance inversée demeure l'ajustement le plus fréquemment utilisé.

### **Présences de plusieurs indices TE pour un même échantillon**

Un enjeu important dans le cadre de la méta-analyse se rapporte à la dépendance possible des indices TE lorsque plusieurs mesures de la variable dépendante sont disponibles au sein d'une même étude ou lorsque plusieurs groupes d'intervention partagent un même groupe de contrôle. Lorsqu'une même étude fournit plusieurs indices TE (en lien avec plusieurs mesures différentes de la variable dépendante), divers choix s'offrent au méta-analyste. Il peut calculer un indice TE moyen ou médian ou encore choisir l'indice associé au type de mesure ou

à la source la plus fréquemment utilisée parmi les autres études de l'échantillon (Durlak, 1995; Lipsey & Wilson, 2001a; Shadish et al., 2002). Rosenthal et Rubin (1986) proposent quant à eux d'inclure dans la méta-analyse un indice TE par construit, source ou type de mesure et de les pondérer ensuite par rapport à la dépendance des mesures ainsi créées. Ce choix implique cependant de tenir compte d'une certaine erreur de mesure supplémentaire. Les méta-analyses présentées dans le tableau 1 n'ont pas choisi cette dernière technique, la majorité d'entre elles, ont préféré utiliser la moyenne des TE se rapportant à un même construit (ex. : Grossman & Hugues, 1992; Farrington & Welsh, 2003 ; Sweet & Appelbaum, 2004).

Dans le cas d'études primaires comprenant plusieurs groupes d'intervention, mais un seul groupe de contrôle, le chercheur peut calculer l'indice TE en contrastant un seul des groupes d'intervention (généralement le groupe à intervention maximale) avec le groupe de contrôle. Il peut aussi faire la moyenne entre les différents groupes d'intervention (ce dernier choix est cependant à éviter) ou encore utiliser les indices TE de chaque groupe d'intervention en pondérant pour la dépendance des groupes (Gleser & Olkin, 1994). Tobler et al. (2000) ont choisi de pondérer pour la dépendance des groupes lorsqu'ils avaient plusieurs groupes d'intervention pour un seul groupe de contrôle alors que Wilson et al. (2001) n'ont utilisé que le groupe d'intervention maximale.

### Degré de signification

La popularité grandissante de la méta-analyse a fait ressurgir un débat ancien à propos de l'utilisation du seuil de signification (valeur  $p$  de .05) comme critère de décision eu égard à la présence ou à l'absence d'une relation entre deux variables (Cohen, 1994; Schmidt & Hunter, 2003). Certains auteurs soutiennent en effet que l'utilisation exclusive de la valeur  $p$  .05, autrement appelée seuil alpha de 5% (associé à l'erreur de type 1), dans la recherche scientifique est devenue une maladie (Olkin, 1990). Selon eux, ce qui était à la base une convention est devenu un dogme qui ne laisse plus de place à la nuance : si notre valeur  $p$  est plus petite que .05, nous pouvons conclure à l'existence d'une relation, sinon la prétention à l'existence d'une relation est rejetée. Outre le caractère dichotomique de l'utilisation du .05, d'autres problèmes sont présents en ce qui concerne les degrés de signification. Premièrement, ces derniers sont influencés par la taille de l'échantillon : des analyses effectuées sur de très grands échantillons augmentent la probabilité d'obtenir une valeur  $p$  significative (Abelson, 1995). Deuxièmement, le degré de signification ne contrôle pas l'erreur de type II, soit le fait d'accepter l'hypothèse nulle alors qu'elle est fautive. Autrement dit, certains chercheurs concluent qu'il n'existe pas de relation entre deux variables alors que cette relation est malgré tout probable (Hedges, 1991). Pour Schmidt et Hunter (2003), une telle interprétation des résultats de recherche pourrait mener à de fausses conclusions. Ils sont d'avis que d'autres procédures doivent être utilisées, comme les indices TE ou les intervalles de confiance qui contiennent toutes les informations fournies par le degré de signification (Cohen, 1994), mais qui, en plus, tiennent compte de l'erreur

de type II. Au niveau des synthèses des connaissances, la méta-analyse, basée sur la TE, comble plusieurs des lacunes soulevées précédemment.

