



Commission
européenne

CONCEVOIR ET COMMANDER DES ÉVALUATIONS D'IMPACT CONTREFACTUELLES

**GUIDE PRATIQUE À L'INTENTION DES AUTORITÉS
DE GESTION DU FSE**

COMMISSION EUROPÉENNE

Direction générale de l'emploi, des affaires sociales et de l'inclusion
Direction G — Fonds, programmation et mise en œuvre
Unité G5 — Mieux légiférer

Contact: Linda Adamaite

Courriel: EMPL-G5-UNIT@ec.europa.eu ou Linda.Adamaite@ec.europa.eu

Commission européenne B-1049 Bruxelles

CONCEVOIR ET COMMANDER DES ÉVALUATIONS D'IMPACT CONTREFACTUELLES

GUIDE PRATIQUE À L'INTENTION DES AUTORITÉS DE
GESTION DU FSE

Manuscrit achevé en septembre 2021

L'édition 2021 de « Concevoir et commander des évaluations d'impact contrefactuelles – Guide pratique à l'intention des autorités de gestion du FSE » a été rédigée par Jochen Kluve (Humboldt-Universität zu Berlin), Andrea Naldini et Marco Pompili (Ismeri Europa). Il s'agit d'une adaptation de l'édition 2013 des orientations (ISBN 978-92-79-28238-6; DOI 10.2767/94454) préparée par Stephen Morris (unité d'évaluation et de recherche des politiques, Université métropolitaine de Manchester), Herta Tödting-Schönhofer (Metis GmbH, Vienne) et Michael Wiseman (Institut George Washington de politique publique).

Le présent document a été élaboré pour la Commission européenne, mais ne reflète que le point de vue de ses auteurs. La Commission ne peut être tenue responsable des conséquences de la réutilisation qui pourrait en être faite. De plus amples informations sur l'Union européenne sont disponibles sur internet (<http://www.europa.eu>).

Luxembourg: Office des publications de l'Union européenne, 2022

© Union européenne, 2022



La politique de réutilisation des documents de la Commission européenne est mise en œuvre sur la base de la décision 2011/833/UE de la Commission du 12 décembre 2011 relative à la réutilisation des documents de la Commission (JO L 330 du 14.12.2011, p. 39). Sauf indication contraire, la réutilisation de ce document fait l'objet d'une licence sous Creative Commons Attribution 4.0 International (CC-BY 4.0) (<https://creativecommons.org/licenses/by/4.0/>). Cela signifie que la réutilisation est autorisée moyennant citation appropriée de la source et indication de toute modification.

Pour toute utilisation ou reproduction d'éléments qui ne sont pas la propriété de l'Union européenne, l'autorisation peut devoir être obtenue directement auprès des titulaires de droits respectifs.

