

Robert R. Haccoun¹

Université de Montréal

La méta-analyse Principes et concepts

Résumé

Cet article décrit les principaux concepts et principes de la méta-analyse, une technique quantitative servant à faire les recensements de la documentation empirique. L'article présente une technique qui domine dans ce champ disciplinaire; celle de Hunter et Schmidt (1990). Les indicateurs qui doivent être calculés aussi bien que ceux qui servent à interpréter les résultats sont présentés et discutés.

Mots-Clefs: Méta-analyse, tendance centrale, taille de l'effet, artefacts, modérateurs, atténuation, variances restreintes, erreur d'échantillonnage, taille de l'échantillon

Meta-analysis Concepts and principles

Abstract

Meta analysis is fast taking a dominant role in the conduct of literature reviews. This article describes the basic principles which underlie the technique and then describes in further detail one of the most powerful of meta analytic techniques: the one proposed by Hunter and Schmidt (1990). Key issues relevant to the conduct, reporting and to the interpretation of meta analytic results are presented and discussed.

Key words: Meta-analysis, central tendency, effect size, artifacts, moderators, attenuation, variance restriction, sampling error, sample size

¹ Les tirés à part sont disponibles en écrivant à l'auteur : Département de psychologie, Université de Montréal, C.P. 6128, succursale Centre-Ville, Montréal, Québec, Canada H3C 3J7.

L'expérience démocratique facilite-t-elle l'harmonie sociale? Est-ce que suivre un processus thérapeutique contribue à vivre plus heureux? Une approche dynamique est-elle préférable à une approche comportementale? Les personnes satisfaites au travail sont-elles plus productives? La personnalité des individus est-elle un déterminant du succès professionnel? La médecine douce est-elle efficace?

Les recensions de documentation existent, pour faire le point sur ce genre de questions, phénomènes ou concepts dans le but principal de dégager les pistes d'avenir.

Chaque année des milliers d'études empiriques sont publiés en sociologie, en psychologie, en pharmacologie, etc. Si cela témoigne de la richesse et de la vigueur de la recherche dans tous les champs disciplinaires force est de reconnaître combien il devient difficile de se faire une idée intégrée de l'état des connaissances.

Heureusement, certains collègues dévoués s'attellent à cette tâche. Ces valeureux collègues identifient et font le sommaire de chaque article, afin d'intégrer des conclusions et d'en extraire une vue d'ensemble. Si chaque étude représente un arbre, nos collègues recenseurs nous aident à voir la forêt.

C'est un travail de moins par lequel nous sommes presque tous passés. Et, fort de cette expérience, les étudiants, les professeurs, les praticiens et les chercheurs sont tous reconnaissants lorsque quelqu'un d'autre s'en charge! Il serait tellement plus agréable si ce pénible processus pouvait être simplifié.

La tâche de recension est certes difficile et laborieuse, impliquant des mois de travail consacré à l'identification et à l'intégration de résultats de recherches qui, tout en portant sur le même phénomène, ne se servent pas exactement du même protocole de recherche, ni des mêmes instruments de mesure et qui ne font pas toujours porter leur étude sur les mêmes populations. Par conséquent les résultats des différentes études portant sur la même problématique montrent quasi invariablement des incohérences, voir des contradictions. Si certaines études démontrent que le traitement X est supérieur au traitement Y, il est certain que d'autres études ne démontrent aucune différence entre les traitements et que d'autres « prouvent » le résultat inverse.

La réplication scientifique est d'une grande valeur et est fortement encouragée dans les écrits épistémologiques. Mais la réalité demeure que les répliques sont rares. Un chercheur qui se propose de faire exactement ce qui a déjà été fait aurait de la difficulté à trouver du financement ou de voir son projet de thèse accepté! Les différences entre les résultats d'études sont-elles donc tributaires de ces variations entre les recherches ou pointent-elles l'instabilité inhérente au phénomène à l'étude?

Pourtant, en psychologie, comme dans toutes les autres disciplines empiriques, les recensions de documents se publient à chaque année et chacune en

tire des conclusions. Comment est-ce possible? Faisons un survol du processus de recension d'une documentation.