### Test d'homogénéité

Le test d'homogénéité dont il a déjà été question plus haut est utilisé par plusieurs méta-analystes comme le critère permettant de choisir le modèle d'analyse approprié (à effets fixes, à effets aléatoires ou à effets mixtes). Or, Lipsey et Wilson (2001b) rapportent que, pour les méta-analyses de moins de 100 études, le test d'homogénéité manque de puissance statistique. Il ne serait donc pas adéquat. D'autres chercheurs sont plus drastiques (Hedges, 1991; Cohen, 1994; Schmidt & Hunter, 2003) et condamnent l'utilisation du test d'homogénéité puisqu'il est basé sur le seuil de signification ( $.05$ ), pourfendu par ces auteurs (voir section degré de signification). Selon eux, une stricte adhérence aux résultats obtenus sur ce test ne peut en aucun cas être aussi valide que le jugement du chercheur quant à la nature des données constituant la méta-analyse ainsi qu'au type d'inférences qu'il souhaite faire (à l'univers des possibles ou à l'échantillon actuel).

### Modèles d'analyse

Des trois modèles présentés précédemment, celui à effets fixes est le plus fréquemment utilisé, dû en grande partie à sa relative simplicité. Dans le cas des méta-analyses recensées dans le tableau 1, 14 de celles-ci ont fait l'usage de ce modèle, alors que seulement 3 ont utilisé un modèle aléatoire et une seule un modèle mixte. Or, Shadish et al. (2002) mettent en garde les chercheurs quant au choix de ce modèle. Selon eux, l'objectif de la plupart des méta-analyses devrait être de pouvoir généraliser les résultats à l'univers des possibles, impliquant ainsi un modèle à effets aléatoires. Abondant dans le même sens, Durlak (1995) et Schmidt et Hunter (2003) proposent d'utiliser uniquement des modèles à effets aléatoires ou mixtes. Enfin, d'autres auteurs (Hedges & Vevea, 1998; Hedges, 1983; Hedges & Olkin, 1985; Lipsey & Wilson, 2001a) postulent que le choix du modèle doit dépendre en grande partie du type d'inférence désiré. Ainsi, un modèle à effets fixes est intéressant pour émettre des inférences sur les données observées sans les généraliser à une population de référence plus générale (inférences conditionnelles). À l'inverse, un modèle à effets aléatoires doit être utilisé lorsque le chercheur souhaite généraliser ses résultats à une population de référence plus générale (inférences inconditionnelles). La majorité des méta-analystes (ex. : Rispens et al., 1997) dans le domaine de la prévention qui justifient le choix de leur modèle affirment qu'ils souhaitent davantage faire des inférences conditionnelles qu'inconditionnelles. Cela peut expliquer la faible utilisation par ces derniers des modèles mixtes ou aléatoires.

### Considérations pratiques

Une des forces de l'indice TE est sa facilité de transposition en signification pratique (Cooper, 1991; Lipsey & Wilson, 1993; McCartney & Rosenthal, 2000; Prentice & Miller, 1992; Rosnow & Rosenthal, 1988). À cet égard, Rosnow et

Rosenthal (2003) donnent l'exemple d'une étude sur l'efficacité de l'aspirine comme agent préventif des crises cardiaques. Dans cette étude, les chercheurs ont décidé à l'unanimité d'arrêter les essais cliniques après avoir constaté à mi-parcours que l'indice TE associé à l'administration de l'aspirine en regard de la rechute était de .034. Même si un indice TE de .034 est minime selon les critères de Cohen (1977), pour ces patients, la prise d'aspirine diminuait de 3,4% les risques d'avoir une nouvelle crise cardiaque. Au plan clinique, ce résultat est loin d'être négligeable étant donné la nature du problème traité et le faible coût du traitement.

De leur côté, Rosenthal et Rubin (1982) ont développé un index appelé le Binomial Effect Size Display (BESD) afin d'évaluer la signification pratique des indices TE. Le BESD compare deux groupes (par exemple, le groupe expérimental et le groupe de contrôle) selon un taux de succès relatif. Pour ce faire, il suffit de dichotomiser les résultats de la variable dépendante (ex. : réussite ou échec) et de calculer la différence entre le groupe expérimental et le groupe contrôle en terme de succès. Rosenthal (1994) a ainsi calculé le BESD dans le cadre d'une méta-analyse portant sur les effets de la psychothérapie. Il a montré qu'avec un  $r$  modéré de .32, une personne ayant suivi une psychothérapie augmentait de 34% ses chances d'en retirer des effets bénéfiques.

