

Manuel des normes du What Works Clearinghouse

Version 4.0 (Octobre 2017)

Institute of Education Sciences

What Works Clearinghouse Standards Handbook

Traduit en français par
Nathalie ROQUES (juillet 2019)

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2017, October). *What Works Clearinghouse: Standards Handbook (Version 4.0)*. Retrieved from <http://whatworks.ed.gov>

https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

Note concernant la traduction

Cette traduction est libre et n'engage en aucune manière le What Works Clearinghouse et ses organismes de tutelle.

Le document original, écrit par des américains, s'appuie sur des études effectuées aux USA. Pour faciliter la lecture par des français, le niveau des élèves a lui aussi été traduit (ainsi les élèves de 8^{ème} grade américains sont devenus des élèves de 4^{ème}).

La mise en forme de la version originale a été le plus souvent laissée inchangée (usage des guillemets, mots en italique, ...). Certaines expressions américaines ont été conservées dans le texte, notamment les titres d'ouvrages et les noms des organismes américains. Dans quelques cas, une traduction les accompagne.

Seuls les deux premiers chapitres de ce manuel ont été partiellement traduits, et les passages non traduits sont signalés par le symbole .../....

Vous trouverez également une traduction partielle du *What Works Clearinghouse Procedures Handbook* (Version 4.0). Ces documents présentent à eux deux les règles et normes qui encadrent et soutiennent tous les textes et outils publiés par le WWC.

Ces traductions sont en accès libre sur le site www.mathadoc.fr

Table des matières

I. INTRODUCTION	7
II. ESSAIS CONTRÔLÉS RANDOMISÉS ET ÉTUDES QUASI-EXPÉRIMENTALES	11
A. L'unité d'affectation est le sujet	11
Étape 1. Design de l'étude : l'appartenance au groupe d'intervention et au groupe de comparaison est-elle déterminée par un processus d'affectation aléatoire ?	13
Étape 2. Attrition de l'échantillon : la combinaison de l'attrition globale et de l'attrition différentielle est-elle élevée ?	16
Perte d'échantillon qui n'est pas considérée comme une attrition	18
Mesurer l'attrition lorsque certaines données sont imputées	20
Étape 3. Équivalence initiale : une équivalence est-elle établie au niveau initial pour les groupes de l'échantillon analysé ?	20
B. L'unité d'affectation est le groupe	22
Critères de sélection pour déterminer si l'unité d'affectation de l'étude est le groupe	23
Procédure d'examen de l'effet probant d'une intervention sur des sujets (étapes 1 à 4)	25
Étape 1. L'étude est-elle un ECR dont l'unité d'affectation est le groupe avec une attrition faible au niveau du groupe ?	26
Étape 2. Existe-t-il un risque de biais dû aux sujets « nouveaux arrivés » dans les groupes ?	26
Étape 3. Existe-t-il un risque de biais dû à des non-réponse de sujets ?	28
Étape 4. L'étude établit-elle l'équivalence initiale des sujets pour les groupes de l'échantillon analysé ?	29
Procédure d'examen de l'effet probant d'une intervention sur des groupes (étapes 5 à 7)	29
Étape 5. L'échantillon des sujets analysés est-il représentatif des groupes ?	30
Étape 6. L'étude est-elle un ECR avec une attrition faible au niveau des groupes ?	31
Étape 7. L'étude établit-elle l'équivalence initiale des groupes de l'échantillon analysé ?	31

Figures

Figure I.1. Étapes de l'examen systématique des études dans les manuels du WWC

Figure II.1. Classement WWC des ECR et des EQE dont l'unité d'affectation est le sujet

Figure II.2. Attrition et biais potentiels

Figure II.3. Traitement des exclusions d'échantillons dans les ECR par le WWC

Figure II.4. Procédure d'examen des études dont l'unité d'affectation est le groupe

Tableaux

Tableau II.2. Taille d'effet (valeur absolue) initiale

Liste des abréviations

I. INTRODUCTION

Il est essentiel que les décideurs en matière d'éducation aient accès aux meilleures données probantes concernant l'efficacité des programmes, des politiques et des pratiques en matière d'enseignement. Cependant, il peut être difficile, long et coûteux pour ces décideurs d'avoir accès aux conclusions d'études pertinentes sur l'efficacité de ces interventions¹ et d'en tirer les conclusions qui s'imposent. Le centre de documentation *What Works Clearinghouse* (WWC) répond à ce besoin d'informations synthétiques et dignes de confiance en identifiant les recherches effectuées sur des interventions dans le domaine de l'enseignement, en évaluant la qualité de ces recherches, en résumant et en diffusant les preuves issues d'études respectant les normes édictées par le WWC.

Le WWC est une initiative de l'Institut des sciences de l'éducation (IES) du département de l'Éducation des États-Unis, créé en vertu de la loi de 2002 (Education Sciences Reform Act). Il s'agit d'un volet important de la stratégie de l'IES qui consiste à mettre à profit les recherches, les évaluations et les statistiques rigoureuses et pertinentes dans le but d'améliorer le système de l'éducation nationale. La mission du WWC est d'être **une source centrale et fiable de données scientifiques probantes sur « ce qui marche » (*What Works*) dans le domaine de l'enseignement**. Le WWC examine la recherche concernant les interventions visant à améliorer les résultats pertinents sur le plan éducatif, notamment celles qui concernent les élèves et le personnel enseignant.

Les processus d'examen systématiques menée par le WWC constituent la base de bon nombre des documents et des outils produits par cette institution. Elles appliquent et suivent des normes et des procédures cohérentes, objectives et transparentes, tout en assurant une couverture complète de la littérature pertinente existante. Les examens systématiques menée par le WWC suivent cinq étapes :

1. *Développement du protocole de l'examen*. Un protocole formel est élaboré pour chaque examen, y compris pour chaque centre d'intérêt du WWC (par exemple, l'alphabétisation des adolescents ou les mathématiques à l'établissement primaire) afin de définir les paramètres de la recherche qu'il convient d'inclure dans le champ de l'examen (par exemple les caractéristiques de la population et les types d'interventions) ; la recherche documentaire (par exemple les mots-clés utilisés pour interroger les bases de données) ; et toute application spécifique des normes à un sujet donné (par exemple, établissement des seuils acceptables d'attrition des échantillons et spécification des caractéristiques qui permettent de déterminer l'équivalence initiale des groupes étudiés).
2. *Identification de la littérature pertinente*. Les études sont rassemblées grâce à une recherche exhaustive de la littérature de recherche publiée et non publiée disponible. Cette recherche s'appuie entre autres sur des bases de données électroniques.
3. *Sélection des études*. Les manuscrits sont, dans un premier temps, examinés pour évaluer leur éligibilité afin de déterminer s'ils répondent à la question de recherche

¹ Une intervention est donc constituée par un ou plusieurs éléments caractérisés comme étant un programme, une politique ou une pratique.

initiale, fournissent des preuves potentiellement crédibles de l'efficacité d'une intervention et s'inscrivent dans le cadre du protocole d'examen.

4. *Examen des études*. Chaque étude éligible est examinée par rapport aux normes élaborées par le WWC. Le WWC utilise un processus d'examen structuré pour évaluer la validité des résultats présentés dans ces études en relation avec l'efficacité des interventions sur l'enseignement. Les normes développées par le WWC se concentrent sur la validité causale au sein de l'échantillon de l'étude (validité interne) plutôt que sur l'estimation de la réplication potentielle des résultats de l'étude et de son élargissement éventuel à d'autres contextes (validité externe)².
5. *Présentation des résultats*. Les détails de l'examen et ses conclusions sont résumés sur le site Web du WWC, et souvent dans une publication du WWC. Pour nombre de ses documents et outils³ produits, le WWC combine les résultats d'études individuelles pour donner naissance à des mesures synthétisant l'efficacité d'une intervention, en présentant notamment l'ampleur des résultats et leur niveau de preuve.

En outre, le WWC examine certaines études en dehors du processus d'examen systématique, par exemple dans le cas où certaines études reçoivent une attention particulière de la part des médias. Ces examens sont également guidés par un protocole d'examen et utilisent les mêmes normes et procédures de présentation des résultats.

Ce *What Works Clearinghouse Standards Handbook* (Version 4.0) traduit ici par *Manuel des normes WWC* fournit une description détaillée des normes utilisées par le WWC pour examiner les études (étape 4 ci-dessus). Les étapes 1 à 3 et l'étape 5 quant à elles sont décrites dans un autre document intitulé *What Works Clearinghouse Procedures Handbook*⁴. Depuis octobre 2017, ces deux documents remplacent le document unique utilisé antérieurement, le *What Works Clearinghouse Procedures and Standards Handbook* (version 3.0, publiée en mars 2014)⁵. La figure I.1 montre comment les étapes du processus d'examen systématique du WWC sont réparties entre les deux manuels.

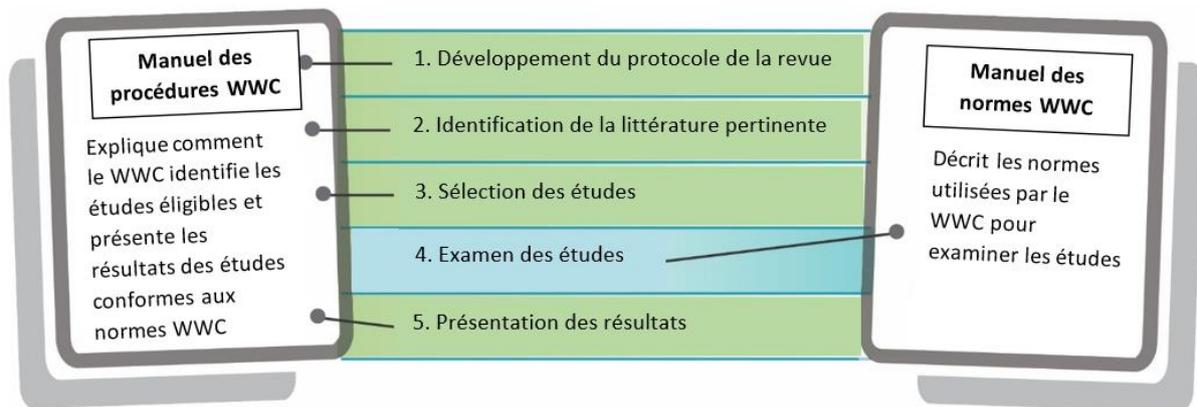
² Ces normes concernent le design (ou conception) des études qui comparent des groupes.

³ Il s'agit notamment des *Rapports d'intervention*, des *Guides des pratiques* et du site internet Find What Works mis à disposition du grand public librement.

⁴ Dont la traduction partielle est intitulée *Manuel des procédures WWC*

⁵ Une nouvelle version des deux manuels est en cours de rédaction et sera effective à l'automne 2019.