Couverture: © Shutterstock, 2021

Version PDF

ISBN 978-92-76-40728-7

doi: 10.2767/917059

KE-02-21-992-FR-N

Table des matières

INTRODUCTION: CONTEXTE ET OBJET DU GUIDE	5
CHAPITRE 1. CONCEPT ET APPROCHES	11
1.1. L'ESSENCE DE LA MÉTHODE CONTREFACTUELLE	11
1.2. POURQUOI LES ÉVALUATIONS CONTREFACTUELLES SONT-ELLES IMPORTANTES?	12
1.3. POURQUOI LES ÉVALUATIONS CONTREFACTUELLES SONT-ELLES TECHNIQUEMENT DIFFICILES?	13
1.4. APERÇU DES CONCEPTIONS ET DES APPROCHES DES EIC	15
1.5. COMMENT INTÉGRER UNE EIC DANS UN CADRE D'ÉVALUATION PLUS LARGE?	16
CHAPITRE 2. CONSIDÉRATIONS PRATIQUES POUR L'ÉLABORATION D'UNE EIC	23
2.1. SÉLECTIONNER LES INTERVENTIONS POUR L'ÉVALUATION D'IMPACT	27
2.1.1. Hiérarchisation des interventions pour l'évaluation d'impact.....	28
2.1.2. Sélectionner des interventions qui sont susceptibles d'être soumises à une méthode contrefactuelle	30
2.2. QUESTIONS D'ÉVALUATION ET VARIABLES DE RÉSULTAT	35
2.2.1. Quels sont les buts et les objectifs de l'intervention?	35
2.2.2. Quel est l'objectif de l'évaluation?	35
2.3. DONNÉES PERMETTANT D'IDENTIFIER LE GROUPE TÉMOIN ET DE MESURER LES VARIABLES DE RÉSULTAT.....	40
2.3.1. Les données adéquates sont-elles disponibles ou peuvent-elles être mises à disposition?	40
2.3.2. Comment le groupe «traité» est-il défini?.....	53
2.3.3. Facteurs à prendre en considération lors de l'identification d'un groupe témoin	55
2.3.4. Quels genres de considérations en matière de données interviennent dans le système d'évaluation?	59
2.3.5. Quelles sont les principales contraintes de l'analyse des données et des résultats?.....	60
2.3.6. Liste de contrôle afin de vérifier la préparation et la faisabilité de l'EIC.....	63
2.4. MÉTHODE D'EIC À APPLIQUER	64
2.5. CALENDRIER ET BUDGET.....	65
2.5.1. Quelles sont les ressources disponibles?	65
2.5.2. À quel moment l'intervention doit-elle être évaluée?.....	69
2.6. MISE EN ŒUVRE DE L'EIC.....	72
2.6.1. Sélection de l'évaluateur.....	72
2.6.2. Supervision de l'EIC	74
2.6.3. Établissement de rapports.....	74
2.6.4. Exploitation des résultats.....	76
CHAPITRE 3. COMMENT SÉLECTIONNER LA MÉTHODE APPROPRIÉE POUR RÉALISER UNE EIC	79
3.1. RANDOMISATION – L'APPROCHE EXPÉRIMENTALE	79
3.2. APPROCHES NON ALÉATOIRES OU QUASI EXPÉRIMENTALES	83
3.2.1. Groupe cible et groupe témoin sans répartition aléatoire.....	83
3.2.2. Appariement par score de propension (propensity score matching).....	84
3.2.3. Double différence.....	87
3.2.4. Approche de la régression par discontinuité	90
3.2.5. Variables instrumentales	94
CHAPITRE 4. FAIRE ÉVOLUER LE PROGRAMME D'EIC	100
4.1. AMÉLIORER LES NIVEAUX DE COMPRÉHENSION DES PARTIES PRENANTES.....	100

4.2. DÉVELOPPEMENT DES CAPACITÉS.....	101
4.3. S'ATTAQUER AUX OBSTACLES JURIDIQUES.....	104
4.4. VERS DES APPROCHES PLUS ORIENTÉES VERS L'AVENIR	104
4.5. ÉLARGISSEMENT DU CHAMP D'APPLICATION DES EIC	105
GLOSSAIRES	110
ACRONYMES.....	110
DÉFINITIONS	111
LISTE DES ANNEXES.....	120
ANNEXE 1. LECTURES COMPLÉMENTAIRES	120
ANNEXE 2. SUGGESTIONS DE COURS RELATIFS AUX EIC EN LIGNE	124
ANNEXE 3. ÉVALUATION D'IMPACT CONTREFACTUELLE - EXEMPLES MENTIONNÉS DANS LE GUIDE.....	125