L'approche traditionnelle du processus de recension de la documentation

Le « recenseur-moine » identifie la documentation pertinente par l'entremise des banques de données informatisées, tel que PsychLit ou Psy Abstract. Des mots clés sont utilisés pour réaliser cette recherche préliminaire. Souvent plusieurs essais sont requis avant de pouvoir cerner un inventaire suffisamment précis des articles pertinents. Ensuite, le chercheur consulte chaque article afin d'en extraire les conclusions. Puisque la somme d'information à travers les articles disponibles est invariablement immense et disparate, le chercheur simplifie l'information. Afin de l'aider dans ce processus, des fiches sommaires décrivant chaque écrit sont préparées : typiquement, le chercheur note le résultat en transcrivant la statistique principale de l'étude : le test de signification statistique. L'étude A démontre que le groupe qui reçoit le traitement est « significativement » plus fort sur la mesure dépendante; l'étude B n'indique aucune différence etc. Il est facile de s'imaginer le néophyte, confronté à cette cacophonie de résultats, se tenant la tête avec une main et décapant sa boîte d'aspirine avec l'autre!

Mais, le chercheur qui n'en est pas à ses premières armes, ne se décourage pas. Il comprend la nécessité de se faire une idée (ou conclusion) préliminaire qui définit la tendance dominante parmi les résultats des études. Généralement, il identifiera une tendance lourde. Si un nombre important d'études rapporte des résultats « significatifs », il tirera la conclusion que l'effet existe. Sinon la conclusion inverse prédominera.

Ensuite, les conclusions sont nuancées en y intégrant les résultats divergeants. Combien de fois avons-nous lu un recensement des écrits qui concluent que « la documentation démontre qu'en général X et Y sont en relation ou que le traitement X est supérieur au traitement Y, mais qu'il existe des résultats divergents et qu'il faudra entreprendre d'autres recherches pour tirer des conclusions définitives » ?

Mais ce n'est pas tout! Il faut aussi expliquer les résultats divergents. Prenons l'exemple simplifié où l'étude A détecte une différence entre les traitements X et Y tandis que l'étude B montre aucune différence. Comment traiter ce constat dans une recension?

L'une ou l'autre de deux explications est invoquée par les chercheurs. La première explication porte sur la substance : le traitement X est supérieur au traitement Y dans la situation (ou avec la population) de l'étude A mais pas dans la situation B. Ce qui mène à la conclusion que le phénomène à l'étude se comporte de différentes façons dans différentes situations. Par exemple, la thérapie comportementale est supérieure à la thérapie dynamique dans les cas où le problème est très précis et associé à un trouble comportementale particulier, tel que

l'agoraphobie. Lorsque le problème est plus diffus une approche plus existentielle est préférée. Est-ce que le thérapeute devrait faire usage d'une intervention comportementale? Cela dépend du problème.

Lorsqu'un phénomène ne se produit que dans un contexte, qu'avec une population distincte ou ne se révèle que lorsque certains instruments de mesures sont utilisés, nous disons que cette incohérence dans les résultats est attribuable à une *variable modératrice*. Ainsi, le type de problème (agoraphobie versus trouble existentiel) est une variable modératrice. Les variables modératrices servent à fournir une *explication substantielle* des résultats.

La deuxième explication pour les divergences entre les études est d'ordre méthodologique. Ainsi l'étude A (ou l'étude B) comporte des choix (ou des erreurs) de méthodologies qui expliquent la différence entre leurs résultats. Un exemple très fréquent et très simple provient du nombre de sujets utilisés dans les études. Lorsque le N est faible les chances de trouver des différences statistiquement significatives sont plus faibles que lorsque le nombre de sujet est plus important. Par conséquent, nous pouvons parfois dire que certaines études ne montrent pas la même tendance que les autres parce qu'elles se servent d'un nombre insuffisant de sujet.

Dépendant de l'explication que le recenseur choisit (substantielle ou méthodologique) deux conclusions dramatiquement différentes sont possibles. Dans le premier cas, celui de l'explication substantielle, la conclusion sera que le traitement X n'est applicable que dans certaines situations., tandis que dans le deuxième scénario, celui de l'explication méthodologique, le chercheur tirera la conclusion que le *vrai* résultat est que le traitement X est supérieur au traitement Y et que les autres résultats auraient été dans le même sens si les auteurs avaient évité l'erreur en question.

Sommaire des étapes dans un recensement de documentation

Ainsi le recensement traditionnel des écrits comprend les étapes suivantes :

1. Identification de la documentation.
2. Analyse de chaque résultat et identification des tendances dominantes.
3. Identification des variations ou incohérences dans les résultats des études.
4. Attribution des incohérences à des erreurs méthodologiques ou à des facteurs substantiels.
5. Une liste de recommandations pour les recherches futures.

Critique de la procédure traditionnelle de recensement des écrits et principes méta analytiques

Une fois la documentation identifiée, le recensement des écrits exige l'identification de la tendance globale aussi bien que l'identification et l'explication des divergences. Simple en apparence, cette démarche est remplie d'embûches, très sujette aux biais du recenseur et très susceptible à l'erreur.