Figure I.1. Étapes de l'examen systématique des études dans les manuels du WWC



Ce *Manuel des normes WWC* fournit une description détaillée des normes utilisées par le WWC lors de l'examen des études répondant aux critères d'éligibilité, et qui mettent en œuvre notamment l'un des designs éligibles par le WWC : l'essai contrôlé randomisé (ECR), l'étude au design quasi expérimental (EQE), l'étude au design à régression par discontinuité (ERD) et l'étude de cas unique (ECU). Les études utilisant d'autres designs ne sont pas examinées par le WWC. Le WWC considère les essais contrôlés randomisés (ECR) et les études suivant des design quasi expérimentaux (EQE) comme des études dont la conception est basée sur l'étude de groupes. Les études examinées par rapport aux normes WWC reçoivent l'un des trois classements suivants indiquant le niveau de preuve de l'étude : *conforme sans réserve aux normes WWC*⁶, *conforme avec réserve aux normes WWC* ou *non conforme aux normes WWC* (*Meets WWC Design Standards Without Reservations, Meets WWC Design Standards With Reservations, or Does Not Meet WWC Design Standards*).

.../...

⁶ La mention « group design » qui caractérise les ECR et les EQE est sous-entendue dans la traduction par la suite. Il existe aussi des normes pour les ERD et ECU (non traduit ici).

II. ESSAIS CONTRÔLÉS RANDOMISÉS ET ÉTUDES QUASI-EXPÉRIMENTALES

Ce chapitre décrit les éléments de base concernant l'examen des études dont le design se fonde sur l'étude d'une (ou plusieurs) intervention sur un (ou plusieurs) groupes : les essais contrôlés randomisés (ECR) et les études à designs quasi expérimentaux plus simplement dénommée par la suite études quasi-expérimentales (EQE). Alors que les ECR s'appuient sur une assignation aléatoire pour former des groupes d'intervention et de comparaison⁷, les EQE forment ces groupes en utilisant des méthodes autres que la randomisation. Les normes sont présentées séparément pour les études affectant des sujets (tels que des élèves) à une condition et les études affectant des groupes⁸ (telles que des classes ou des établissements) à une condition. Le chapitre se termine par des directives spécifiques pour l'examen d'études utilisant diverses approches analytiques communes (non traduit ici).

Bien que le design des études s'appuyant sur la régression par la discontinuité (ERD) soit parfois considéré comme basé sur une étude de groupes, le WWC applique des normes distinctes pour examiner les ERD éligibles. Si une valeur seuil sur une mesure connue est utilisée pour affecter des sujets aux groupes d'intervention et de comparaison, l'étude peut alors être éligible pour être examinée en tant que ERD. Les critères d'éligibilité et les normes WWC pour l'examen des ERD sont décrits au chapitre III (non traduit ici).

A. L'unité d'affectation est le sujet

Dans cette section, nous décrivons les trois étapes pour examiner les ECR et les EQE qui affectent des sujets à la condition d'intervention ou de comparaison :

Étape 1 : évaluation du design de l'étude

Étape 2 : évaluation du taux d'attrition de l'échantillon et

Étape 3 : évaluation de l'équivalence initiale des groupes d'intervention et de comparaison (avant l'intervention).

Pour être classé au niveau le plus élevé du classement WWC des études adoptant un design basé sur l'étude de groupes, à savoir *conforme sans réserve aux normes WWC*⁹, l'étude doit être un ECR avec un faible taux d'attrition. Une EQE ou un ECR avec une attrition élevée peuvent être classée au niveau *conforme avec réserve aux normes WWC* s'ils satisfont à l'obligation d'équivalence initiale des groupes selon laquelle les groupes d'intervention et de comparaison doivent être similaires avant l'intervention. Les EQE ou les ECR présentant un taux d'attrition élevé qui ne satisfont pas à l'exigence d'équivalence initiale seront classés

⁷ On parle aussi souvent de groupe traitement et de groupe contrôle.

⁸ On parle aussi souvent de grappes (*cluster*).

⁹ Sous-entendu (y compris dans la suite du texte) qu'il s'agit des normes concernant les ECR et les EQE (d'autres normes sont utilisées par le WWC pour définir l'éligibilité d'études suivant d'autres designs).

comme *non conforme aux normes WWC* (Figure II.1). Chaque description d'une étape de la procédure d'examen se termine par un résumé de l'ensemble des résultats possibles, et pour chaque résultat possible les lecteurs seront orientés vers l'étape suivante appropriée.

Toutefois, les études dont l'unité d'affectation est le sujet et qui satisfont aux exigences décrites dans les étapes 1 à 3 doivent également satisfaire à deux exigences supplémentaires pour pouvoir être classée *conforme sans réserve aux normes WWC* ou *conforme avec réserve aux normes WWC*.

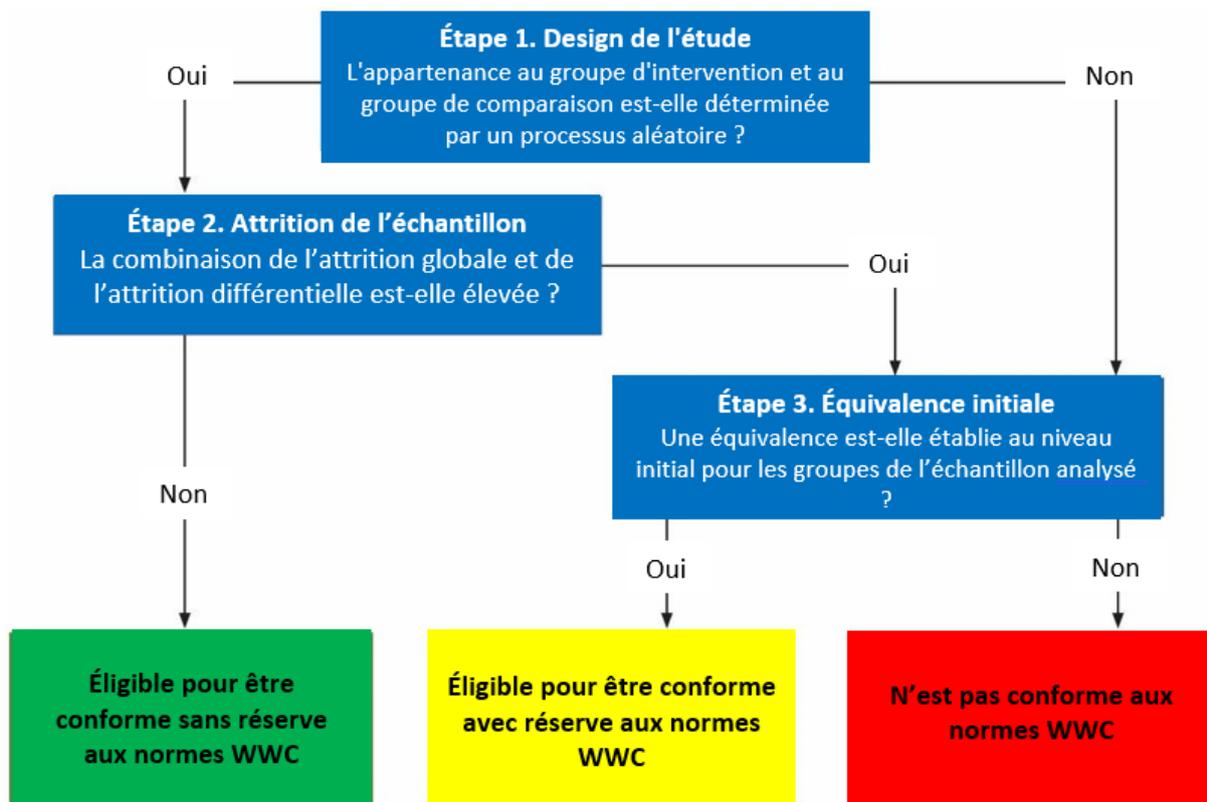
Ces exigences supplémentaires, décrites au chapitre IV (non traduit ici), stipulent que l'étude doit :

- Examiner au moins une mesure de résultat éligible répondant aux exigences de l'examen, et
- Être exempte de facteurs de confusion.

En outre, lorsque les études utilisent certaines approches analytiques, y compris des analyses de scores de propension, des analyses dans lesquelles les sujets sont observés sur plusieurs périodes, des méthodes pour traiter les données manquantes, ou incluent des covariables endogènes, des directives et des normes supplémentaires peuvent s'appliquer comme décrit dans la section C (non traduit ici).

.../...

Figure II.1. Classement WWC des ECR et des EQE dont l'unité d'affectation est le sujet



Note : pour être classée *conforme sans réserve aux normes WWC* ou *conforme avec réserve aux normes WWC*, l'étude doit également satisfaire aux exigences du chapitre IV, elle doit notamment examiner au moins un indicateur de résultat éligible qui réponde aux exigences de l'examen et être exempte de facteurs de confusion.

Étape 1. Design de l'étude : l'appartenance au groupe d'intervention et au groupe de comparaison est-elle déterminée par un processus d'affectation aléatoire ?

Essais contrôlés randomisés (ECR)

La caractéristique distinctive d'un ECR est l'affectation aléatoire des sujets de l'étude à l'un des deux groupes différenciés selon qu'ils reçoivent ou non l'intervention. Les chercheurs peuvent utiliser plusieurs méthodes pour procéder à une affectation aléatoire. Par exemple, les méthodes acceptables d'affectation aléatoire peuvent être associées à un regroupement des élèves en blocs avant randomisation, à la mise en œuvre d'un sous-échantillonnage aléatoire, à l'affectation de sujets à des groupes avec des probabilités différentes et à la formation de groupes de tailles différentes.

Pour que l'affectation soit valide, les sujets doivent être attribués entièrement par hasard et avoir une probabilité non nulle d'être assignés à chaque groupe. Il n'est pas nécessaire que les sujets aient une probabilité égale d'être affectés à chacun des groupes, et la probabilité d'être affectée à un groupe particulier peut varier d'un sujet à l'autre. Cependant, si les sujets sont affectés à un groupe avec des probabilités différentes (c'est-à-dire, si la probabilité d'être affecté à un groupe diffère pour les sujets affectés au même groupe), alors les résultats doivent être ajustés de manière à tenir compte de ces différentes probabilités d'affectation (voir la discussion sur le deuxième type d'ECR compromis dans la sous-section suivante). Cette exigence s'applique également si la probabilité d'affectation à un groupe varie d'un bloc à l'autre dans le cadre d'affectation aléatoire stratifiée.

ECR compromis

Lorsque la validité d'un processus d'affectation aléatoire ou l'analyse d'un processus d'affectation aléatoire par ailleurs bien exécuté est compromise, l'étude est examinée à l'aide de la procédure utilisée pour les EQE. Un ECR affectant des sujets à la condition d'intervention ou de comparaison peut être compromis de quatre manières.