LISTE DES ENCADRÉS

ENCADRÉ 1 EXEMPLE D'UNE EIC VISANT À COMPARER LES AVANTAGES ET LES COÛTS.....	21
ENCADRÉ 2 EIC INTÉGRÉE DANS UN CADRE PLUS LARGE	25
ENCADRÉ 3 QUESTIONS POSÉES LORS DE LA SÉLECTION DES INTERVENTIONS POUR UNE EIC	27
ENCADRÉ 4 TYPES D'INTERVENTIONS ET DE GROUPES CIBLES LES PLUS COURANTS CHOISIS POUR LES EIC DU FSE	30
ENCADRÉ 5 DÉFINIR LES GROUPES TÉMOINS.....	34
ENCADRÉ 6 EXEMPLES DE DONNÉES UTILISÉES POUR LES EIC	42
ENCADRÉ 7 EXEMPLES DE BASES DE DONNÉES INTÉGRÉES POUR LES EIC	43
ENCADRÉ 8 CADRE RÉGLEMENTAIRE DE L'UE RELATIF AU TRAITEMENT DES DONNÉES À CARACTÈRE PERSONNEL	47
ENCADRÉ 9 PROTECTION ET ÉCHANGE DES DONNÉES	52
ENCADRÉ 10 QUESTIONS STRATÉGIQUES LIÉES À UN PROGRAMME DE FORMATION	57
ENCADRÉ 11 INTERPRÉTATION DES EFFETS FINAUX.....	58
ENCADRÉ 12 INCERTITUDES CONCERNANT L'INTERPRÉTATION DES RÉSULTATS	62
ENCADRÉ 13 L'EXPÉRIENCE POLONAISE CONCERNANT LES CONFÉRENCES D'ÉVALUATION	78
ENCADRÉ 14 EXEMPLE D'ESSAI ALÉATOIRE MENÉ SUR UN PROJET DU FSE À DESTINATION DES JEUNES.....	81
ENCADRÉ 15 EXEMPLE D'UNE ÉVALUATION PAR L'APPROCHE D'APPARIEMENT.....	86
ENCADRÉ 16 EXEMPLE D'UNE ÉVALUATION PAR L'APPROCHE DE LA DOUBLE DIFFÉRENCE.....	89
ENCADRÉ 17 EXEMPLE D'UNE ÉVALUATION PAR L'APPROCHE DE DISCONTINUITÉ DE LA RÉGRESSION.....	93
ENCADRÉ 18 EXEMPLE D'UNE ÉTUDE RÉALISÉE SELON L'APPROCHE PAR VARIABLES INSTRUMENTALES.....	96
ENCADRÉ 19 EXEMPLE DE PROJET VISANT À RENFORCER LA CULTURE ET LES CAPACITÉS EN MATIÈRE D'EIC.....	103
ENCADRÉ 20 EXEMPLES D'ÉVALUATIONS DANS LE DOMAINE DE L'ÉDUCATION.....	106
ENCADRÉ 21 EXEMPLE D'ÉVALUATION DES EFFETS SUR LES «RÉSULTATS DIFFICILEMENT QUANTIFIABLES» EN ALLEMAGNE.....	108

LISTE DES FIGURES

FIGURE 1 DIFFÉRENTES TÂCHES ET TYPES D'ÉVALUATION.....	17
FIGURE 2 ILLUSTRATION DE L'APPROCHE DU MODÈLE LOGIQUE OU DE LA «THÉORIE DU CHANGEMENT»	18
FIGURE 3 PRINCIPALE SÉQUENCE D'ACTIVITÉS D'UNE EIC.....	24
FIGURE 4 AMPLEUR DES EFFETS MINIMAUX DÉTECTABLES POUR DES TAILLES D'ÉCHANTILLONS DIFFÉRENTES	61
FIGURE 5 CHRONOLOGIE SIMPLIFIÉE DES RÉSULTATS D'UN PROGRAMME DE FORMATION	71
FIGURE 6 CONTRÔLE ALÉATOIRE EN DEUX GROUPES.....	81
FIGURE 7 APPROCHE QUASI EXPÉRIMENTALE STYLISÉE AVEC GROUPE TRAITÉ ET GROUPE TÉMOIN	84
FIGURE 8 ILLUSTRATION DE L'APPROCHE PAR SCORE DE PROPENSION.....	86
FIGURE 9 ILLUSTRATION DE L'APPROCHE DES DOUBLES DIFFÉRENCES.....	89
FIGURE 10 ILLUSTRATION DE L'APPROCHE DE DISCONTINUITÉ DE LA RÉGRESSION.....	91
FIGURE 11 ILLUSTRATION D'UNE APPROCHE PAR VARIABLES INSTRUMENTALES.....	94

LISTE DES TABLEAUX

TABLEAU 1 CONTENU RECOMMANDÉ D'UN SYSTÈME D'ÉVALUATION.....	25
TABLEAU 2 TYPES DE DONNÉES ET SOURCES	42
TABLEAU 3 STRUCTURE DES PRINCIPAUX COÛTS D'UNE EIC	69
TABLEAU 4 INFORMATIONS DE BASE À INCLURE DANS UNE FICHE PRÉSENTANT L'EIC.....	76
TABLEAU 5 COMPARAISON DES CARACTÉRISTIQUES ESSENTIELLES DES PRINCIPALES APPROCHES D'EIC	97
TABLEAU 6 CARACTÉRISTIQUES DES EIC MENTIONNÉES DANS LE GUIDE COMME EXEMPLES	125