Le nombre de sujets. Comment définir la tendance centrale parmi un ensemble d'étude? La façon la plus simple et la plus répandue consiste à comparer la fréquence des études qui produisent des résultats statistiquement significatifs et celle des autres puis de tirer les conclusions quant aux résultats les plus fréquents ou typiques. Cependant, le problème est que le test de signification statistique résulte de deux éléments qui sont indissociables : (a) la taille de la différence (ou de la relation) et (b) le nombre d'observation. Un même niveau de différence obtenue dans deux études, une comprenant beaucoup de sujets, l'autre peu, a de fortes chances d'être « significatif » dans le premier cas mais non significatif dans le second. Par conséquent, le recenseur traditionnel pourrait conclure de façon erronée, qu'il existe des différences de substance entre ces deux études alors qu'en réalité, il ne s'agit que d'une différence du nombre de sujets!

Le mode versus la moyenne. Le deuxième problème provient de paramètres moins évidents mais tout aussi importants. Une des choses les plus importantes à comprendre en recherche est que le résultat obtenu dans chaque étude représente une *estimation* du résultat véritable qui existe. Par exemple, si je trouve une corrélation entre l'anxiété et la performance à une tâche, la taille de cette corrélation n'est qu'un estimé de la véritable corrélation qui existe dans la nature entre ces deux paramètres. Une deuxième étude produira des résultats qui seront eux aussi un estimé de cette réalité. Les probabilités sont fortes que ces deux corrélations ne seront pas les mêmes. En fait, si nous faisons une infinité d'études sur la même relation et que nous placions chaque résultat sur un graphique, les résultats se répartiraient sur une courbe normale. Chaque étude que nous examinons lors d'un recensement représente donc un point de ce graphique. La *moyenne* de l'ensemble de ces divers points représente le meilleur estimé que nous avons de la réalité. Le recenseur traditionnel qui tire sa conclusion à partir du résultat le plus typique (ou fréquent) le fera à partir du *mode* (valeur typique) de la distribution de résultat et non pas en fonction de sa moyenne. Or, le mode est un estimé de moins bonne qualité que la moyenne. Par conséquent, le chercheur narratif a plus de chances de produire une conclusion erronée.

L'erreur d'échantillonnage. De plus, le chercheur narratif se doit de faire une interprétation des divergences entre les études. Or, une partie importante de cette divergence provient de l'aléa. Souvenons-nous que chaque étude ne représente qu'un estimé de la population de tous les résultats possibles. Par conséquent,

lorsque l'analyste traditionnel identifie les résultats divergents, il est probable qu'au moins une partie de ces différences est attribuable à des facteurs aléatoires qui ne nous renseignent pas sur la véritable divergence entre les études. Les statisticiens nomment ceci l'erreur d'échantillonnage.

L'erreur de la mesure. Toutes les mesures, et en particulier les mesures psychologiques, sont sujettes à l'erreur. En pratique cela veut dire qu'au moins une partie de la réponse qu'une personne nous donne, par exemple à un test d'intelligence ou de personnalité, est erronée. L'erreur de mesure est fortement influencée par la fidélité des mesures. La *fidélité* est un indice du degré avec lequel la réponse fournie par une personne est le fruit de l'aléa versus la conséquence de quelque chose de stable chez l'individu. Prenons le cas de la stabilité des mesures (« test-retest »). La réponse que vous donneriez à une mesure de satisfaction au travail aujourd'hui sera au moins un peu différente de celle que vous fournirez un mois plus tard, même lorsque le véritable niveau de satisfaction est invariant. Donc, la réponse de chaque personne est en partie tributaire de sa satisfaction, le « vrai » sentiment de satisfaction que vous éprouvez, et en partie le résultat d'événements ponctuels et aléatoires.

Une mesure psychologique plus sensible à ces événements aléatoires, par exemple un questionnaire qui ne comprend qu'un petit nombre de questions, produira des résultats moins fidèles que d'autres. Maintenant, les études qui sont incluses dans un recensement ne se servent pas toutes des mêmes mesures. Certaines études font appel à des mesures de meilleures qualités que d'autres. Celles qui se servent de mesures moins fidèles produiront toujours des résultats moins probants que les études qui font appel à une meilleure instrumentation, c'est à dire, moins entachée d'erreurs. Par conséquent, la différence que nous observons entre deux études n'est peut-être que le reflet de ce choix d'instruments. Si ces deux études avaient fait appel à une instrumentation de qualité égale, les résultats obtenus auraient pu être le même. Le recenseur traditionnel qui n'aurait pas prêté attention à ce problème aurait tiré la fausse conclusion que les résultats des deux études sont divergents.