- L'ECR est compromis lorsqu'il inclut des sujets dans l'échantillon utilisé pour estimer les résultats (c'est l'échantillon analysé) qui n'ont pas été affectés de façon aléatoire.
- L'ECR est compromis si les sujets sont affectés de manière aléatoire à un groupe avec des probabilités différentes, mais que les résultats sont basés sur une analyse ne prenant pas en compte les différentes probabilités d'affectation. Considérons une étude qui effectue une affectation aléatoire à partir de deux blocs d'élèves. L'étude inclut le même nombre d'élèves dans les deux blocs, mais les élèves du bloc A ont un niveau élevé au départ, tandis que les élèves du bloc B ont un niveau faible au départ. L'étude affecte 70% des élèves du bloc A à la condition d'intervention, mais n'attribue que 30% des étudiants du bloc B à la condition d'intervention. Dans ce cas, le groupe d'intervention comprend 70% d'élèves très performants, tandis que le groupe de comparaison comprend 70% d'élèves peu performants. Si les données sont analysées sans prendre en compte les différentes probabilités d'affectation, les groupes dissemblables peuvent donner l'impression que l'intervention a un effet positif, même si ce n'est pas le cas.

.../...

Si les auteurs de l'étude décrivent un processus d'affectation aléatoire suggérant des probabilités d'affectation variables, mais n'effectuent aucun ajustement, l'ECR est compromis et l'étude est examinée à l'aide de la procédure d'évaluation des EQE.

- L'ECR est compromis lorsque l'enquêteur modifie l'appartenance de sujets à un groupe après une assignation aléatoire. Prenons une étude dans laquelle certains sujets affectés à la condition d'intervention n'ont pas reçu l'intervention, mais sont restés dans l'étude. Par exemple, certains élèves initialement affectés à une classe mettant en œuvre la condition d'intervention peuvent en fait fréquenter une classe différente qui a implémenté la condition de comparaison. Si les auteurs de l'étude analysent ces sujets en tant que membres du groupe de comparaison, en se basant sur le fait qu'ils n'ont pas reçu l'intervention, l'assignation aléatoire est compromise. Toutefois, si les auteurs de l'étude analysaient ces sujets en tant que membres du groupe d'intervention en tenant compte de leur affectation d'origine (analyse en intention de traiter), l'intégrité de l'affectation aléatoire serait maintenue. En d'autres termes, tous les sujets ne doivent pas nécessairement recevoir la condition qui leur est attribuée, mais ils doivent tous être analysés en fonction de la condition initialement à laquelle ils ont été affectés. Notez que les études qui traitent de la non-conformité¹⁰ en calculant l'effet moyen du traitement sur les sujets traités (*complier average causal effects (CACE)*) peuvent être éligibles à un examen en s'appuyant sur les normes décrites dans la section D de ce chapitre (non traduit ici).

- L'ECR est compromis lorsque lorsqu'un auteur de l'étude manipule l'échantillon analysé pour exclure certains sujets en fonction d'événements survenus après l'introduction de l'intervention lorsqu'il existe un lien clair entre le statut du groupe et le motif de l'exclusion. Un lien clair est présent lorsque l'exclusion est basée sur une mesure qui pourrait être en lien avec l'affectation à la condition d'intervention ou à la condition de comparaison. Tous les cas d'exclusions effectués par les auteurs ne rentrent pas dans cette catégorie, comme illustré dans les exemples suivants. Ces exemples illustrent les trois manières selon lesquelles le WWC traite les cas d'exclusions, résumées par la figure II.3 : (1) l'exclusion compromet l'ECR, (2) l'exclusion est considérée comme une attrition ou (3) l'exclusion est considérée comme ignorable (c'est-à-dire qu'elle n'est pas comptabilisée comme attrition et ne compromet pas l'ECR) :

- o **ECR compromis.** Si une intervention risque de modifier l'assiduité des élèves (par exemple, elle pourrait influencer de manière plausible sur la motivation des élèves à venir en classe) et que les auteurs de l'étude excluent de l'analyse les élèves présentant un taux élevé d'absentéisme, l'ECR est compromis. Ce résultat est représenté par l'encadré rouge de la figure II.3.

- o **Attrition.** Supposons que les auteurs de l'étude regroupent les élèves par paires et affectent aléatoirement un élève de chaque paire à la condition d'intervention. Si les données manquaient pour l'un des élèves de la paire, l'exclusion des deux élèves de la paire (ou de tout autre bloc de randomisation plus important) de l'analyse ne compromettrait pas l'affectation aléatoire, car il n'y a pas de lien clair entre l'intervention et l'attrition de la paire. Dans cet exemple, la paire exclue est comptabilisée en tant qu'attrition, ce qui ne

¹⁰ En tenant compte par exemple des élèves du groupe intervention qui n'ont pas été soumis à l'intervention.

compromet pas un ECR et est décrit en détail à l'étape 2 ci-dessous. Ce résultat est représenté par la case jaune de la figure II.3.

o **Ignorable** (n'est pas comptabilisé comme une attrition et ne compromet par l'ECR). Certains exemples d'exclusions ne sont considérés ni comme une attrition ni comme une source de compromission. Par exemple, si les auteurs de l'étude excluent des élèves au hasard de la collection des données ou excluent de l'échantillon analysé les élèves partageant une certaine caractéristique mesurée avant l'introduction de l'intervention (par exemple, avoir bénéficié de programmes d'enseignement individualisé avant l'étude), ces exclusions ne compromettent pas l'affectation aléatoire. De plus, les sujets exclus peuvent être retirés du calcul de l'attrition car ils étaient basés sur une caractéristique avérée avant l'intervention. Ce résultat est représenté par la case verte dans la figure II.3, et la distinction entre ce type d'exclusion et les exclusions comptabilisées en tant qu'attrition est examinée plus en détail à l'étape 2 de la sous-section sur la perte de sujets d'un échantillon qui n'est pas considérée comme une attrition.

Le WWC considère qu'un ECR n'est compromis que lorsque le chercheur analyse des données relatives à l'un de ces quatre cas. Certaines procédures de randomisation valides peuvent produire des groupes d'intervention et de comparaison qui semblent différents et dont ces différences sont le fait du hasard. Le WWC ne considère pas que ces différences dues au hasard compromettent l'ECR, et ces études sont examinées en utilisant la procédure d'examen habituelle pour les ECR valides. En outre, si une étude rapporte plusieurs résultats, dont seulement quelques-uns sont reconnus par le WWC comme pouvant compromettre un ECR, les résultats préservant l'intégrité de l'affectation aléatoire peuvent être examinés à l'aide de la procédure des ECR valides.

Étude quasi expérimentale (EQE)

Une étude peut être considérée comme une EQE si elle compare les résultats des sujets d'un groupe d'intervention aux résultats des sujets d'un groupe de comparaison, et que la détermination de la composition des deux groupes ne repose pas sur une affectation aléatoire. Les groupes peuvent être définis à travers une variété de procédures et l'étude être éligible à l'examen par le WWC tant que les groupes sont exclusifs, ce qui signifie qu'un sujet ne peut être analysé qu'en tant que membre d'un seul groupe. L'affectation à l'intervention peut dépendre à la fois de caractéristiques observables et non observables. Par exemple, des élèves peuvent être éligibles à un programme se déroulant après l'établissement, mais seuls certains peuvent choisir d'y participer. Les élèves qui n'ont pas choisi de participer sont désignés comme groupe de comparaison. Dans ce cas, les caractéristiques des groupes d'intervention et de comparaison diffèrent. Les deux groupes peuvent différer sur les caractéristiques que les chercheurs ont pu mesurer, telle que des scores à certains tests, ou sur des caractéristiques que les chercheurs n'ont pas été en capacité de mesurer, telle que la motivation. Même dans le cas où les caractéristiques mesurées se montreraient équivalentes, il pourrait exister des différences entre des caractéristiques non mesurées qui pourraient introduire un biais dans l'estimation de l'effet de l'intervention. Un biais est une différence systématique entre l'effet réel de l'intervention et l'effet estimé, qui peut conduire à des conclusions erronées sur les effets de l'intervention. Pour cette raison, les EQE ne peuvent pas être classées comme *conformes sans réserve aux normes WWC*, mais peuvent être classées comme *conformes avec réserve aux normes WWC*.

Procédure d'examen du WWC pour l'étape 1 de l'examen des études dont l'unité d'affectation est le sujet

- Si les sujets ont été placés dans chaque condition d'étude par le biais d'un processus d'affectation aléatoire valide et que l'ECR n'a pas été compromis, passer à l'étape 2.
- Si les sujets n'ont pas été placés dans chaque condition de l'étude selon un processus d'affectation aléatoire valide, mais que l'étude est éligible pour un examen en tant que EQE, passer à l'étape 3.

Étape 2. Attrition de l'échantillon : la combinaison de l'attrition globale et de l'attrition différentielle est-elle élevée ?

L'attrition se produit lorsqu'un résultat n'est pas disponible pour tous les sujets initialement affectés aux groupes d'intervention et de comparaison. Même les ECR bien conçus peuvent enregistrer des taux et des modes d'attrition de l'échantillon qui compromettent la comparabilité initiale des groupes d'intervention et de comparaison et qui peuvent éventuellement conduire à des estimations de l'efficacité de l'intervention biaisées. L'attrition conduit à un biais lorsqu'elle est liée au résultat qui intéresse l'auteur. Pour les ECR, le WWC s'inquiète à la fois de l'attrition globale (c'est-à-dire du taux d'attrition de l'ensemble de l'échantillon, mesuré par le pourcentage de l'échantillon randomisé perdu) et de l'attrition différentielle (c.-à-d. la différence en points de pourcentage des taux d'attrition pour les groupes d'intervention et de comparaison), car les deux types d'attrition contribuent au biais potentiel de l'effet estimé.

Attrition forte ou faible

La norme WWC qui permet d'évaluer l'effet de l'attrition, repose sur un modèle théorique du biais d'attrition et des hypothèses empiriques. Le modèle décrit les biais potentiels en fonction des taux d'attrition globale et différentielle et de la relation entre l'attrition et les résultats. Pour déterminer les valeurs raisonnables qui permettent d'évaluer l'étendue du biais d'attrition potentiel dans une étude, le WWC a formulé des hypothèses sur la relation entre l'attrition et les résultats compatibles avec les conclusions de plusieurs essais randomisés en sciences de l'éducation.