Parmi les méta-analyses recensées dans le tableau 1, la participation au groupe expérimental augmentait le taux de succès entre 7 et 80% comparativement au groupe de comparaison. Cette importance pratique se retrouve parfois même au niveau des interventions produisant des petits indices TE. Par exemple, l'indice TE moyen des interventions visant à prévenir la récidive chez des jeunes contrevenants dans la méta-analyse de Lipsey et Wilson (1998) est de .12. Concrètement, ce résultat se traduit par une réduction de 12% du taux de récidive par rapport au groupe de comparaison. Considérant les coûts humains, financiers et sociaux de la délinquance juvénile, cette donnée suggère que les programmes peuvent certes être grandement améliorés, mais que ce qui se fait actuellement n'est pas complètement inutile.

L'évaluation des répercussions pratiques, à l'aide d'outils tels le BESD, peut également servir de base à une analyse coûts/bénéfices (McCartney & Rosenthal, 2000). En effet, une intervention peu coûteuse, comme l'aspirine pour la prévention des crises cardiaques, peut s'avérer avantageuse même si son indice TE est relativement faible. Dans le domaine de la prévention, une méta-analyse sur les effets des programmes implantés en milieu scolaire ciblant les comportements d'agressivité, révèle que, même si les programmes de démonstration (c.-à-d. implanté par une équipe de chercheurs) présentaient un indice TE modeste de .25, cela se traduirait par une réduction de 7% des batailles (Wilson et al., 2003). Pour une commission scolaire, ce pourcentage signifie, par exemple, une diminution substantielle des coûts associés aux interventions en cas de crise.

## Discussion

La méta-analyse constitue une avancée méthodologique importante dans la façon de synthétiser et d'intégrer les résultats de multiples études sur un thème donné. Par sa relative nouveauté, elle prête flanc à la critique et à la controverse même si elle continue de présenter plusieurs avantages sur les autres méthodes de synthèse des recherches. Voici donc, en guise de conclusion, un survol imagé de certaines limites, mais aussi des principales forces de la méta-analyse.

### Limites

*Un déchet dans un joli sac demeure un déchet.* La qualité de la méta-analyse est fonction de la qualité générale des études primaires qui la composent. Ainsi, une méta-analyse n'incluant que des études comprenant des devis sans groupe de comparaison ne pourrait jamais prétendre à la même robustesse de conclusion qu'une méta-analyse incluant des études primaires aux devis expérimentaux avec répartition aléatoire des sujets.

*Il ne faut pas mélanger des pommes et des oranges (sauf si on veut faire une salade de fruits).* Pour certains (voir Eysenck, 1978), il ne devrait pas être possible de calculer un indice TE moyen, du fait que l'on compare des études aux échantillons, aux mesures et à l'opérationnalisation des variables dépendantes différents. Par exemple, la délinquance est une problématique qui a retenu l'attention de plusieurs méta-analystes. Les construits retrouvés dans les études primaires pour définir cette problématique ainsi que les instruments utilisés pour la mesurer présentent une grande variabilité. Les instruments de mesure peuvent être des rapports de police, des observations, des questionnaires ou des entrevues. Les données peuvent provenir de sources officielles, d'observateurs indépendants, des parents, des enseignants ou des adolescents eux-mêmes. De plus, certains auteurs considèrent la délinquance comme étant des comportements, d'autre des attitudes ou encore une personnalité. La critique d'Eysenck (1978) affirme qu'il est presque impossible de donner un sens commun à ces informations disparates.

*L'association ne fait pas le larron.* Les résultats d'une méta-analyse demeurent toujours corrélationnels (associatifs), ce qui ne permet pas de faire des inférences causales (Cooper et Hedges, 1994). En effet, seuls les devis expérimentaux contrôlant pour toutes les variables confondantes peuvent établir un lien de causalité. Même si la méta-analyse comporte des études primaires de nature expérimentale, la sélection des études, elle, n'est pas soumise au processus de la distribution aléatoire étant donné que le méta-analyste doit composer avec les études qui ont déjà été menées. L'attribution causale est par conséquent impossible.

*Une trouvée pour dix de perdues.* Le nombre d'études faisant partie de l'échantillon d'une méta-analyse est toujours limité et le méta-analyste ne peut jamais avoir accès à toutes les études possibles (voir la section biais de publication). De plus, même si toutes les études sur un sujet pourraient être incluses dans une méta-analyse, cela ne signifie pas pour autant que ces études sont représentatives

de l'univers des possibles quant aux situations, aux traitements ou aux contextes auxquels le chercheur voudrait idéalement généraliser les résultats (Cook et al., 1992).