La restriction de la variance. Supposons que j'étudie la relation entre le niveau socio-économique des écoliers et leur succès scolaire : l'étude A montre une relation importante entre ces deux variables alors que l'étude B n'en montre pas. Clairement, il y a une contradiction. Supposons maintenant que l'on vous dit que l'étude B a été exécutée dans une commission scolaire qui regroupe des quartiers très riches aussi bien que des quartiers très pauvres, tandis que l'étude A a été exécutée exclusivement avec des écoliers provenant de milieu défavorisé. De fait, une relation peut être détectée seulement lorsque les variables à l'étude démontrent de la variabilité. La variabilité du niveau socio-économique de l'échantillon B est très restreinte. Par conséquent, la corrélation observée avec l'échantillon B sera « restreinte » et plus petite que celle détectée dans l'étude A. Le recenseur narratif

qui aurait conclu que la relation entre statut socio-économique et réussite scolaire est incohérente aurait fait une erreur. La différence observée est causée par un problème lié à l'échantillon de recherche choisie et non pas à une incohérence au sein du phénomène à l'étude.

Le processus méta-analytique²

Le recensement méta-analytique débute de la même manière que sa contrepartie traditionnelle : par une identification des études pertinentes pour le champ d'études en question. Naturellement, seules les études empiriques et quantitatives peuvent être analysées avec la procédure.

Préparation des données.

Chaque étude est examinée afin d'en extraire le résultat d'intérêt. Chaque étude méta-analysable doit fournir les moyennes obtenues dans chacun des groupes, leurs écarts-type et le nombre de sujets. Souvent, les auteurs ne fournissent pas ces données. Mais cela n'est pas une limite importante à condition de connaître le test statistique (t , F , r , etc.) qui forme la base des conclusions. Ainsi, les statistiques de différences entre les groupes (t , F ou le chi deux) ou les indices de relations, tel que le coefficient de corrélation (r), sont suffisantes. La recension méta analytique ne fait aucune utilisation des indices de significations statistiques ($p < .05$, $p < .01$, etc.) et il n'est pas utile d'en tenir compte. Pour chaque étude, nous aurons alors le résultat principal exprimé par une statistique sommaire.

La métrique commune

Le résultat obtenu dans chaque étude doit maintenant être exprimé sous une forme commune. Toutes les statistiques (r , t , etc.) sont mathématiquement équivalentes et peuvent être traduites d'une structure à une autre. Par exemple, il est facile de transformer un résultat exprimé suivant la statistique de différence d'étudiants (le t) en F ou en chi-deux (χ^2) ou en coefficient de corrélation. Les formules de transformations sont disponibles en consultant Hunter et Schmidt (1990).

En général, ma préférence est de les traduire toutes en coefficient de corrélation. Je préfère cette métrique commune *pour faciliter l'interprétation*, puisque cette corrélation au carré (r^2) représente la proportion de variance expliquée par le phénomène à l'étude. Par exemple, une corrélation de 0,50 qui existerait entre le niveau de stimulation intellectuelle et le QI des enfants indiquerait que le QI pourrait être augmenté de 25% (0,50²). Mais il n'est pas nécessaire de suivre mon exemple ici. Toutes les autres métriques produiront les

² L'objectif du texte vise l'explication plutôt que l'enseignement de la méta-analyse, aussi certaines simplifications lui sont imposées.

mêmes conclusions bien que certaines soient plus difficiles à interpréter que les autres.

Détermination de la tendance globale.

La détermination de la tendance globale se fait en calculant une moyenne. Puisque les résultats obtenus pour les études sont toutes sur la même métrique, il devient facile de calculer la tendance centrale qui se dégage de la distribution des résultats. Bien que nous puissions faire appel au mode ou à la médiane, la moyenne arithmétique est le meilleur indicateur. Ainsi, la moyenne est calculée de la manière habituelle : $(\sum X) / K$ où $\sum X$ représente la somme des résultats obtenus dans chaque étude et où K représente le nombre d'études.

Mais avant de construire la moyenne, il faut d'abord solutionner un problème. Il est clair que toutes les études ne se servent pas du même nombre de sujets. Ainsi, nous nous devons d'avoir plus confiance dans les études qui se servent de vastes échantillons comparativement à celles qui ne se servent que de quelques sujets³. Il faudrait rendre compte de cette réalité. Pour ce faire, nous n'avons qu'à pondérer les résultats par le nombre de sujets sur lequel ils sont construits. En pratique, cela se fait facilement en multipliant chaque valeur obtenue par son nombre de sujets, en faisant la somme pour les études, puis en divisant par le nombre total de sujets à travers les études. Ainsi, le calcul de la moyenne méta-analytique devient : $(\sum nX) / nK$

La moyenne méta analytique est la valeur la plus importante pour une méta analyse. Elle indique la taille de la relation qui existe dans la population au sujet du phénomène à l'étude.