Le WWC a évalué des niveaux de biais attendus associés à différentes combinaisons de taux d'attrition globaux et différentiels dans le cas où l'on peut s'appuyer sur une hypothèse optimiste et dans le cas où l'on peut s'appuyer sur une hypothèse prudente en ce qui concerne le lien entre l'attrition et les résultats. La figure II.2 illustre les combinaisons générant des niveaux de biais attendus tolérables (région verte), potentiellement tolérables (région jaune) et inacceptables (région rouge). Un niveau de biais tolérable est défini comme diminuant le résultat d'une taille d'effet de 0,05 écart-type ou moins, ce qui représente environ 2 centiles pour un élève se situant au 50^{ème} centile. Par exemple, si les résultats rapportés dans une étude suggèrent que l'intervention passera l'élève du 50^{ème} centile au 60^{ème} centile (ce qui correspond à une taille d'effet de 0,25 écart-type), l'effet réel de l'intervention devrait déplacer l'élève au 58^{ème} centile (correspondant à une taille d'effet de 0,20 écart-type). Le seuil fixé par le WWC définissant le niveau tolérable de biais est basé sur une vaste consultation d'experts.

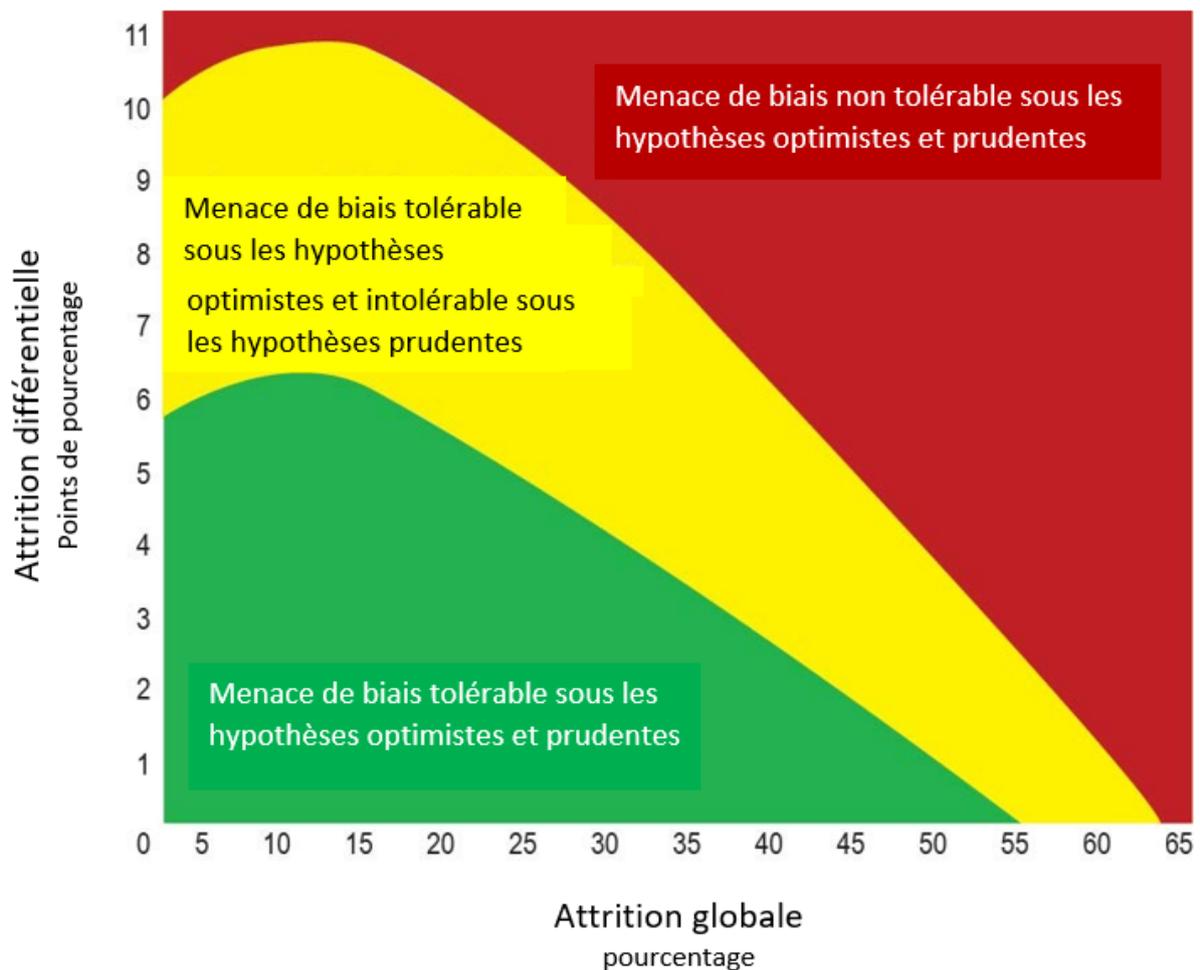
- La zone rouge de la figure II.2 montre des combinaisons d'attrition globale et différentielle qui aboutissent à des niveaux inacceptables de biais potentiels pour les hypothèses optimistes et prudentes.
- La zone verte de la figure II.2 montre des combinaisons d'attrition globale et différentielle qui aboutissent à des niveaux tolérables de biais potentiels pour les hypothèses optimistes et les hypothèses prudentes.

Dans la zone jaune de la figure, la mesure dans laquelle le biais potentiel dépasse 0,05 écart type dépend des hypothèses sélectionnées. Lors de l'élaboration du protocole d'examen, la direction du comité des examinateurs tient compte des types d'échantillons et de la relation probable entre l'attrition et les résultats des études menées le domaine considéré (le personnel du comité des examinateurs est décrit à l'annexe C du *Manuel des procédures WWC* (non traduit)). Le choix des hypothèses (optimistes ou prudentes) est spécifié dans le protocole d'examen ; elles seront appliquées de manière consistante pour toutes les études incluses dans l'examen.

- Si la direction du comité des examinateurs a des raisons de croire qu'une grande partie de l'attrition est exogène aux interventions examinées (c'est-à-dire sans lien avec l'intervention), des hypothèses plus optimistes concernant la relation entre l'attrition et le résultat peuvent être appropriées pour l'examen. Par exemple, la direction du comité des examinateurs peut considérer que les hypothèses optimistes peuvent s'appliquer si elle estime que l'attrition provient le plus souvent du mouvement de jeunes enfants à l'intérieur et à l'extérieur des districts scolaires en raison de la mobilité de la famille ou d'absences de type classique les jours où les évaluations sont effectuées. Dans ce cas, la région jaune sera considérée comme comprenant des combinaisons qui entraînent des niveaux tolérables de biais potentiels, de la même façon que la région verte.
- Si la direction de l'équipe d'examen a des raisons de croire que l'attrition est en grande partie endogène, c'est-à-dire liée à l'intervention, des hypothèses plus prudentes peuvent être appropriées pour l'examen. Par exemple, la direction du comité des examinateurs peut considérer que les hypothèses prudentes doivent s'appliquer lors de l'examen de programme de prévention du décrochage reposant sur une participation volontaire. Dans ce cas, la région jaune sera considérée comme comprenant des combinaisons qui entraînent des niveaux inacceptables de biais potentiels, de la même façon que la région rouge.

.../...

Figure II.2. Attrition et biais potentiels



Note : toutes les combinaisons d'attrition différentielle et d'attrition globale ne sont pas possibles pour une étude donnée. Le protocole d'examen spécifiera quel ensemble de valeurs limites d'attrition s'applique.

Lorsque la combinaison des taux d'attrition globaux et différentiels donne lieu à des niveaux inacceptables de biais potentiels - la région rouge, ainsi que la région jaune si l'on pose des hypothèses plus prudentes - le WWC qualifie cette attrition d'élevée. Lorsque la combinaison des taux d'attrition globaux et différentiels donne lieu à des niveaux tolérables de biais potentiels - la région verte, ainsi que la région jaune si l'on pose des hypothèses plus optimistes - le WWC qualifie cette attrition de faible. Par conséquent, le choix d'hypothèses optimistes ou prudentes donne lieu à un ensemble spécifique de combinaisons de taux d'attrition globaux et différentiels définissant les taux d'attrition faible et élevé à appliquer de la même manière pour toutes les études dans un domaine donné.

.../...

Perte d'échantillon qui n'est pas considérée comme une attrition

Toute perte d'échantillon après l'affectation aléatoire n'est pas incluse dans les calculs d'attrition :

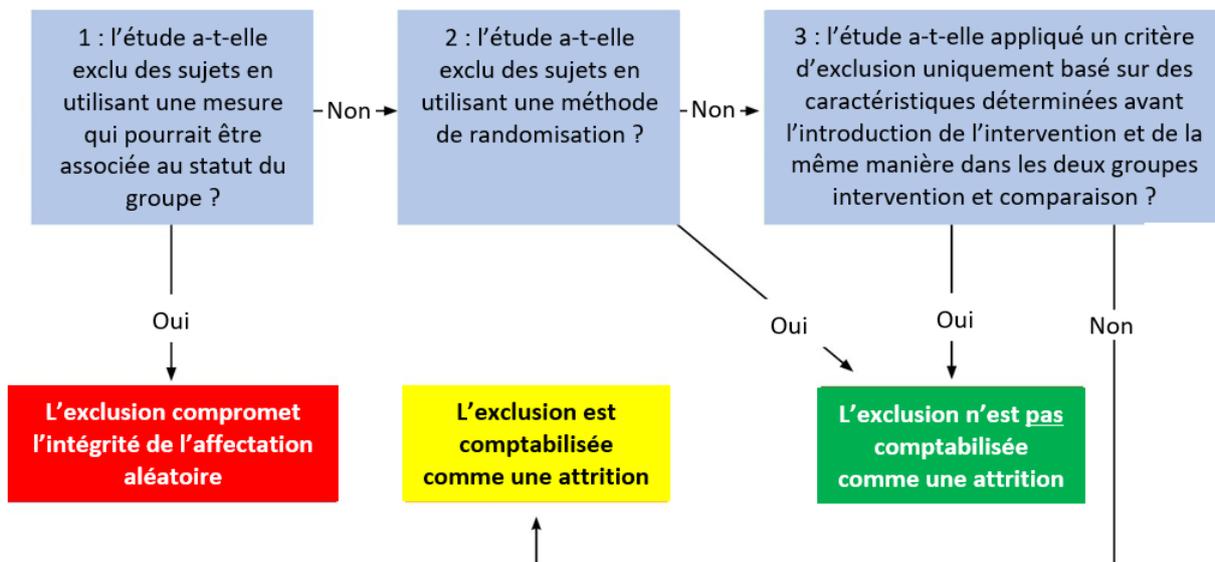
- La perte de sujets de l'échantillon après une affectation aléatoire en raison de phénomènes naturels, tels que des ouragans ou des tremblements de terre, n'est pas considérée comme une attrition lorsque la perte est susceptible d'affecter de la même

manière les membres des groupes d'intervention et de comparaison. Toutefois, lorsque la perte d'échantillon due à un tel phénomène a été concentrée dans un groupe, la perte sera considérée comme une attrition.