Introduction: contexte et objet du guide

La période de programmation 2021-2027 commence par l'expérience dramatique de la pandémie de COVID-19 et de la crise économique qui a suivi. L'Union européenne (UE) a intensifié ses efforts financiers et politiques pour aider à la relance des économies nationales et accroître l'emploi. Dans ce cadre, le Fonds social européen plus (FSE+) joue un rôle essentiel dans la fourniture d'une aide généralisée aux chômeurs, en mettant l'accent sur les jeunes et les femmes, ainsi que dans le soutien aux interventions contre la pauvreté des enfants, en promouvant une meilleure éducation et l'inclusion sociale des groupes sociaux les plus faibles dans l'ensemble de l'UE. La diversité des objectifs poursuivis par le FSE+ et la nécessité d'obtenir des résultats rapides dans les domaines de l'emploi et de l'inclusion sociale exigent une allocation efficace des ressources. Les approches fondées sur des données probantes en matière d'élaboration des politiques sont de plus en plus importantes et l'évaluation est un instrument fondamental pour orienter les politiques publiques.

Guide à l'intention des autorités de gestion mettant à jour l'expérience de la période 2014-2020

Au cours de la période 2014-2020, la Commission européenne (CE) a soutenu la capacité d'évaluation des États membres (EM) et des autorités de gestion (AG) et a encouragé le recours à l'évaluation d'impact contrefactuelle (EIC). Les résultats positifs constatés dans l'EIC sont des preuves tangibles de ce qui a été réalisé par le FSE et de ce qui aurait été accompli sans son soutien. Un nombre important d'EIC ont d'abord été programmées dans les plans d'évaluation (PE), puis réalisées au cours de la mise en œuvre des programmes du FSE. Dans de nombreux cas, les EIC ont rencontré des difficultés au cours de leur préparation et de leur mise en œuvre, ou sont restées des expériences isolées et n'ont pas été incluses dans un cadre d'évaluation systématique.

Le présent guide s'adresse aux autorités de gestion et aux autres organismes chargés de la mise en œuvre des interventions et programmes financés par le FSE+ et vise à aider à la planification, à la conception et à la commande des EIC. Il prend en considération les expériences passées et fournit des conseils pratiques sur certaines des questions clés à prendre en considération lors de l'élaboration d'une évaluation d'impact contrefactuelle. Le guide met à jour les précédentes orientations pour la période 2014-2020, en mettant davantage l'accent sur les difficultés rencontrées lors de la mise en œuvre pratique d'une EIC. Néanmoins, les aspects méthodologiques sont examinés et, si possible, simplifiés et intégrés au moyen d'exemples actualisés sélectionnés à partir des évaluations réalisées par le FSE dans la période 2014-2020.

Les EIC répondent aux questions essentielles qui permettent la prise de décisions stratégiques fondées sur des données factuelles: quels sont les effets produits par les interventions et «qu'est-ce qui fonctionne?». Elles cherchent à obtenir la preuve que les interventions financées par le FSE sont effectivement responsables des changements dans la situation des participants et des réalisations découlant des interventions. Réalisées

Une EIC répond à la question «qu'est-ce qui fonctionne?»

correctement, les EIC démontrent l'«effet net»¹ ou l'impact d'une intervention et permettent aux responsables politiques d'éliminer toute autre explication des changements éventuellement observés dans la situation ou les réalisations des participants. Lorsque les EIC fournissent des estimations de la présence et de l'ampleur de l'effet net, celles-ci contiennent nécessairement une mesure de l'incertitude, en fonction de la précision méthodologique et des informations disponibles. La nature des preuves fournies par les EIC permet aux responsables politiques d'évaluer l'efficacité des interventions, de comparer les interventions et d'évaluer leurs résultats respectifs. Les EIC apportent en outre une importante contribution aux analyses de rentabilité ou d'efficacité.