Le problème de la variance entre les études

Mais la moyenne est un indicateur omnibus qui ne comprend pas toute l'information requise pour faire le point sur un champ disciplinaire ou sur un phénomène. Il faut aussi étudier la divergence entre les résultats. La divergence implique une différence entre les résultats. Une façon pratique de concevoir ces différences consiste à les définir comme une variance. C'est à dire, chaque résultat obtenu fait partie d'une distribution de résultats, laquelle distribution est décrite partiellement par la moyenne (déjà calculée). Nous pouvons donc calculer cette variance en faisant appel à la formule traditionnelle de la variance, soit $\sum (X-M)^2 / K-1$ ou M est la moyenne de tous les résultats, X est le résultat de chaque étude et K le nombre total d'études. Mais nous devons encore une fois considérer le problème du nombre de sujets en pondérant chaque valeur par le N qui lui est associé. La formule finale devient donc $\sum n(X-M)^2 / nK-1$.

3 Prenons deux études qui posent comme hypothèse que les Québécois sont de plus grande taille que les Québécoises. Dans une étude l'échantillon de recherche est composé d'une femme et d'un homme, tandis que dans l'autre toute les personnes vivant au Québec font partie de l'échantillon. Laquelle est plus crédible ($N = 2$ ou $N = 7\ 000\ 000$) ?

Signification de la moyenne et de la variance inter étude

À date, nous avons fait le calcul de deux valeurs : (a) la moyenne obtenue dans les études au sujet du phénomène et (b) la variance entre les résultats obtenus. La moyenne représente le meilleur estimé que nous avons de la taille du phénomène dans la population. Ainsi, la moyenne est sans doute le résultat le plus important qui peut émerger d'une méta analyse.

La variance quant à elle, définit, dans les faits, le degré de désaccords entre les résultats des études. Ainsi une variance de zéro, implique que toutes les études produisent exactement les mêmes résultats, une variance très grande impliquant que les études sont en fort désaccord entre elles. Ainsi, la deuxième statistique importante pour la méta analyse est la variance des résultats. Comme nous avons vu, cette statistique se calcule avec autant de facilité que la moyenne.

Le calcul de ces deux indicateurs (la moyenne et la variance inter étude) représente la base de la méta analyse, et en fait le processus méta analytique pourrait être terminé à ce stade. Mais il y a un hic! La variance inter étude représente le degré avec lequel les études sont en accord (ou en désaccord) entre elles. Donc, dès que la variance inter étude est plus grande que zéro, nous serions contraints de conclure que les études sont en désaccord et nous serions alors également contraints d'expliquer ce résultat en invoquant des arguments substantiels ou méthodologiques. Mais aurions-nous raison de le faire?

La réponse à cette question est « non », du moins pas avec l'information disponible. Souvenons-nous que chaque résultat des études est un estimé (indépendant) du phénomène à l'étude. C'est à dire que chaque résultat obtenu est un point tiré de la distribution des résultats possibles. Donc, une partie de la variance inter étude est attribuable au simple fait que les études sont des échantillons tirés d'une population de résultats. Nous avons indiqué que les statisticiens définissent ce concept sous la rubrique d'*erreur d'échantillonnage*. Donc, nous pourrions maintenant dire que la variance inter étude calculée à partir des résultats produits par les différentes études, est composée par (au moins) deux composantes : la variance véritable plus la variance due à l'erreur d'échantillonnage. Nous pourrions alors écrire : la variance totale (calculée) = la variance « véritable » + la variance attribuable à l'erreur d'échantillonnage. Ou, plus simplement : le désaccord calculé entre les études = le véritable désaccord + l'erreur due à l'aléa. Nous voyons donc que la variance inter étude que nous avons calculée ne peut pas être interprétée car elle mélange deux composantes : la réalité et l'aléa. Le défi consiste à isoler la réalité en la séparant de l'aléa.

Il n'est pas possible de calculer directement le véritable désaccord inter étude (la variance véritable). Heureusement, l'erreur d'échantillonnage est une statistique qui elle est calculable. La formule, disponible par exemple dans Hunter et Schmidt, 1990, est à première vue assez redoutable, mais il n'en est rien et elle s'applique

extrêmement facilement. Cette statistique prend le nom de *variance due à l'erreur d'échantillonnage*.