- L'échantillon exclu lors de l'analyse des données de résultat pour seulement un sous-ensemble de l'échantillon initial n'est pas considéré comme une attrition si (1) le sous-échantillon du groupe d'intervention ou du groupe de comparaison a été sélectionné au hasard, ou (2) le sous-échantillonnage était basé sur des caractéristiques clairement déterminées avant l'introduction de l'intervention et appliqué uniformément dans les groupes d'intervention et de comparaison. Par exemple, les étudiants exclus pour avoir suivis des programmes d'enseignement individualisés avant l'étude ne seraient pas comptabilisés dans le calcul de l'attrition. Le WWC considère que les caractéristiques qui ont peu de chances de changer avec le temps, notamment le sexe et la race ou l'appartenance ethnique, ont été déterminées avant l'introduction de l'intervention, même lorsque les chercheurs ont collecté ces données ultérieurement.

Le WWC suppose que la perte d'échantillon dont la cause serait différente de celles décrites ci-dessus pourrait influencer sur les résultats, et inclut cette perte d'échantillon dans les calculs de l'attrition. Les règles du WWC relatives à la manière dont l'exclusion des échantillons peut affecter l'évaluation d'un ECR sont résumées dans la figure II.3. Cela inclut les exemples d'exclusions susceptibles de compromettre l'ECR décrit à l'étape 1 ci-dessus (cadre rouge), les exemples d'exclusion ne correspondant pas à une attrition fondée sur les critères ci-dessus (cadre vert) et tous les autres exemples d'exclusion considérés comme des attritions (cadre jaune).

Figure II.3. Traitement des exclusions d'échantillons dans les ECR par le WWC



Une caractéristique peut être déterminée après l'affectation aléatoire mais ne pas être affectée par le statut du groupe. La réponse aux questions des encadrés 1 et 3 peut par conséquent être « non ». Par exemple, la réponse aux trois questions serait « non » dans

une étude où le chercheur a exclu les élèves bénéficiant d'un programme d'éducation individualisé mis en place peu après l'introduction de l'intervention, et que l'intervention n'a aucun lien avec la manière dont les élèves sont identifiés pour participer aux programmes d'éducation individualisés. Dans ce cas, l'échantillon exclu sera considéré comme une perte d'effectif, car certaines caractéristiques des élèves ont pu être déterminées après l'introduction de l'intervention mais ne sont probablement pas liées à celle-ci. Toutefois, s'il existe un lien clair entre l'intervention et la participation des élèves au programme d'éducation individualisé, l'exclusion de l'échantillon compromettrait l'ECR. Par exemple, l'exclusion compromettrait l'ECR si l'intervention permettait aux enseignants d'identifier les élèves qui pourraient avoir besoin de programmes d'éducation individualisés.

Mesurer l'attrition lorsque certaines données sont imputées

Lorsqu'une étude manque de données indispensables pour calculer les résultats, les chercheurs peuvent remplacer les données manquantes par des données imputées d'une manière ou d'une autre, plutôt que d'exclure les sujets pour lesquels les données de l'échantillon analysé sont manquantes. Les sujets de l'échantillon dont les données sont manquantes puis imputées sont considérés comme perdus lors du calcul de l'attrition. En utilisant cette approche, le résultat du calcul du taux d'attrition est le même quelle que soit la manière dont les auteurs traitent les données manquantes. Par exemple, si une étude analyse les données de 100 sujets, dont 90 ont des données mesurées et les 10 restants ont des données imputées par les chercheurs, le taux d'attrition global est de 10%. Voir la section C (non traduite ici) de ce chapitre pour plus d'informations sur la manière dont le WWC examine les études avec des données initiales ou des résultats, et qui sont soit manquantes soit imputées.

Procédure d'examen du WWC pour l'étape 2 de l'examen des études dont l'unité d'affectation est le sujet

- Si l'ECR présente une combinaison des taux d'attrition d'échantillon globaux et différentiels qui répondent aux critères du taux d'attrition faible, alors l'étude peut être classée *conforme sans réserve aux normes WWC*. Pour recevoir ce classement, l'étude doit également satisfaire aux exigences décrites au chapitre IV (non traduit ici), qui stipulent que l'étude doit examiner au moins une mesure de résultat éligible répondant aux exigences de l'examen, et être exempte de facteurs de confusion.
- Si l'ECR a des niveaux d'attrition d'échantillons inconnus ou élevés, passer à l'étape 3.

Étape 3. Équivalence initiale : une équivalence est-elle établie au niveau initial pour les groupes de l'échantillon analysé ?

Les ECR avec un taux d'attrition élevé, les ECR compromises et toutes les EQE ne peuvent pas recevoir le classement WWC le plus élevé en raison de l'incertitude entourant la similarité des groupes d'intervention et de comparaison avant l'introduction de l'intervention. Pour ces études, l'équivalence des groupes d'intervention et de comparaison

sur des caractéristiques spécifiques mesurées au départ (c'est-à-dire avant l'introduction de l'intervention) doit être évaluée pour l'échantillon analysé (à savoir les sujets des groupes d'intervention et de comparaison utilisés pour estimer les résultats). Les caractéristiques sur lesquelles le WWC se repose pour évaluer l'équivalence initiale sont spécifiées dans le protocole d'examen.

Si la différence rapportée d'une caractéristique initiale spécifiée est supérieure à 0,25 écart-type en valeur absolue, calculée sur la base de la variation de cette caractéristique dans l'échantillon regroupant les sujets du groupe d'intervention et dans l'échantillon regroupant les sujets du groupe de comparaison, le WWC considère que les groupes d'intervention et de comparaison ne sont pas équivalents. Pour les différences dans les caractéristiques initiale spécifiées comprises entre 0,05 et 0,25 écart-type, l'analyse doit inclure un ajustement statistique acceptable pour les caractéristiques initiales afin de satisfaire à l'exigence d'équivalence initiale. Les différences inférieures ou égales à 0,05 écart-type ne nécessitent aucun ajustement statistique (tableau II.2). Le chapitre VI du *Manuel de procédures WWC* décrit les formules utilisées par le WWC pour calculer ces différences en nombre d'écarts-types, ou tailles de l'effet, à la fois pour des mesures continues et dichotomiques.

Tableau II.2. Taille d'effet¹¹ (valeur absolue) initiale

$0.00 \leq \text{Taille d'effet initiale} \leq 0.05$	$0.05 < \text{Taille d'effet initiale} \leq 0.25$	$ \text{Taille d'effet initiale} > 0.25$
Satisfait à l'exigence d'équivalence initiale	Requiert un ajustement statistique pour satisfaire à l'exigence d'équivalence initiale	Ne satisfait pas à l'exigence d'équivalence initiale

Les ajustements statistiques que le WWC considère comme acceptables dépendent de la relation entre la caractéristique initiale et le résultat. En général, lorsque le WWC exige qu'une analyse inclue un ajustement statistique pour une caractéristique initiale spécifiée dans le protocole d'examen, la caractéristique doit être incluse dans l'analyse au niveau du sujet de manière à prendre en compte la corrélation entre la mesure initiale et le résultat. Plusieurs techniques sont acceptables pour répondre à cette exigence, notamment l'ajustement par la régression et l'analyse de la covariance.

Cependant, lorsque la caractéristique initiale est la même que le résultat, d'autres approches qui ne mettent pas en œuvre une estimation de la corrélation, peuvent également être acceptables. Ces méthodes incluent l'utilisation de scores de gain, l'utilisation de l'ajustement faisant intervenir la différence des différences¹² ou l'inclusion d'effets fixes au niveau individuel. Si les auteurs n'effectuent pas d'ajustement par eux-mêmes, le WWC peut effectuer son propre ajustement en faisant intervenir la différence des différences pour permettre à l'étude de satisfaire à l'exigence de mise en œuvre d'un ajustement statistique (voir l'annexe E du *Manuel des procédures WWC*).

.../...

¹¹ Effect Size (ES)

¹² Un ajustement de la différence des différences implique de soustraire la différence initiale à la différence dans les résultats mesurés après l'intervention.

Lorsque le WWC ne requiert pas d'ajustement statistique (parce que l'étude est un ECR à faible attrition ou que les différences initiales sont inférieures ou égales à 0,05 écart-type), les auteurs peuvent ajuster leurs analyses en utilisant des approches autres que celles que le WWC juge acceptables dans le but de satisfaire à l'exigence de mise en œuvre d'un ajustement statistique. En outre, bien que les normes WWC nécessitent des ajustements statistiques dans des circonstances limitées et uniquement pour certaines caractéristiques spécifiées, les auteurs peuvent effectuer un ajustement pour toutes les données initiales disponibles dans leurs analyses.

.../...

Procédure d'examen du WWC pour l'étape 3 de l'examen des études dont l'unité d'affectation est le sujet

- Si l'étude satisfait aux exigences d'équivalence initiale des caractéristiques spécifiées dans le protocole d'examen pour les groupes d'intervention et de comparaison analysés (y compris en appliquant des ajustements statistiques acceptables le cas échéant), elle est éligible pour être classée *conforme avec réserve aux normes WWC*. Pour recevoir ce classement, l'étude doit également satisfaire aux exigences décrites au chapitre IV (non traduit ici), qui stipulent que l'étude doit examiner au moins une mesure de résultat éligible répondant aux exigences de l'examen, et être exempte de facteurs de confusion.
- Si l'étude ne satisfait pas à l'exigence d'équivalence initiale des caractéristiques spécifiées dans le protocole d'examen pour les groupes d'intervention et de comparaison analysés, alors l'étude est classée comme *non conforme aux normes WWC*.

B. L'unité d'affectation est le groupe

Les études de recherche dans lesquelles des sujets (tels que des élèves) sont affectés à la condition d'intervention ou de comparaison en tant que groupes (encore appelés grappes ou *clusters*) sont devenues plus courantes en recherche dans le domaine de l'éducation. Cette affectation de groupes peut prendre différentes formes, notamment des élèves peuvent être regroupés autour d'un ou plusieurs enseignants, les élèves peuvent être regroupés au sein de classes, les élèves peuvent être regroupés au sein d'établissements, les enseignants peuvent être regroupés au sein d'établissements ou les classes peuvent être regroupées au sein d'établissements.

Les études peuvent mettre en œuvre une affectation aléatoire des groupes tout en utilisant des informations récoltées auprès des sujets au sein de ces groupes pour estimer les effets. Dans ces études, les effets observés de l'intervention peuvent être influencés à la fois par les effets de l'intervention sur les sujets mais aussi par la modification éventuelle de la composition des groupes. Par exemple, une intervention très attrayante peut attirer des élèves d'autres classes ou d'autres établissements entre le moment de l'affectation aléatoire et le moment où les résultats sont mesurés. Le WWC examine les études dont l'unité d'affectation est le groupe dans le but de déterminer si les effets observés de l'intervention peuvent être considérés de manière crédible comme étant uniquement dus aux effets de l'intervention sur les sujets, ou si des changements dans la composition des groupes ont

également pu affecter les résultats. Si des modifications de composition ne peuvent pas être exclues, l'étude peut toujours satisfaire aux normes du WWC en ce qui concerne les preuves des effets de l'intervention sur les groupes, mais ne peut pas atteindre le classement WWC le plus élevé¹³.