## Forces

*Est-ce bien construit?* Puisqu'elle se base habituellement sur des opérationnalisations différentes d'un construit, la méta-analyse fournit des indications sur la précision des divers instruments utilisés dans la mesure d'un phénomène. Ainsi, à qualité méthodologique égale, il est possible de vérifier si certaines sources, indicateurs ou méthodes pour mesurer la variable dépendante, dans ce cas-ci les effets du programme, produisent des résultats plus élevés ou précis que d'autres.

*Une question de puissance... statistique.* La méta-analyse augmente la puissance des études individuelles (Hunter & Schmidt, 1990). En effet, elle réduit l'erreur standard associée à la taille de l'effet moyen pondéré en diminuant l'intervalle de confiance autour de celui-ci (Cohn & Becker, 2003). Par ailleurs, l'indice TE tient compte de l'erreur de type II, c'est-à-dire affirmer qu'il n'y a pas de relation entre deux variables alors qu'il y en a une, contrairement aux tests basés sur le degré de signification (Cohen, 1977, 1994).

*Une compétence transversale.* Un des principaux objectifs de la méta-analyse est de pouvoir tirer des conclusions quant à la généralisation des résultats d'effets principaux au-delà des conditions, des traitements ou de l'époque se rapportant aux études primaires sur lesquelles repose la méta-analyse (Campbell, 1972 in Cook et al., 1992). Par l'analyse des modérateurs et par le choix d'un modèle à effets aléatoires ou mixtes, une bonne méta-analyse réussit généralement à bien remplir ce mandat.

*Le contrôle de la qualité.* Une des forces de la méta-analyse par rapport aux autres techniques de synthèse des connaissances repose sur sa capacité de tenir compte de la qualité méthodologique des études qui la composent, soit en introduisant un facteur de pondération, soit en évaluant son impact à l'aide de tests de modération. Cela a pour avantage de court-circuiter le dilemme quant à l'inclusion ou non des études de faible qualité méthodologique, tout en maximisant les sources primaires de données.

## Perspectives futures

Au-delà des différents enjeux et des contributions actuelles de la méta-analyse au corpus des connaissances entourant les interventions préventives et curatives dans plusieurs domaines de l'inadaptation psychosociale, cette dernière est porteuse d'une mission plus grande encore. De par ses qualités, son influence peut s'exercer à plusieurs niveaux.

*Les concepteurs de programmes de prévention.* Sous sa forme actuelle ou future, la méta-analyse contribue à l'identification des « ingrédients actifs » d'un

programme de prévention, à l'examen des facteurs de risque et de protection en cause dans un problème d'adaptation, au repérage des moments d'intervention les plus opportuns, à la reconnaissance des clientèles les plus susceptibles de bénéficier d'une intervention et à la mise à jour des conditions favorables à une mise en œuvre optimale des programmes (Cook et al., 1992; Lipsey & Wilson, 2001b). Bref, la méta-analyse permet de répondre aux questions relatives aux effets principaux des interventions et à celles concernant leurs conditions d'application, que ce soit en terme de caractéristiques des sujets, des intervenants, du milieu, de la qualité de la mise en œuvre ou de la durée du traitement dans une perspective d'intervention différentielle (Kellam & Van Horn, 1997).

*Les décideurs publics.* Les décideurs doivent mettre en place des politiques de santé publique ayant une incidence directe sur la vie des gens, surtout auprès des clientèles les plus vulnérables. Ayant peu de temps pour rassembler et consulter la documentation les renseignant sur les meilleures pratiques possibles, les décideurs doivent se tourner vers des sources leur permettant de se faire une idée rapide de la situation. Dans cette optique, la méta-analyse constitue un outil de choix pour influencer les politiques sociales par sa concision et sa fiabilité.

L'arrimage de la méta-analyse à des études coûts-bénéfices ne pourra que contribuer à augmenter son influence tout en permettant de sélectionner les meilleurs programmes en fonction des disponibilités budgétaires (Grumlich, 1990 in Price, 1997).

*Pour les théoriciens.* L'utilisation de la méta-analyse comme outil de validation de modèles théoriques d'intervention ou de prédiction, quoiqu'encore peu fréquente, demeure une voie d'avenir. Dès lors, il serait important de capitaliser davantage sur les forces de la méta-analyse eu égard à sa possible contribution aux théories de l'inadaptation, particulièrement celles ancrées dans le paradigme de la psychopathologie développementale, actuellement dominant dans le domaine de la prévention et de l'intervention. En ce qui concerne les méta-analyses sur les effets des programmes de prévention et d'intervention, cet objectif pourrait être atteint si davantage d'études primaires inclue l'examen de médiateurs susceptibles d'expliquer les effets des programmes qu'elles ont comme mission d'évaluer. Cela permettrait ensuite d'utiliser les indices TE associés à ces mesures comme médiateurs dans les méta-analyses.