Le présent guide est publié alors que le FSE+ fait face à des défis sans précédent. Compte tenu de l'augmentation considérable des ressources de l'UE en faveur des investissements et de l'emploi prévue par le paquet NextGenerationEU, il est essentiel que les responsables politiques mesurent et comprennent les effets des interventions dont ils sont responsables. Des fonds publics doivent être alloués à des interventions plus productives et plus efficaces afin d'accélérer la reprise et de réduire les déséquilibres sociaux. Il incombe donc à ces responsables qui puisent dans les ressources du FSE+ de justifier leurs choix en démontrant que leurs interventions sont efficaces et ont des effets positifs pour les citoyens. Le meilleur moyen d'y parvenir est de réaliser un plus grand nombre d'EIC de haute qualité.

Le FSE+ est le principal instrument européen de soutien à l'emploi et à l'inclusion sociale. Dans la période de programmation 2014-2020, le FSE a alloué près de 125 milliards d'EUR aux politiques actives du marché du travail, de l'éducation et de l'inclusion sociale mises en œuvre à travers des programmes opérationnels (PO) dans les 28 États membres. Comme le prévoit le règlement (UE) n° 1303/2013 portant dispositions communes, les évaluations «sont effectuées dans le but d'améliorer la qualité de la conception et de la mise en œuvre des programmes et d'évaluer leur efficacité, leur efficience et leur impact».

Dans la période de programmation 2021-2027, les performances et les résultats continueront d'être examinés². Pour cela, il faudra renforcer les systèmes et les capacités actuels de suivi et d'évaluation, y compris les mécanismes de collecte de données. Les plans d'évaluation resteront obligatoires et une plus grande importance sera accordée à l'évaluation d'impact. Étant donné que plusieurs méthodes peuvent être utilisées pour mesurer les effets des opérations soutenues par le FSE+, il revient aux autorités de gestion de choisir la méthode ou la combinaison de méthodes la mieux adaptée pour répondre aux exigences réglementaires. La quantification rigoureuse des effets produits par les interventions suppose également l'élaboration de scénarios contrefactuels.

Orientation vers les résultats et évaluation de haute qualité

¹ L'effet net ou l'impact est la différence entre l'effet total (ou brut) et ce qui aurait été réalisé en l'absence d'intervention. L'effet net peut également être négatif, lorsque l'intervention est moins efficace que les dynamiques du marché. Voir la page 9 pour de plus amples détails.

² RÈGLEMENT (UE) 2021/1060 DU PARLEMENT EUROPÉEN ET DU CONSEIL du 24 juin 2021 portant dispositions communes relatives au Fonds européen de développement régional, au Fonds social européen plus, au Fonds de cohésion, au Fonds pour une transition juste et au Fonds européen pour les affaires maritimes, la pêche et l'aquaculture, et établissant les règles financières applicables à ces Fonds et au Fonds «Asile, migration et intégration», au Fonds pour la sécurité intérieure et à l'instrument de soutien financier à la gestion des frontières et à la politique des visas.

L'accent mis sur les bonnes performances et l'orientation sur les résultats est une caractéristique importante de la nouvelle réglementation. Des stratégies et des techniques d'évaluation de haute qualité sont essentielles pour rassembler les connaissances indispensables permettant aux EM de différencier les interventions qui «fonctionnent» de celles qui ne fonctionnent pas. Il est capital de renforcer la qualité des évaluations et d'acquérir des données fiables sur la valeur ajoutée.

En principe, le point de départ pour rassembler des données probantes sur l'efficacité des interventions stratégiques est simple. Les conditions incluent:

- la détermination du problème à résoudre;
- la détermination des instruments à utiliser pour résoudre le problème;

une formule qui relie les instruments et les résultats.

Afin d'évaluer un programme ou un instrument de financement, il est essentiel de disposer d'indicateurs clairs et mesurables des moyens mis en œuvre et des réalisations et résultats propres au problème. Il est courant de fixer des objectifs à la fois pour les réalisations et les résultats et de comparer les accomplissements réels aux objectifs. Le suivi sert à corrélérer les moyens engagés aux réalisations et résultats obtenus au fil du temps et à apporter un retour d'informations en matière de gestion. La théorie d'intervention sous-jacente révèle souvent des résultats intermédiaires qui peuvent également faire l'objet du suivi. Cependant, déterminer, à partir de ce point de départ, si une intervention donnée fonctionne n'est pas chose aisée.