Parmi les études qui utilisent comme unité d'affectation le groupe, certaines analysent les résultats au niveau du sujet et d'autres analysent les résultats au niveau des groupes (résultats au niveau du sujet qui ont été agrégés au niveau des groupes), mais la distinction entre les effets d'une intervention sur les groupes et ses effets sur les sujets n'est pas fondée sur l'unité d'analyse. Il est possible qu'une analyse des données au niveau des groupes satisfasse aux normes WWC en ce qui concerne la preuve des effets sur les sujets, de même qu'une analyse des données au niveau des sujets satisfasse aux normes WWC en ce qui concerne la preuve des effets sur les groupes.

Cette section présente les critères selon lesquels les estimations des effets présentés par des études dont l'unité d'affectation est le groupe peuvent être classées *conformes sans réserve aux normes WWC*, *conformes avec réserve aux normes WWC* ou *non conformes aux normes WWC*. Le WWC examine d'abord les preuves de l'effet d'une intervention sur les sujets (étapes 1 à 4). Si un effet sur les sujets ne peut pas être démontré de manière crédible, le WWC examine les preuves de l'intervention sur les groupes (étapes 5 à 7), dans le cas où des changements dans la composition des sujets au sein des groupes ont pu influencer l'effet observé. Chaque étape consiste à répondre à une question sur le design de l'étude. La réponse à chaque question mène aux étapes suivantes qui devront être suivies dans le cadre de la procédure d'examen (Figure II.4). Dans les étapes ci-dessous, les évaluations de l'attrition et de l'équivalence initiale utiliseront les mêmes normes que celles décrites ci-dessus à la section A du présent chapitre, à quelques exceptions près.

Les études dont l'unité d'affectation est le groupe et qui satisfont aux exigences décrites dans les étapes 1 à 7 peuvent être classées comme *conforme sans réserve aux normes WWC* ou *conformes avec réserve aux normes WWC*. Pour recevoir ce classement, l'étude doit également satisfaire aux exigences décrites au chapitre IV (non traduit ici), qui stipulent que l'étude doit examiner au moins une mesure de résultat éligible répondant aux exigences de l'examen, et être exempte de facteurs de confusion.

Critères de sélection pour déterminer si l'unité d'affectation de l'étude est le groupe

Une étude doit être examinée à l'aide des normes applicables aux études dont l'unité d'affectation est le groupe lorsqu'elle remplit deux conditions : (1) les sujets sont affectés à la condition d'intervention ou de comparaison en tant que groupes, et (2) les résultats sont mesurés pour les sujets de ces groupes et peuvent être analysés en tant que données au niveau sujet ou en tant que moyennes au niveau des groupes.

¹³ Bien que la mesure des effets d'une intervention sur les établissements ou d'autres groupes puisse répondre à des questions politiques importantes, le WWC se concentre principalement sur la preuve des effets sur les élèves ou sur les résultats concernant le personnel enseignant jugés pertinents pour l'amélioration des résultats des élèves.

Sur la base de ces deux critères, ni la méthode d'estimation de l'effet ni le niveau d'agrégation des données ne permettent de déterminer si l'étude doit être considérée comme une étude dont l'unité d'affectation est le sujet ou est le groupe. Prenons une étude qui assigne de manière aléatoire les établissements à des conditions intervention ou comparaison et telle que le résultat d'intérêt soit la réussite des élèves mais que les données soient agrégées par groupe scolaire (par exemple, les performances moyennes des élèves d'un établissement donnée) aux fins d'analyse. L'étude remplit la première condition car elle attribue une condition (intervention ou comparaison) aux établissements. L'étude remplit la deuxième condition car la mesure des résultats a été effectuée sur des sujets au sein d'établissements scolaires et il importe peu que l'étude agrège les données au niveau de l'établissement. Autrement dit, cette étude sera considérée comme une étude dont l'unité d'affectation est le groupe, même si l'unité d'analyse (données de performance agrégées par établissement) est alignée sur l'unité d'affectation (établissement), car les données agrégées représentent en réalité les résultats mesurés au niveau sujet.

Voici trois autres exemples d'application de ces critères de sélection dans les ECR :

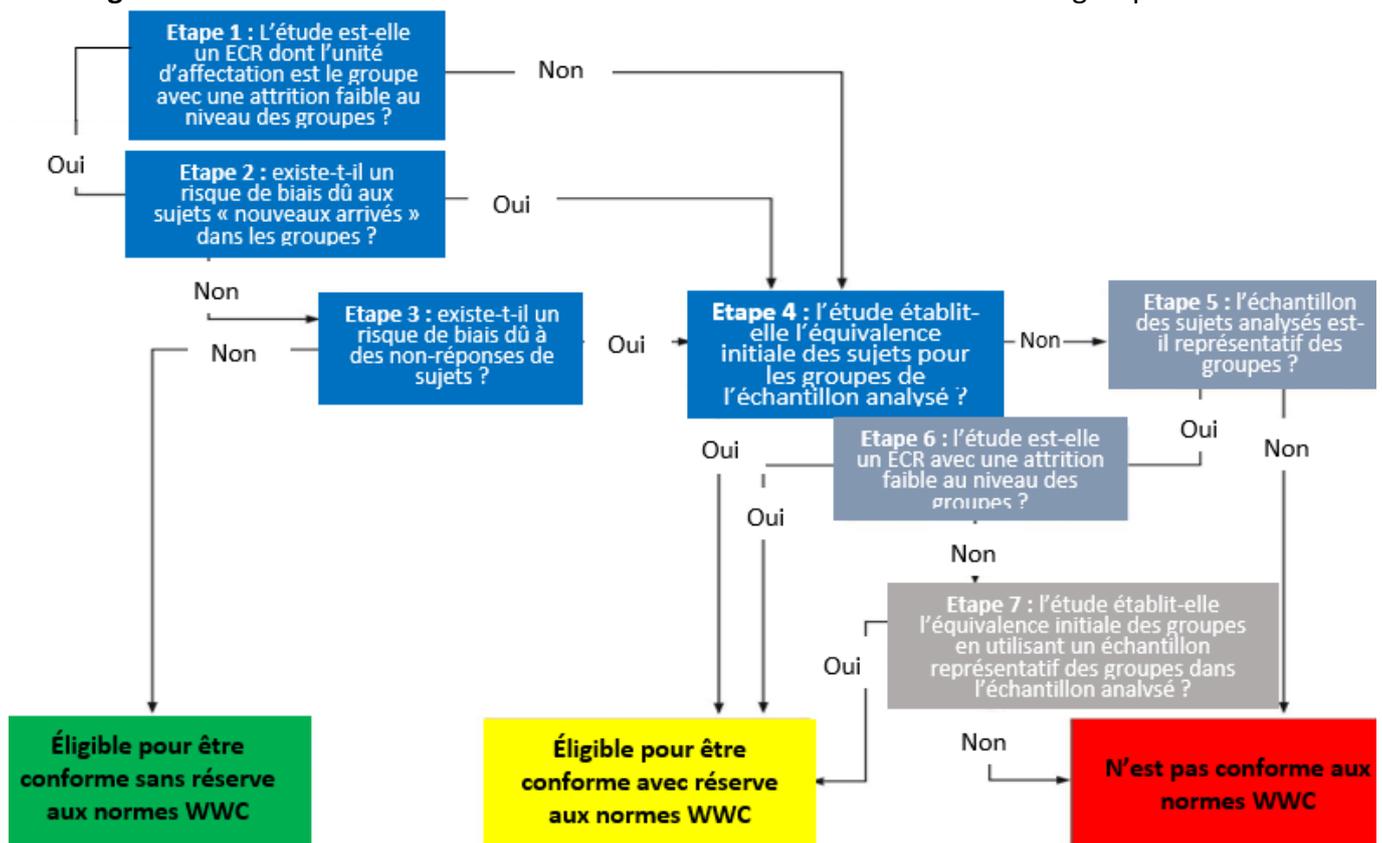
- Si une étude affecte au hasard des enseignants à une condition et que le résultat d'intérêt est un résultat concernant l'élève (tel que sa réussite à des tests), alors l'étude sera considérée comme une étude dont l'unité d'affectation est le groupe. L'unité d'affectation est l'enseignant, et le résultat a été mesuré pour chaque élève, qu'il soit agrégé au niveau de l'enseignant pour analyse ou non.
- Si une étude assigne au hasard des enseignants à une condition et que le résultat d'intérêt est un résultat concernant l'enseignant (tel que le maintien des effectifs), alors l'étude doit être caractérisée comme une étude dont l'unité d'affectation est le sujet. L'unité d'affectation est l'enseignant et le résultat a été mesuré sur les enseignants.
- Si une étude randomise à la fois des groupes et des sujets, elle sera considérée comme une étude qui affecte les sujets. Par exemple, une étude peut randomiser les salles de classe sur une condition donnée, et affecter aléatoirement les élèves dans les classes (ou dans l'ordre inverse). Dans ce cas, l'unité d'affectation est l'élève.

Les deux critères de sélection (les sujets sont affectés à la condition en tant que groupes et les résultats sont mesurés pour les sujets au sein des groupes) s'appliquent également aux EQE, et la description de l'étude peut fournir des indications concernant l'unité d'affectation avérée. Par exemple, « les établissements mettant en place l'intervention ont été comparés à des établissements ne mettant pas en place l'intervention » illustre une situation dans laquelle le groupe (l'établissement) est l'unité d'affectation. Les protocoles d'examen peuvent également clarifier la manière d'identifier l'unité d'affectation dans des scénarios communs à un domaine. Sinon, le WWC identifie la plus grande unité d'étude qui ne contient que les membres soumis à une condition. Par exemple, si une étude examine l'effet d'une intervention sur le rendement des élèves dans un établissement et que chaque classe n'a que des élèves soumis à l'intervention ou n'a que des élèves contrôles, l'unité d'affectation est le groupe. D'un autre côté, si certaines classes ont à la fois des élèves soumis à l'intervention et des élèves non soumis à l'intervention, l'unité d'affectation est alors le sujet. De même, une étude qui a examiné les effets d'un programme de prévention

du décrochage en comparant les taux d'abandon scolaire dans les établissements soumis à l'intervention aux taux des établissements n'ayant pas été soumis à l'intervention, est une étude où l'unité d'affectation est le groupe. L'unité d'affectation est l'établissement et le résultat a été mesuré sur les élèves.

Les résultats d'une étude qui répond à ces deux critères de sélection courent le risque d'être influencés par des changements éventuels dans la composition des sujets au sein des groupes. Ils devraient être considérés comme provenant d'une étude dont l'unité d'affectation est le groupe et être examinés à l'aide des étapes suivantes. Si une étude dont le design est basé sur l'étude de groupes (ECR et EQE) ne répond pas à ces deux critères, elle doit être considérée comme une étude d'affectation au niveau sujet.