Au niveau de l'intervention, Lipsey (1997) propose d'utiliser une théorie de l'intervention psychosociale non seulement pour développer des interventions, mais aussi pour conceptualiser des méta-analyses. Il explique que les méta-analystes « bricolent » souvent des explications théoriques post-hoc, alors qu'elles devraient en constituer le cadre d'analyse. Pour Lipsey (1997), une théorie de l'intervention devrait contenir au moins les trois éléments suivants : un paradigme de changement, les types de problématiques visées et les types de programmes qui ont été développés. À notre connaissance, seuls Durlak et Wells (1997) ont réalisé une méta-analyse se basant sur les paramètres proposés par Lipsey (1997).

*Pour les chercheurs et les méta-analystes.* Un défi pour les futures méta-analyses réside dans la complexification théorique, méthodologique et statistique des études primaires. Il faudra développer les méthodes afin de synthétiser des données multivariées, tels les bêtas de régression, les corrélations partielles, etc. Quoique certaines techniques existent déjà, elles devront être raffinées davantage. Par ailleurs, des modèles statistiques plus complexes permettant de mieux tenir compte des divers paramètres influençant les résultats devront être utilisés, comme les analyses multi-niveaux (Wan den Noortgate & Onghena, 2003), les modèles d'équations structurales ou les analyses de trajectoires (Cook et al., 1992; Becker & Schram, 1994, Lipsey & Wilson, 2001a). Enfin, nous proposons, dans la lignée d'autres auteurs (voir entre autres Abelson, 1995; Lipsey & Wilson, 2001a; Shadish et al., 2002) que la présentation des indices TE devienne obligatoire lors de la publication d'une étude évaluative.

### Conclusion

Malgré tous les avantages qui y sont associés, la méta-analyse n'est pas une panacée ou une réponse miracle. C'est pourquoi, en aucun cas, elle ne peut se substituer à de bonnes études primaires (Cooper & Hedges, 1994). Répondant à des besoins différents, ces deux types d'études sont complémentaires, l'une et l'autre s'alimentant mutuellement. En effet, les conclusions issues des méta-analyses peuvent guider la conception des études primaires et celles-ci vont en retour permettre la réalisation de nouvelles méta-analyses plus raffinées et plus complètes.

Somme toute, la méta-analyse est une méthode efficace de synthèse des connaissances lorsqu'elle est utilisée adéquatement. Elle permet de tirer des conclusions solides quant à la direction et à la force des effets observés. Cependant, il ne faut pas oublier que la méta-analyse est une technique statistique et que sa bonne utilisation dépend non seulement du méta-analyste responsable de sa réalisation, mais aussi de l'esprit critique des chercheurs et praticiens qui en consomment les résultats et les conclusions. À l'instar de Shadish et al. (2002), nous croyons qu'il importe de rester critique face à une méta-analyse, comme nous le sommes pour n'importe quelle méthode scientifique. Ce texte visait précisément à connaître mieux cet outil dont le potentiel est élevé pour la psychoéducation comme pour plusieurs autres disciplines.