Le besoin de résultats...

Il est nécessaire de compléter les pratiques d'évaluation existantes par des approches qui produisent des preuves bien plus solides des effets nets ou des impacts des interventions. Mesurer les réalisations revient à rendre compte des fonds utilisés. Les EIC répondent à la question fondamentale de savoir si une intervention est efficace. Alors que les EIC cherchent à établir un lien de causalité entre les interventions et les résultats, d'autres méthodes d'évaluation essentiellement théoriques et d'évaluation des processus peuvent être nécessaires pour repérer les mécanismes de causalité sous-jacents et veiller à ce que les impacts produits en un lieu fournissent une base de connaissances permettant une reproduction de l'intervention ailleurs.

... et de preuves des effets nets

Au cours de la période de programmation 2014-2020, tous les États membres et les autorités de gestion ont adopté un plan d'évaluation décrivant les objectifs, les activités, les éléments organisationnels et les produits de l'évaluation. Les plans d'évaluation prévoient plusieurs types d'évaluations, comme l'évaluation globale continue des programmes, ainsi que des évaluations thématiques visant à répondre à des questions d'évaluation spécifiques ou ciblant des axes particuliers ou des priorités d'investissement d'un PO. Il ressort de l'analyse des plans d'évaluation liés aux 177 programmes du FSE, à la fin de l'année 2018³, que près de 132 évaluations d'impact contrefactuelles étaient prévues dans cette période de programmation, un chiffre qui montre les progrès réalisés et l'intérêt accru pour les méthodes contrefactuelles dans la période de programmation 2014-2020 par rapport à la période 2007-2013⁴.

Expérience des EM en matière d'EIC

³ Voir [Ismeri Europa – Ecorys – Institute for Employment Studies, 2019](#).

⁴ Voir [Bratu C. et al., 2014](#).

Selon le répertoire des projets du Helpdesk Évaluation⁵, 1 795 évaluations ont été réalisées jusqu'en juin 2021, dont 1 001 portaient sur des programmes FSE/IEJ ou des programmes FSE/IEJ et FEDER (675 et 326 respectivement). Sur les 1 001 évaluations couvrant le FSE, 323 étaient des évaluations d'impact et 106 d'entre elles ont appliqué une méthode contrefactuelle: 82 évaluations d'impact sur 234 concernaient des programmes FSE/IEJ et 24 évaluations d'impact sur 89 portaient sur des programmes FSE/IEJ et FEDER⁶.

Le recours à l'EIC augmente

Au cours des premières années de la période de programmation actuelle, les évaluations contrefactuelles des programmes FSE/IEJ et/ou FEDER se sont concentrées sur la période de programmation précédente, tandis que, depuis 2018, presque toutes les EIC évaluent les effets des interventions de la période 2014-2020⁷. Vingt États membres sur vingt-sept ont réalisé des évaluations contrefactuelles de la programmation concernant le FSE, ce qui reflète également, dans ce cas, une adoption plus large de cette approche par rapport au passé.

Bien que les approches méthodologiques des EIC varient d'un État membre à l'autre, l'approche contrefactuelle a été la plus souvent utilisée pour évaluer l'impact des politiques actives du marché du travail (formation, incitations, aide à la recherche d'emploi, expériences professionnelles), mais dans une bien moindre mesure dans d'autres domaines, tels que les interventions visant à soutenir les établissements d'enseignement et les étudiants, et les politiques relatives aux questions sociales, telles que les interventions visant à lutter contre la pauvreté ou l'exclusion sociale.

D'un point de vue méthodologique, l'appariement par score de propension est la technique la plus fréquemment utilisée, les autres méthodes étant moins répandues. Les EIC se concentrent souvent sur les effets à court ou moyen terme (6 ou 12 mois), tandis que les effets à long terme sont moins fréquemment évalués.