Comme nous l'avons vu ci-dessus, la variance totale (V_t) représente la somme de deux termes : la variance véritable (V_v) plus la variance attribuable à l'erreur d'échantillonnage (V_e). Donc $V_t = V_v + V_e$. Nous avons maintenant réalisé deux calculs importants : (a) La variance inter-étude totale (V_t) et (b) l'erreur d'échantillonnage (V_e). Nous pouvons maintenant isoler la variance véritable qui, elle, représente (du moins pour l'instant) le niveau de désaccords inter étude qui importe. Le calcul de la variance véritable devient très simple : $V_v = V_t - V_e$ ou la variance véritable (le vrai désaccord) est égale à la variance inter étude totale calculée à partir des résultats d'études moins la variance attribuable à l'erreur d'échantillonnage. Cette nouvelle statistique, qui prend le nom de *variance nette* est extrêmement importante car elle représente la statistique de base qui sert à l'interprétation des différences entre les études. Lorsque la variance nette tombe à zéro (i.e., les désaccords entre les études sont égaux à l'erreur d'échantillonnage) nous devons conclure qu'il n'y a pas de véritables divergences entre les résultats d'études 4. Le recenseur méta analytique fera porter ces conclusions au sujet des désaccords dans la recherche, en interprétant la variance nette et non pas la variance totale. Par contraste, le recenseur traditionnel devra interpréter la variance totale, ce qui, de fait, est une erreur qui pourrait conduire à de fausses conclusions.

L'exécution de ces quatre calculs : moyenne, variance totale, variance d'échantillonnage et variance nette définissent la méta analyse dans sa forme la plus simple. Les implications de ces calculs sont importantes : dans un recensement narratif typique, les auteurs n'ont pas le choix que d'attribuer les différences entre les études soit à des causes substantielles soit à des causes méthodologiques. Cela mènera souvent à une sur-interprétation des divergences. Typiquement, le recenseur conseillera de nouvelles recherches ou recommandera certaines méthodologies. Cela est une erreur importante car, au moins une partie des différences observées n'est attribuable ni à des erreurs méthodologiques, ni à des paramètres de substance mais tout simplement à l'aléa. La méta analyse permet donc au recenseur d'éviter une sur-interprétation des résultats et l'empêchera de recommander des nouvelles études ou stratégies de recherches qui, sommes toutes, ne sont pas nécessaires.

Ces calculs étant réalisés, la méta-analyse rudimentaire pourrait prendre fin. Sauf que d'autres paramètres connus doivent maintenant être pris en considération.

4 Il est possible, en pratique, que la variance d'échantillonnage soit plus forte que la variance totale. Cela voudrait dire que les résultats d'études sont plus proches, les uns des autres que ceux auxquels nous pourrions nous attendre normalement. Dans ce cas, la variance nette prendrait une valeur négative. Une valeur négative est interprétée comme zéro. Par conséquent, lorsque l'on obtient une variance nette de zéro (ou négative) il faut conclure que toutes les études produisent les mêmes résultats.

La correction pour l'atténuation

Nous savons que la fidélité d'un instrument de mesure a un impact certain sur la relation qui existe entre le trait mesuré et n'importe quelle autre caractéristique. Elle la réduit. Prenons le cas extrême ou la fidélité d'un instrument de mesure est égale à zéro. Cela implique que les performances produites par les sujets telles que mesurées par l'instrument changera d'une administration à une autre, sans pour autant que le trait change. Par exemple, la même personne pourrait obtenir un QI de 140 (fortement au-dessus de la moyenne) à partir d'un test de QI parfaitement non fidèle et obtenir un QI de 80 (fortement sous la moyenne) une semaine plus tard. Par conséquent, la relation entre cette mesure et n'importe quelle autre sera faible, et cela indépendamment de la véritable relation qui existe entre le QI et d'autres variables (tel que le succès académique). Donc, si une étude produit une relation forte entre, par exemple, le QI et le succès académique, et une autre indique une relation faible, cette différence peut être le fruit de trois possibilités : la relation QI-succès est (a) instable (interprétation substantielle); (b) attribuable à l'aléa (erreur d'échantillonnage) ou (c) attribuable à une différence dans la fidélité des mesures.

Afin de distinguer entre ces possibilités le recenseur méta analytique procède à un calcul fort simple qui prend le nom de *correction pour l'atténuation* 5. Le résultat obtenu pour chaque étude est maintenant « corrigé » statistiquement. En pratique ces corrections nous indiquent les résultats qui auraient été obtenus par les études si elles avaient toutes fait appel à des mesures identiques en terme de leurs qualités métrologiques.