## Références

- Abelson, R.P. (1995). *Statistics as Principled Argument*. Hillsdale, NJ: Lawrence Erlbaum.
- Ang, R.P., & Hughes, J.N. (2002). Differential benefits of skills training with antisocial youth based on group composition: A meta-analytic investigation. *School Psychology Review, 31*, 164-185.
- Bakermans-Kranenburg, M.J., van Ijzendoorn, M.H., & Juffer, F. (2003). Less is more: Meta-analysis of sensitivity and attachment interventions in early childhood. *Psychological Bulletin, 129*, 195-215.
- Banger-Drowns, R.L., Wells-Parker, E., & Chevillard, I. (1997). Assessing the methodological quality of research in narrative reviews and meta-analyses. In K.J. Briant, M. Windle & S.G. West (Eds.), *The science of prevention: Methodological advances for alcohol and substance abuse research* (pp. 405-429). Washington, DC: American Psychological Association.
- Baron, R.M., & Kenny, D.A. (1986). The moderator mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.
- Becker, B.J. (1992). Model of science achievement: Forces affecting male and female performance in school science. In T.D. Cook, H. Cooper, D.S. Cordray et al. (Eds.), *Meta-analysis for explanation* (pp. 209-282). New-York: Russell Sage Foundation.
- Becker, B.J. (1994). Combining significance levels. In H. Cooper & L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 215-230). New-York: Russell Sage Foundation.
- Becker, B.J., & Schram, C.M. (1994). Examining explanatory models through research synthesis. In H. Cooper & L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 357-382). New-York: Russell Sage Foundation.
- Begg, C.B. (1994). Publication bias. In H. Cooper & L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 399-410). New-York: Russell Sage Foundation.
- Bushman, B.J. (1994). Vote-counting procedures in meta-analysis. In H. Cooper & L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 193-214). New-York: Russell Sage Foundation.
- Chalmers, I., Hegdes, L.V., & Cooper, H. (2002). A brief history of research synthesis. *Evaluation & the Health Professions, 25*, 12-37.
- Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences (Revised Edition)*. New-York: Academic Press.
- Cohen, J. (1994). The earth is round ( $p < .05$ ). *American Psychologist, 49*, 997-1003.
- Cohn, L.D., & Becker, B.J. (2003). How meta-analysis increases statistical power. *Psychological Methods, 8*, 243-253.
- Coie, J.D., Watt, N.F., West, S.G., Hawkins, J.D., Asarnow, J.R., Markman, H.J., Ramey, S.L., Shure, M.B., & Long, B. (1993). The science of prevention. *American Psychologist, 48*, 1013-1022.
- Cook, T.D. (1991). Meta-analysis: It's potential for causal description and causal explanation within program evaluation. G. Albrecht & H-U. Otto (Eds.), *Social prevention and the social sciences* (pp. 245-286). Berlin: Walter de Gruyter.
- Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand-McNally.
- Cook, T.D., Cooper, H., Cordray, D.S., Hartmann, H., Hedges, L.V., Light, R.J., Louis, T.A., & Mosteller, F. (1992). *Meta-analysis for explanation*. New-York: Russell Sage Foundation.
- Cooper, H. (1991). An introduction to meta-analysis in the integrative research review. In G. Albrecht & H-U. Otto (Eds.), *Social prevention and the social sciences* (pp. 287-304). Berlin: Walter de Gruyter.
- Cooper, H., & Hedges, L.V. (1994). Potential and limitations of research synthesis. In H. Cooper & L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 521-530). New-York: Russell Sage Foundation.

- Cooper, H., & Hedges L.V. (1994). *The handbook of research synthesis*. New-York: Russell Sage Foundation.
- Corcoran, J., O'Dell Miller, P., & Bultman, L. (1997). Effectiveness of prevention programs for adolescent pregnancy: A meta-analysis. *Journal of Marriage and the Family*, 59, 551-567.
- Dew, M.A., Bromet, E.J., & Brent, D. (1987). A quantitative literature review of the effectiveness of suicide prevention centers. *Journal of Consulting and Clinical Psychology*, 55, 239-244.
- Dishion, T.J., McCord, J., & Poulin, F. (1999). When intervention harm. *American Psychologist*, 54, 755-764.
- DuPaul, G.J., & Eckert, T.L. (1997). The effects of school-based interventions for attention deficit hyperactivity disorder: A meta-analysis. *School Psychology Review*, 26, 5-27.
- Durlak, J.A. (1995). Understanding Meta-Analysis. In L.G. Grimm & P.R. Yarnold (Eds.), *Reading and Understanding Multivariate Statistics* (pp. 319-353). Washington, DC: American Psychological Association.
- Durlak, J.A., & Wells, A.M. (1997). Primary prevention mental health programs for children and adolescents: A meta-analytic review. *American Journal of Community Psychology*, 25, 115-152.
- Eysenck, H.J. (1978). An exercise of mega-silliness. *American Psychology*, 33, 517.
- Farrington, D.P., & Welsh, B.C. (2003). Family-based prevention of offending: A meta-analysis. *Australian and New Zealand Journal of Criminology*, 36, 127-151.
- Field, A.P. (2003). Can meta-analysis be trusted? *Psychologist*, 16, 642-645.
- Furukawa, T.A., Cipriani, A., Barbui, C., Brambilla, P., & Watanabe, N. (2005). Imputing response rates from means and standard deviations in meta-analysis. *International Clinical Psychopharmacology*, 20, 49-52.
- Glass, G.V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher*, 5, 3-8.
- Glesser, L.J., & Olkin, I. (1994). Stochastically dependant effect sizes. In H. Cooper & L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 339-356). New-York: Russell Sage Foundation.
- Greenhouse, J.B., & Iyengar, S. (1994). Sensitivity analysis and diagnostics. In H. Cooper & L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 383-398). New-York: Russell Sage Foundation.
- Grossman, P.B., & Hughes, J.N. (1992). Self-control intervention with internalizing disorders: A review and analysis. *School Psychology Review*, 21, 229-245.
- Hall, J.A., & Rosenthal, R. (1991). Testing for moderator variables in meta-analysis: Issues and methods. *Communication Monographs*, 58, 437-448.
- Hedges, L.V. (1983). A random-effects model for effect sizes. *Psychological Bulletin*, 2, 388-395.
- Hedges, L.V. (1991). Methodological aspects of synthesis of social prevention research. In G. Albrecht & H-U. Otto (Eds.), *Social prevention and the social sciences* (pp. 353-380). Berlin: Walter de Gruyter.
- Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- Hedges, L.V., & Pigott, T.D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods*, 9, 426-445.
- Hedges, L.V., & Vevea, J.L. (1998). Fixed- and random-effect models in meta-analysis. *Psychological Methods*, 3, 486-504.
- Hunter, J.E., & Schmidt, F.L. (1990). *Methods of Meta-Analysis : Correcting Error and Bias in Research Findings*. Newbury Park, CA: SAGE.
- Kellam, S.G., & Van Horn, Y.V. (1997). Life course development, community epidemiology, and preventive trials: A scientific structure for prevention research. *American Journal of Community Psychology*, 25, 177-188.
- Lewis, R.J., & Vosburgh, W.T. (1988). Effectiveness of kindergarten intervention programs. *School Psychology International*, 9, 265-275.
- Lipsey, M.W. (1995). What do we learn from 400 research studies on the effectiveness