L'analyse des EIC effectuée dans le cadre du projet du Helpdesk Évaluation a révélé des limites, notamment: une similarité insuffisante entre les personnes figurant dans les groupes témoins et celles dans les groupes de soutien, la faible fiabilité des données, par exemple lorsqu'elles proviennent d'enquêtes ex post, la réalisation d'entretiens comptant sur la capacité des répondants à se rappeler des événements avec précision, ou l'utilisation d'échantillons de petite taille produisant des résultats statistiquement insignifiants, etc. Dans certains cas, les rapports d'évaluation ne contiennent pas suffisamment d'informations sur les étapes, les choix et les limites méthodologiques, qui sont des éléments nécessaires pour informer les personnes autres que celles ayant commandé l'évaluation des effets des mesures examinées et leur permettre de mieux connaître lesdits effets.

mais des limites sont encore fréquentes

En résumé, malgré les progrès réalisés ces dernières années, la réalisation des EIC exige une expertise technique et une volonté politique. Le présent guide apporte des arguments en faveur des EIC et aborde certaines des questions sur lesquelles les AG doivent se pencher pour réussir leur mise

Guide à l'intention des praticiens

⁵ Le projet du Helpdesk, financé par la DG REGIO et la DG EMPL, recueille des informations sur les évaluations publiées en ligne depuis le 1^{er} janvier 2015 sur les sites internet des AG. Le recensement des évaluations est effectué par un réseau d'experts nationaux et elles sont résumées et évaluées en fonction de la qualité et de la fiabilité des résultats.

https://ec.europa.eu/regional_policy/en/policy/evaluations/member-states/

⁶ Sur les 106 évaluations d'impact contrefactuelles, 90 ont été examinées et résumées par le projet du Helpdesk.

⁷ Vingt-neuf des trente-trois EIC réalisées jusqu'en 2018 concernent la période de programmation précédente.

en œuvre. Au-delà des considérations d'ordre pratique des EIC, il se concentre également sur des questions plus larges qu'il faudrait éventuellement aborder pour améliorer la qualité des évaluations d'impact. Le guide contient un certain nombre de recommandations de base qui, espérons-le, seront utiles aux AG, mais l'un des principaux messages est que la réalisation des EIC nécessite une planification minutieuse de la préparation des données (en particulier des données administratives), des buts et des objectifs clairs, etc., afin d'éviter d'éventuels problèmes lors de la phase de mise en œuvre ou des évaluations de qualité médiocre.

Le guide contient également des exemples d'évaluations et de pratiques destinés à aider le lecteur. Si les évaluations recueillies par le projet du Helpdesk Évaluation constituent la principale source d'information pour identifier les exemples et les pratiques présentés dans le présent guide, d'autres sources ont également été prises en considération, notamment les évaluations réalisées par le Centre commun de recherche (centre de recherche sur les évaluations d'impact, CRIE) et, dans une mesure plus limitée, les publications universitaires. Dans la mesure du possible, les exemples et les expériences figurant dans le guide ont trait aux interventions du Fonds social européen.

Le guide s'articule autour de quatre chapitres.

Le **premier chapitre** aborde la nature des EIC et explique pourquoi elles sont importantes. Il donne une vue d'ensemble introductive des approches de l'EIC, en insistant sur la distinction essentielle entre les approches expérimentales et quasi expérimentales. Il présente également les difficultés techniques et pratiques qui doivent être résolues lors de la mise en œuvre d'une EIC. Une attention est également accordée, d'une manière générale, aux types de questions stratégiques qui pourraient être traitées par les EIC et au rapport entre les méthodes d'EIC et d'autres approches de l'évaluation (par exemple, les approches basées sur la théorie, l'évaluation des processus et l'analyse de l'efficience).

*Une articulation
reflétant les étapes
d'une évaluation*

Le **deuxième chapitre** se penche sur une série de questions que les AG devraient examiner lorsqu'elles conçoivent et mettent en œuvre des EIC. Le présent guide présente certains des problèmes majeurs qui se posent fréquemment aux responsables de l'élaboration des EIC et propose quelques recommandations sur la manière d'y faire face. Les questions servent à guider ceux qui projettent de commander des EIC d'interventions financées par le FSE.