La correction pour les variances inégales

Poursuivons l'exemple de la relation entre le QI et le succès académique. Encore une fois nous trouvons deux études où une montre une relation de taille, l'autre non. Maintenant supposons que la mesure du QI dans l'étude A ne démontre aucune variance (tous les étudiants obtiennent essentiellement le même QI) tandis que pour l'étude B la variance est « forte ». Cela n'est pas un exemple farfelu. La première étude aurait pu être réalisée dans une école qui exerce un contrôle très étroit sur les admissions, n'acceptant que les élèves très doués tandis que la deuxième pourrait avoir lieu dans une école publique ou une grande diversité d'étudiants serait inscrit. La relation existante entre le QI et le succès académique serait proche de zéro pour l'étude A et plus importante pour l'étude B. Ici nous aurions noté une différence entre les études que nous aurions pu attribuer à un ensemble de cause substantielle quand, en réalité, les résultats divergeant ne sont que le simple reflet d'un problème d'homogénéité différentiel des échantillons de

5 Ci-dessus, j'avais indiqué ma préférence pour l'utilisation des coefficients de corrélation comme métrique commune. Un avantage additionnel de cette métrique est qu'elle permet l'application d'une formule très simple pour faire la correction pour l'atténuation.

recherche. En termes plus techniques, nous faisons face ici à un problème de *restriction des variances*. Il importe de bien comprendre que la restriction des variances aura un effet quasi certain sur les résultats d'études. Il faudra donc traiter ce problème en corrigeant chaque résultat d'étude. Suite à l'application de ces corrections, chaque étude de la recension aura un niveau égal de restriction des variances. Par conséquent la différence entre les études ne pourra plus être attribuée à la restriction des variances inégales.

Nous avons maintenant décortiqué l'équation finale de la méta analyse :

$$\text{Variance totale} = \text{variance vraie} + \text{variance d'échantillonnage} + \text{atténuation} + \text{restriction des écarts}$$

La procédure méta analytique

Le recenseur méta analytique doit donc, une fois les résultats d'études colligés, procéder aux étapes suivantes : (a) corrections de chaque résultat pour l'atténuation et pour la restriction des variances; (b) calcul de la moyenne pondérée des résultats obtenus; (c) calcul de la variance totale pondérée; (d) calcul de la variance d'échantillonnage et (e) calcul de la variance nette. Les résultats obtenus à l'étape b (moyenne) et l'étape e (variance nette) sont les deux résultats méta analytique fondamentaux. Lorsque la variance nette est égale à zéro, la moyenne obtenue dans l'ensemble des résultats des recherches représente la « véritable » taille du phénomène dans la population. En reprenant l'illustration de la relation QI-succès académique, en obtenant une corrélation moyenne de, disons 0,50, et une variable nette de zéro nous serions contraints de conclure que le succès académique est expliqué à 25% (0,50²) par le QI et que *cet estimé ne varie par d'une étude à l'autre*. Dans un tel cas, le recenseur découvrirait qu'il ne lui est pas nécessaire de présenter une section décrivant les recherches futures. Il pourrait tirer une conclusion solide.

La question des modérateurs

Plusieurs méta analyses concluent, suite aux calculs, qu'il n'y a pas de désaccords dans la documentation; mais il arrive, plus souvent, que malgré tous les calculs, il reste de la variance nette. Ici, nous sommes contraints de dire que le phénomène à l'étude est incohérent ou qu'il opère de façons différentes dans des conditions variées. Le recenseur fait face maintenant à l'énorme problème d'expliquer pourquoi. Là encore, la méta analyse offre une technique relativement simple qui aidera le chercheur dans cette tâche. Nous parlons ici de la recherche des modérateurs.

Un *modérateur* est une variable qui altère (modère) la relation qui existe entre deux autres variables. Par exemple, la relation entre le comportement social et le nombre d'enfants dans la famille pourrait être différente pour les garçons et les filles. Le genre serait donc une variable modératrice.

Lorsque la variance nette d'une méta analyse est plus grande que zéro il faudra conclure à l'existence de variables modératrices⁶. Comment identifier les variables modératrices? Il faut de l'art et de la technique!

Le recenseur qui conclut à l'existence de variance nette poursuit donc son analyse. En fait deux étapes sont requises : (a) identification préliminaire d'une variable modératrice potentielle; (b) mise en marche d'un processus technique qui confirme (ou non) que le modérateur potentiel en est un.