- of treatment with juvenile delinquents? In J. McGuire (Ed.), *What works: Reducing reoffending - Guidelines from research and practice* (pp. 63-78). New-York: John Wiley and Sons.
- Lipsey, M.W. (1997). What can you build with thousands of bricks? Musings on the cumulation of knowledge in program evaluation. *New Directions for Evaluation*, 76, 7-23.
- Lipsey, M.W. (2002). Meta-analysis and program outcome evaluation. *Socialvetenskaplig Tidskrift*, 9, 194-208.
- Lipsey, M.W. (2003). Those confounds moderators in meta-analysis: Good, bad, and ugly. *Annals, AAPSS*, 598, 69-81.
- Lipsey, M.W., & Wilson, D.B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48, 1181-1209.
- Lipsey, M.W., & Wilson, D.B. (1998). Effective intervention for serious juvenile offenders: A synthesis of research. In R. Loeber & D.P. Farrington (Eds.), *Serious and violent juvenile offenders: Risk factors and successful interventions* (pp. 313-345). Thousand Oaks, CA: SAGE.
- Lipsey, M.W., & Wilson, D.B. (2001a). *Practical meta-analysis*. Thousand Oaks: SAGE Publications.
- Lipsey, M.W., & Wilson, D.B. (2001b). The way in which intervention studies have personality and why it is important to meta-analysis. *Evaluation & the Health Professions*, 24, 236-254.
- Lochman, J.E. (2001). Issues in prevention with school-aged children: Ongoing intervention refinement, developmental theory, prediction and moderation, and implementation and dissemination. *Prevention and Treatment*, 4, 1-7.
- Mann, C.C. (1990). Meta-analysis in the breach. *Science*, 249, 476-480.
- Mann, C.C. (1994). Can meta-analysis make policy? *Science*, 266, 960-962.
- McCartney, K., & Rosenthal, R. (2000). Effect size, practical importance and social policy for children. *Child Development*, 71, 173-180.
- Olkin, I. (1990). History and goals. In K.W. Wachter & M.L. Straf (Eds.), *The future of meta-analysis* (pp. 3-10). New-York: Russell Sage Foundation.
- Prentice, D.A., & Miller, D.T. (1992). When small effects are impressive. *Psychological Bulletin*, 112, 160-164.
- Price, R.H. (1997). In praise of a cumulative prevention science. *American Journal of Community Psychology*, 25, 169-176.
- Quinn, M.M., Kavale, K.A., Mathur, S.R., Rutheford, R.B., & Forness, S.R. (1999). A meta-analysis of social skill interventions for students with emotional or behavioural disorders. *Journal of Emotional and Behavioral Disorders*, 7, 54-64.
- Raudenbush, S.W. (1994). Random-effects models. In H.Cooper & L.V. Hedges (Eds.), *Handbook of research synthesis* (pp. 301-322). New-York: Russell Sage Foundation.
- Reed, J.G., & Baxter, P.M. (1994). Using reference databases. In H.Cooper & L.V. Hedges (Eds.), *The handbook of research synthesis* (pp.57-70). New-York: Russell Sage Foundation.
- Ripens, J., Aleman, A., & Goudena, P.P. (1997). Prevention of child sexual abuse victimization: A meta-analysis of school programs. *Child Abuse & Neglect*, 21, 975-987.
- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, 86, 638-641.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: SAGE Publications.
- Rosenthal, R. (1994). Parametric measures of effect size. In H.Cooper & L.V. Hedges (Eds.), *Handbook of research synthesis* (pp. 231-244). New-York: Russell Sage Foundation.
- Rosenthal, R. (1995). Writing meta-analytic review. *Psychological Bulletin*, 118, 183-192.
- Rosenthal, R. (1996). Meta-analysis: Concepts, corollaries and controversies. In Adair, J.G, Bélanger, D. & Dion, K.L. (Eds.), *Advances in psychological science: Social, personal and cultural aspect* (pp. 371-383). Actes du XXVI Congrès international de psychologie.