Figure II.4. Procédure d'examen des études dont l'unité d'affectation est le groupe



Note : Pour être classée *conforme sans réserve aux normes WWC* ou *conforme avec réserve aux normes WWC*, l'étude doit également satisfaire aux exigences décrites au chapitre IV (non traduit ici), qui stipulent que l'étude doit examiner au moins une mesure de résultat éligible répondant aux exigences de l'examen, et être exempte de facteurs de confusion.

Procédure d'examen de l'effet probant d'une intervention sur des sujets (étapes 1 à 4)

Les quatre étapes suivantes décrivent la procédure d'examen visant à évaluer la crédibilité des preuves concernant une étude qui cherche à évaluer les effets d'une intervention sur

des sujets. Pour pouvoir être classée *conforme sans réserve aux normes WWC*, un ECR dont l'unité d'affectation est le groupe doit limiter les biais potentiels résultant de changements éventuels dans la composition des groupes entre eux ou de changements au sein des groupes après l'affectation aléatoire. Les ECR présentant un risque de biais dû à ces changements de composition et toutes les EQE dont l'unité d'affectation est le groupe peuvent toujours être éligibles pour être classées *conforme avec réserve aux normes WWC* pour satisfaire aux normes WWC en ce qui concerne la preuve des effets de l'intervention sur les sujets si l'étude satisfait à l'exigence de l'équivalence initiale des sujets des groupes d'intervention et de comparaison analysés. Une étude peut satisfaire aux normes WWC en ce qui concerne les effets sur les sujets quel que soit le niveau d'agrégation des données utilisées dans l'analyse. En particulier, les résultats basés sur une analyse des données sur les performances des élèves agrégées au niveau de l'établissement peuvent satisfaire aux normes WWC en ce qui concerne la preuve des effets d'une intervention sur les élèves si elle satisfait aux exigences des étapes 1 à 4.

Étape 1. L'étude est-elle un ECR dont l'unité d'affectation est le groupe avec une attrition faible au niveau du groupe ?

Afin d'obtenir la classification la plus élevée, l'étude doit être un ECR dont l'unité d'affectation à une condition est le groupe et présente un taux d'attrition faible au niveau des groupes, défini par les limites spécifiées dans le protocole d'examen applicable et représenté figure II.2. L'attrition au niveau des groupes mesure la perte de groupes entiers de l'échantillon randomisé. Un groupe est perdu lorsqu'il ne fournit aucune donnée de résultat au moment de l'analyse. La perte de sujets au sein des groupes est évaluée dans une étape distincte ci-dessous (étape 3).

Procédure d'examen du WWC pour l'étape 1 de l'examen des études dont l'unité d'affectation est le groupe

- Si l'étude est un ECR avec une attrition faible au niveau des groupes, passer à l'étape 2.
- Si l'étude est un ECR avec une attrition élevée ou inconnue au niveau des groupes, un ECR compromise ou une EQE, passer à l'étape 4.

Étape 2. Existe-t-il un risque de biais dû aux sujets « nouveaux arrivés » dans les groupes ?

Afin de recevoir la classification la plus élevée, un ECR dont l'unité d'affectation est le groupe doit limiter le risque de biais dû aux sujets nouveaux arrivés (*joiners*) dans le groupe après le moment de l'assignation aléatoire. Si l'étude inclut des nouveaux arrivés dans l'échantillon analysé, l'estimation de l'effet de l'intervention sur les résultats au niveau des sujets pourrait être biaisée si les sujets ayant intégré des groupes d'intervention diffèrent systématiquement de ceux qui ont intégré des groupes de comparaison. Ce risque de biais peut varier selon les domaines d'activité et les interventions, et également en fonction du temps écoulé depuis l'assignation aléatoire. Par conséquent, le protocole d'examen identifiera les nouveaux arrivés qui présenteraient un risque de biais s'ils étaient inclus dans

l'échantillon analysé dans un ECR de ce type (le WWC ne considère jamais les nouveaux arrivés comme un risque de biais lorsqu'ils sont exclus de l'échantillon analysé), en fonction du moment où ils ont rejoint les groupes, des caractéristiques de l'intervention et de l'unité d'affectation. La procédure doit être spécifiée dans le protocole d'examen et appliquée de manière constante à toutes les études incluses dans l'examen.

Certains nouveaux arrivés peuvent intégrer des groupes après une affectation aléatoire, mais avant que les sujets n'aient connaissance des conditions attribuées de manière aléatoire aux groupes. Le WWC ne considère pas que ces nouveaux arrivés puissent introduire un biais, car les décisions qui ont conduit ces sujets à rejoindre des groupes n'ont pas pu être affectées par l'intervention. La charge de démontrer que des personnes ne pouvaient pas avoir connaissance de l'intervention incombe aux auteurs de l'étude. Par exemple, une assignation aléatoire d'établissements peut avoir lieu au cours de l'été précédant le début d'une année scolaire, mais l'intervention n'a été annoncée aux familles qu'après le début de l'année scolaire. Les élèves qui ont rejoint les établissements avant l'annonce ne peuvent pas introduire de biais. Les élèves qui ont rejoint les établissements après l'annonce pourraient introduire un biais.

.../...

Le protocole d'examen peut sélectionner différentes options pour lesquelles les nouveaux arrivés sont susceptibles d'introduire un biais pour différents groupes d'interventions et pour différentes unités d'affectation. Par exemple, le protocole d'examen peut indiquer que les nouveaux arrivés n'introduisent pas de biais lorsque l'unité d'affectation est l'établissement, mais que tous les nouveaux arrivés peuvent introduire un biais lorsque l'unité d'affectation est la classe, l'enseignant ou une unité plus petite.

.../...

Une étude qui exclut tous les nouveaux arrivés de l'échantillon analysé ou n'inclut que les nouveaux arrivés qui ne présentent pas de risque d'introduire un biais est réputée limiter le risque de biais dû à l'arrivée de nouveaux sujets et est éligible pour être classée *conforme sans réserve aux normes WWC* si l'étude possède également de faibles taux de non-réponse au niveau des sujets (étape 3). Toutefois, si une étude inclut de nouveaux arrivés dans l'échantillon analysé qui présentent un risque de biais conformément au protocole d'examen, alors le classement le plus élevé pouvant être attribué à cette étude correspond à *conforme avec réserve aux normes WWC*, et pour obtenir ce classement, l'étude doit satisfaire à l'exigence d'équivalence initiale sur les caractéristiques spécifiées dans le protocole d'examen pour les sujets des groupes d'intervention et de comparaison analysés (étape 4).

Procédure d'examen du WWC pour l'étape 2 de l'examen des études dont l'unité d'affectation est le groupe

- Si l'étude soit (1) exclut tous les nouveaux arrivés de l'échantillon analysé soit (2) inclut des nouveaux arrivés dans l'échantillon analysé qui ne présentent pas de risque de biais conformément au protocole d'examen, alors l'étude limite le risque de biais. Passer à l'étape 3.
- Si l'échantillon analysé de l'étude inclus des nouveaux arrivés qui ont été intégrés après que les résultats de l'affectation aléatoire aient été connus et qui, conformément au protocole d'examen, présentent un risque de biais, passer à l'étape 4.

Étape 3. Existe-t-il un risque de biais dû à des non-réponse de sujets ?

Afin de recevoir la note la plus élevée, un ECR dont l'unité d'affectation est le groupe et qui présente un risque limité de biais dû aux nouveaux arrivés doit comporter un taux faible de non-réponse au niveau des sujets ainsi qu'un faible niveau d'attrition au niveau des groupes (évalué ci-dessus, à l'étape 1). La non-réponse au niveau des sujets pour les études dont l'unité d'affectation est le groupe correspond à la différence entre les sujets présents dans un échantillon de référence (décrit ci-dessous) et ceux présents dans l'échantillon analysé au moment de l'évaluation du résultat. Pour les études qui analysent les résultats agrégés au niveau du groupe, les sujets présents dans l'échantillon analysé sont ceux qui contribuent à la mesure des résultats. L'échantillon de référence - l'échantillon de référence à partir duquel la non-réponse est mesurée - peut différer en fonction du risque de biais associé aux nouveaux arrivés. Lorsque l'échantillon de référence est l'échantillon randomisé d'origine, cette étape mesure l'attrition au niveau sujet. Étant donné que l'échantillon de référence peut ne pas être l'échantillon randomisé, le WWC considère cette étape comme une mesure de la non-réponse au niveau sujet.

La non-réponse au niveau sujet est toujours mesurée dans l'échantillon de groupes non perdus. Les sujets des groupes non représentés dans l'échantillon analysé ne contribuent pas à l'échantillon de référence utilisé dans le dénominateur du calcul du taux de non-réponse au niveau sujet.

.../...

Procédure d'examen du WWC pour l'étape 3 de l'examen des études dont l'unité d'affectation est le groupe

- Si l'étude présente de faibles niveaux de non-réponse au niveau des sujets, elle est alors éligible pour être classée *conforme sans réserve aux normes WWC*. Pour recevoir ce classement, l'étude doit également satisfaire aux exigences décrites au chapitre IV (non traduit ici), qui stipulent que l'étude doit examiner au moins une mesure de résultat éligible répondant aux exigences de l'examen, et être exempte de facteurs de confusion.
- Si l'étude présente des niveaux élevés de non-réponse au niveau sujet, passer à l'étape 4.

Étape 4. L'étude établit-elle l'équivalence initiale des sujets pour les groupes de l'échantillon analysé ?

Les ECR dont l'unité d'affectation est le groupe qui présentent un risque élevé de biais (attrition des groupes, nouveaux arrivés intégrés à l'échantillon analysé ou non-réponse au niveau des sujets) et les EQE dont l'unité d'affectation est le groupe, qui satisfont à l'exigence d'équivalence initiale pour l'échantillon des sujets analysé, sont éligibles au classement *conforme avec réserve aux normes WWC*. Les sujets des groupes d'intervention et de comparaison analysés doivent satisfaire aux mêmes exigences que celles qui ont été spécifiées pour l'examen des études avec affectation au niveau des sujets décrites à l'étape 3 de la section A. Pour les études analysant les résultats agrégés au niveau du groupe, les sujets dont les données contribuent à la mesure des résultats doivent satisfaire à cette exigence. Quel que soit le niveau d'analyse, cette exigence d'équivalence initiale doit être satisfaite à l'aide d'écart-types calculés au niveau des sujets. Les moyennes calculées à l'aide de données au niveau du groupe ou de données au niveau des sujets sont acceptables tant que la pondération est cohérente avec la pondération utilisée dans l'analyse. En règle générale, un ajustement statistique requis doit être effectué à l'aide de données au niveau des sujets, de manière à prendre en compte la corrélation au niveau sujet entre la mesure initiale et le résultat.