Pour comprendre cette démarche il faut d'abord saisir ce que nous voulons faire. En observant la variance nette nous constatons que le phénomène opère différemment sous différentes conditions. Le recenseur émet donc une hypothèse qui pourrait se lire comme suit : la différence entre les études est attribuable à la variable X. Pour vérifier cette hypothèse, il fera une méta analyse pour le groupe d'études qui opèrent sous les conditions A et une autre méta analyse pour le groupe d'études qui opèrent sous les conditions B. Dans notre exemple, une méta analyse pourrait être réalisée seulement pour les résultats portant sur les garçons et une autre regroupant seulement les études réalisées avec les filles. Si les variances nettes de ces deux méta analyses tombent à zéro, il faudra conclure que le phénomène opère de manière cohérente pour les filles et de manière cohérente, mais différente, pour les garçons. Ainsi, le recenseur aura correctement identifié la variable qui modère la relation principale qui fait l'objet du travail.

Mais il se pourrait qu'après avoir identifié une hypothèse modératrice et réalisé les calculs requis que la variance nette ne chute pas. Dans ce cas, le recenseur infirme son hypothèse. Dans l'exemple du comportement social vs le nombre d'enfants, lorsque la variance nette demeure inchangée le genre n'est plus une explication de la différence entre les résultats d'études. Il lui faudra maintenant chercher une autre explication mais, au moins, il est en mesure d'éliminer une explication pour la divergence. Peut-être que l'âge des enfants, ou leur statut socio-économique expliquerait mieux les différences inter études. Il s'agirait alors de refaire les méta analyses en utilisant ces autres variables comme variables modératrices. Le processus de recherche de modérateur se poursuivra jusqu'à ce que les variances nettes tombent à zéro ou jusqu'à épuisement des modérateurs potentiels disponibles ou des études.

Sommaire

La procédure méta analytique débute de la même manière que les recensions traditionnelles, soit par un examen de la documentation. Mais là s'arrête les similarités. Le recenseur méta analytique examinera chaque étude afin d'en extraire

⁶ Ici, nous simplifions un peu quand même. D'autres choses, par exemples les erreurs de calculs, les erreurs de codification des données, etc. peuvent aussi se répercuter en gonflant la variance nette.

son résultat quantitatif. Ces résultats seront ensuite convertis en une métrique commune. Chaque résultat d'étude sera ensuite corrigé pour l'atténuation et pour la restriction des variances. Les variances totales et l'erreur d'échantillonnage seront ensuite calculées afin de produire l'estimation de la variance nette. Si cette variance nette est égale à zéro, le processus méta analytique prend fin et le recenseur est en mesure de tirer sa conclusion en interprétant la moyenne obtenue. Si la variance nette est supérieure à zéro, il enclenchera maintenant une procédure de recherche de modérateurs.

Discussion

La mise en œuvre de la méta analyse implique de nombreux calculs mais ils sont tous très simples à exécuter et il existe sur le marché des logiciels qui peuvent les réaliser sans difficultés. L'avantage principal de la méta analyse provient de son habileté à limiter les biais des recenseurs. Un recenseur qui n'affectionne pas une certaine méthodologie ou certains instruments serait tenté d'invoquer ces méthodologies et instruments « indésirables » pour appuyer son explication des incohérences dans les résultats. Mais en faisant appel à la procédure méta analytique, ces préférences et biais seraient soumis à une vérification empirique et ils seraient retenus ou non dépendant des résultats. La limite principale de la méta analyse est qu'elle n'est utilisable que dans les champs disciplinaires qui sont essentiellement quantitatifs. Les données qualitatives ne sont pas utilisables à moins qu'il soit possible de traduire les résultats qualitatifs en métrique quantitative.

En terminant, il m'importe de souligner qu'il est fort probable que seule une minorité de lecteur exécutera une recherche méta analytique dans leur carrière. En revanche, il est fort probable que la plupart des lecteurs auront à lire et à comprendre les méta analyses publiées par d'autres. Il est désormais de plus en plus rare de voir des recensions de documents qui ne se servent pas de procédures méta analytiques. En lisant les articles méta analytiques il importe de faire porter votre attention (dans l'ordre) :

1. à la moyenne inter étude obtenue (souvent le terme « taille de l'effet » est utilisé pour décrire cette statistique), si cette moyenne est faible, l'effet est faible;
2. à la variance nette qui, si elle est supérieure à zéro impliquera une recherche de conditions modératrices et
3. aux modérateurs eux-mêmes.

Référence

Hunter, J.E., & Schmidt, F.L. (1990). *Methods of meta-analysis*. Newberry Park, California : Sage Publications.