- Rosenthal, R., & Rubin, D.B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, 74, 166-169.
- Rosenthal, R., & Rubin, D.B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin*, 99, 400-406.
- Rosnow, R.L., & Rosenthal, R. (1988). Focused tests of significance and effect size estimation in counseling psychology. *Journal of Counseling Psychology*, 35, 203-208.
- Rosnow, R.L., & Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology*, 57, 221-237.
- Schmidt, F.L., & Hunter, J.E. (2003). Meta-analysis. In J.A. Schinka & W.F. Velicer (Eds.), *Handbook of psychology* (pp. 533-554). New-York: John Wiley and Sons.
- Serketich, W.J., & Dumas, J.E. (1996). The effectiveness of behavioural parent training to modify antisocial behavior in children: a meta-analysis. *Behavior Therapy*, 27, 171-186.
- Shadish, W.R. (1992). Do family and marital psychotherapies change what people do? A meta-analysis of behavioral outcomes. In T.D. Cook, H. Cooper, D.S. Cordray et al. (Eds.), *Meta-analysis for explanation* (pp. 129-208). New-York: Russell Sage Foundation.
- Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalised causal inference*. Boston: Houghton Mifflin Company.
- Smith, M.L., & Glass, G.V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752-760.
- Sweet, M.A., & Appelbaum, M.I. (2004). Is home visiting an effective strategy? A meta-analytic review of home visiting programs for families with young children. *Child Development*, 75, 1435-1456.
- Tobler, N.S., Roona, M.R., Ochshorn, P., Marshall, D.G., Streke, A.V., & Stackpole, K.M. (2000). School-based adolescent drug prevention programs: 1998 meta-analysis. *The Journal of Primary Prevention*, 20, 275-336.
- Van Gijsegem, H. (1999). *Us et abus de la mise en mots en matière d'abus sexuels*. Montréal: Méridien.
- Wan den Noortgate, W., & Onghena, P. (2003). Multilevel meta-analysis: a comparison with traditional meta-analytic procedures. *Educational and Psychological Measurement*, 63, 765-790.
- Weissberg, R.P., Kumpfer, K.L., & Seligman, M.E.P. (2003). Prevention that works for children and youth: an introduction. *American Psychologist*, 58, 425-432.
- Weisz, J.R., Weiss, B., Han, S.S., Granger, D.A., & Morton, T. (1995). Effects of psychotherapy with children and adolescent revisited: a meta-analysis of treatment outcome studies. *Psychological Bulletin*, 117, 450-468.
- Wilson, D.B., Gottfredson, D.C., & Najaka, S.S. (2001). School-based prevention of problem behaviors: a meta-analysis. *Journal of Quantitative Criminology*, 17, 247-272.
- Wilson, S.J., & Lipsey, M.W. (2000). Wilderness challenge programs for delinquent youth: A meta-analysis of outcome evaluations. *Evaluation and Program Planning*, 23, 1-12.
- Wilson, S.J., Lipsey, M.W., & Derzon, J.H. (2003). The effects of school-based intervention programs on aggressive behavior: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 71, 136-149.
- Wortman, P.M. (1994). Judging research quality. In H.Cooper & L.V.Hedges (Eds.), *Handbook of research synthesis* (pp. 97-110). New-York: Russell Sage Foundation.