.../...

Procédure d'examen du WWC pour l'étape 4 de l'examen des études dont l'unité d'affectation est le groupe

- Si l'étude satisfait à l'exigence d'équivalence initiale pour l'échantillon des sujets analysé en ce qui concerne les caractéristiques spécifiées dans le protocole d'examen, elle est alors éligible pour être classée *conforme avec réserve aux normes WWC*. Pour recevoir ce classement, l'étude doit également satisfaire aux exigences décrites au chapitre IV (non traduit ici), qui stipulent que l'étude doit examiner au moins une mesure de résultat éligible répondant aux exigences de l'examen, et être exempte de facteurs de confusion.
- Si l'étude ne satisfait pas à l'exigence d'équivalence initiale pour l'échantillon analysé, elle sera examinée afin de déterminer si elle peut satisfaire aux normes WWC qui concernent l'effet de l'intervention sur les groupes. Cette procédure d'examen est décrite dans la section suivante, en commençant par l'étape 5.

Procédure d'examen de l'effet probant d'une intervention sur des groupes (étapes 5 à 7)

Les trois étapes suivantes décrivent la procédure d'examen visant à évaluer la crédibilité des données probantes pour une étude visant à évaluer les effets d'une intervention sur des groupes. Dans ces études, l'estimation de l'effet observé représente potentiellement une combinaison de (1) l'effet de l'intervention sur des sujets et (2) d'un effet composite lié aux sujets des groupes d'intervention et des groupes de comparaison. Par conséquent, les éléments de preuve examinés dans cette section ne sont éligibles que pour être classés *conforme avec réserve aux normes WWC*. Pour recevoir ce classement, une étude qui n'a pas

été classée comme répondant sans ou avec réserve aux normes WWC au moment de son examen pour évaluer si elle satisfaisait aux normes WWC relatives aux effets probants sur les sujets, doit analyser les sujets qui sont représentatifs des groupes de l'échantillon analysé et également, soit être un ECR ayant comme unité d'affectation le groupe et un taux d'attrition faible au niveau des groupes, soit satisfaire à l'exigence d'équivalence initiale des groupes dans les groupes d'intervention et de comparaison analysés. Une étude peut satisfaire aux normes WWC en ce qui concerne les effets sur les groupes, quel que soit le niveau d'agrégation des données utilisé dans l'analyse. En particulier, une étude dont les résultats sont basés sur une analyse des données concernant les performances des élèves peut être conforme aux normes du WWC en ce qui concerne la preuve des effets d'une intervention sur les groupes si elle satisfait aux exigences des étapes 5 à 7.

Étape 5. L'échantillon des sujets analysés est-il représentatif des groupes ?

Si une étude présente de faibles taux de réponse en fin d'intervention ou des taux de réponse très différents entre les sujets des groupes d'intervention et de comparaison, l'effet observé ne permet pas d'estimer de manière crédible l'effet de l'intervention sur les groupes. Par conséquent, le WWC évalue dans quelle mesure les sujets au sein des groupes inclus dans l'échantillon analysé sont représentatifs de tous les sujets présents dans les groupes lors de l'intervention.

Le WWC évaluera la représentativité en utilisant les limites d'attrition spécifiées dans le protocole d'examen applicable et illustrées figure II.2 et, comme pour le calcul de la non-réponse au niveau des sujets, seuls les sujets des groupes non-perdus sont comptabilisés. Le numérateur pour ce calcul d'attrition sera le nombre de sujets présents dans les groupes non perdus durant le suivi (c'est-à-dire au moment approximatif où les résultats ont été mesurés) qui ne contribuent pas à l'estimation de l'effet (non inclus dans l'échantillon analysé). Le dénominateur pour le calcul de l'attrition sera le nombre total de sujets dans les groupes non-perdus lors du suivi. Contrairement à la mesure de la non-réponse au niveau sujet à l'étape 3, l'échantillon de référence du dénominateur permettant de mesurer la représentativité est toujours considéré lors du suivi. Cependant, le moment où s'effectue le décompte de l'échantillon de référence ne doit pas nécessairement être le même que celui correspondant à la mesure des résultats. Par exemple, dans une étude d'affectation au niveau de l'établissement qui a mesuré les résultats à la fin de l'année scolaire, l'échantillon de référence pourrait être le nombre d'inscriptions administratives enregistré à un moment donné de l'année scolaire.

Pour les études qui analysent les résultats agrégés au niveau du groupe, les sujets présents dans l'échantillon analysé sont ceux qui contribuent à la mesure des résultats. L'exigence de représentativité est évaluée en utilisant le nombre de sujets regroupés dans tous les groupes du groupe d'intervention ou du groupe de comparaison (et non pour chaque groupe individuellement).

Procédure d'examen du WWC pour l'étape 5 de l'examen des études dont l'unité d'affectation est le groupe

- Si l'étude présente un faible taux de non-réponse au niveau sujet (représentant l'évaluation de la représentativité), passer à l'étape 6.
- Si l'étude présente un taux de non-réponse au niveau sujet élevé ou inconnu, elle est alors classée comme *non conforme aux normes WWC*.

Étape 6. L'étude est-elle un ECR avec une attrition faible au niveau des groupes ?

Il s'agit de la même évaluation de l'attrition au niveau des groupes que celle effectuée à l'étape 1. Cette étape est répétée car il est possible que les ECR avec une attrition faible au niveau des groupes, les ECR avec une attrition élevée au niveau des groupes et les EQE atteignent l'étape 6.

Procédure d'examen du WWC pour l'étape 6 de l'examen des études dont l'unité d'affectation est le groupe

- Si l'étude est un ECR avec un taux d'attrition de groupe faible, elle est alors éligible pour être classée *conforme avec réserve aux normes WWC*. Pour recevoir ce classement, l'étude doit également satisfaire aux exigences décrites au chapitre IV (non traduit ici), qui stipulent que l'étude doit examiner au moins une mesure de résultat éligible répondant aux exigences de l'examen, et être exempte de facteurs de confusion.
- Si l'étude est un ECR avec une attrition des groupes importante ou inconnue ou une EQE, passer à l'étape 7.

Étape 7. L'étude établit-elle l'équivalence initiale des groupes de l'échantillon analysé ?

Parmi les études qui n'ont pas reçu le classement conforme avec ou sans réserve aux normes WWC lors de l'examen afin de satisfaire aux normes WWC concernant les effets sur les sujets, les ECR dont l'unité d'affectation est le groupe avec une érosion au niveau des groupes élevée ou inconnue et les EQE analysant des groupe doivent satisfaire à l'exigence d'équivalence initiale pour l'échantillon analysé constitué de l'ensemble des groupes d'intervention et des groupes de comparaison pour que l'étude soit éligible pour être classée *conforme avec réserve aux normes WWC*. L'échantillon analysé des groupes est constitué des groupes représentés dans l'échantillon et utilisés pour estimer les résultats.

Les caractéristiques pour lesquelles le WWC doit évaluer l'équivalence initiale des groupes sont spécifiées dans le protocole d'examen et peuvent différer de celles utilisées pour évaluer l'équivalence initiale des sujets. Les exemples de caractéristiques comprennent les niveaux de réussite des élèves, les niveaux d'étude, les caractéristiques démographiques des enseignants ou des élèves dans les établissements et le terrain scolaire (urbain ou rural). Le

protocole d'examen précisera également si les sujets fournissant des données de base dans les groupes utilisées pour évaluer l'équivalence initiale des groupes doivent être les mêmes personnes qui fournissent des données pour obtenir les résultats de l'analyse.

.../...

Quel que soit le niveau d'analyse, l'exigence d'équivalence initiale pour les groupes peut être satisfaite à l'aide de moyennes individuelles ou de groupes, et d'écart-types au niveau sujet ou au niveau des groupes (et ce quelle que soit la combinaison), à condition que la pondération des moyennes soit cohérente avec la pondération utilisée dans l'analyse. Le WWC utilisera des écart-types sujets lorsque cela est possible. Tout ajustement statistique requis doit être effectué à l'aide de données au même niveau que celles utilisées pour évaluer l'équivalence initiale.

De plus, dans le cadre de l'exigence de l'équivalence initiale concernant les groupes de l'échantillon analysé, les sujets dont on utilise les données initiales doivent être représentatifs des groupes contribuant à l'analyse de l'effet, ce qui est évalué en comparant le nombre de sujets contribuant aux données initiales par rapport au nombre de sujets présents dans les groupes au moment de l'évaluation de l'équivalence initiale. Cette évaluation de la représentativité des données initiales est analogue à celle des données de suivi décrite à l'étape 5 et elle est effectuée en utilisant les mêmes seuils d'attrition que ceux utilisés pour des études d'affectation au niveau sujet de la section A. Par exemple, si une étude dont l'unité d'affectation est l'établissement mesure les résultats des élèves de CM1 en 2015 et utilise les élèves de CM1 des mêmes établissements en 2014 pour évaluer l'équivalence initiale, la représentativité initiale serait évaluée en comparant le nombre d'élèves de CM1 inscrits dans les établissements en 2014 qui n'avaient pas contribué aux données initiales par rapport au nombre total d'élèves inscrits (ceux qui ont fourni ou non des données). La taille des échantillons de référence ne doit pas nécessairement être mesurée au moment de la collecte de la mesure de référence. Par exemple, dans une étude dont l'unité d'affectation est l'établissement et qui a recueilli des données initiales à la fin d'une année scolaire et mesuré des résultats durant l'année scolaire suivante, l'échantillon de référence pourrait être l'effectif administratif enregistré à un moment donné de l'année scolaire au cours de laquelle les données initiales ont été collectées.

Procédure d'examen du WWC pour l'étape 7 de l'examen des études dont l'unité d'affectation est le groupe

- Si l'étude satisfait à l'exigence d'équivalence initiale des groupes analysés, ce qui sous-entend que les données initiales sont représentatives, elle est éligible pour être classée *conforme avec réserve aux normes WWC*. Pour recevoir ce classement, l'étude doit également satisfaire aux exigences décrites au chapitre IV (non traduit ici), qui stipulent que l'étude doit examiner au moins une mesure de résultat éligible répondant aux exigences de l'examen, et être exempte de facteurs de confusion.
- Si l'étude ne satisfait pas à l'exigence d'équivalence initiale des groupes analysés, elle est alors classée comme *non conforme aux normes WWC*.

.../...

Liste des abréviations

ECR : essai contrôlé randomisé (*randomized controled trials, RCT*)

ECU : étude de cas unique (*single-case design, SCD*)

EQE : étude quasi-expérimentale (*quasi-experimental design, QED*)

ERD : études de régression par discontinuité (*regression discontinuity design, RDD*)

ES : *Effect Size* (taille d'effet)

WWC : *What Works Clearinghouse*