Les EIC apportent des preuves solides de l'efficacité des fonds. Mais pour cela, elles doivent être bien planifiées et correctement exécutées. Pour ce faire, les AG doivent aborder certaines questions essentielles lorsqu'elles commandent une évaluation. La manière exacte dont les AG examinent ces questions et l'ordre dans lequel elles le font dépendront des aspects pratiques et des modalités institutionnelles en vigueur au sein des États membres. Ce guide met en évidence certaines de ces questions importantes et les porte à l'attention des AG. Les principales questions sont énumérées selon les six étapes du processus d'évaluation: 1) sélectionner l'opération ou les opérations qui feront l'objet de l'évaluation; 2) déterminer les questions d'évaluation et les variables de résultat; 3) analyser les données pour identifier le groupe témoin et mesurer les variables de résultat; 4) sélectionner la méthode d'EIC; 5) définir le calendrier et le budget; 6) mettre en œuvre l'évaluation. Une attention particulière est accordée à la

disponibilité et à la protection des données, une question qui peut réduire sérieusement la possibilité de réaliser une EIC.

Le **troisième chapitre** examine en détail l'étape 4 ci-dessus (sélectionner la méthode d'EIC) et se concentre sur les caractéristiques des principales méthodes appliquées dans les analyses contrefactuelles. Cinq méthodes sont présentées en particulier: la méthode expérimentale, ou la randomisation, l'appariement par score de propension, les doubles différences, l'approche de la régression par discontinuité et les variables instrumentales. Ce chapitre du guide vise non pas à former les lecteurs à l'application des méthodes d'EIC, mais à les familiariser avec les points forts et les points faibles de chaque méthode et à les aider à déterminer la méthode la plus appropriée.

Le **quatrième chapitre** aborde des questions plus larges ayant trait à l'élargissement des perspectives, notamment le besoin de développer la capacité de mener à bien des EIC, tant au sein des AG (responsables politiques et fonctionnaires) que parmi les communautés de chercheurs et d'universitaires et les sociétés de conseil des EM. Ce chapitre traite également de la nécessité de faire face aux obstacles juridiques liés à l'accès aux données et de mettre à jour les EIC des programmes du FSE+ pour les adapter à des conceptions plus avancées afin d'étendre la couverture de la politique en matière d'inclusion sociale et d'éducation, l'estimation des effets des résultats difficilement quantifiables («soft») et la mise en œuvre d'approches de méta-évaluation.

En résumé, le présent guide: 1) apporte des arguments en faveur des EIC, 2) signale les étapes majeures vers une exécution réussie des EIC et 3) présente les EIC comme un élément essentiel du paysage du FSE+. L'objectif ultime est d'améliorer la contribution du FSE+ au bien-être des citoyens européens.

Les auteurs souhaitent reconnaître le soutien et l'assistance précieux fournis par les membres de l'unité G5 de la DG Emploi, Mieux légiférer, dans l'élaboration de ce rapport, en particulier Linda Adamaite, Jeannette Monier et Maria José Cueto Faus. Les auteurs tiennent également à remercier Andrea Pisano et Ernesto Belisario pour leur contribution à la section sur la protection des données et Paweł Hess pour son soutien et sa contribution.

Chapitre 1. Concept et approches

Le présent chapitre aborde des questions fondamentales relatives à la nature et à l'objet de l'approche contrefactuelle en évaluant les effets causaux des interventions. Plus particulièrement, il propose une compréhension de l'essence de l'évaluation d'impact contrefactuelle, notamment lorsqu'elle concerne les types d'opérations cofinancées par le FSE+. Il étudie également le lien entre les approches contrefactuelles et d'autres méthodes d'évaluation et examine pourquoi les EIC sont importantes, particulièrement de nos jours. Enfin, ce chapitre examine les questions stratégiques auxquelles peuvent répondre les EIC et donne un bref aperçu de certaines des principales méthodes contrefactuelles pertinentes pour évaluer les interventions cofinancées par le FSE+ et présentées plus en détail au chapitre 3.

1.1. L'essence de la méthode contrefactuelle

Les EIC cherchent à isoler les effets ou impacts finaux des interventions. Leur particularité réside dans le fait qu'elles visent à confirmer qu'une intervention donnée entraîne un résultat spécifique; en d'autres termes, le résultat concret ne peut être imputé qu'à l'intervention. Pour ce faire, les EIC isolent l'intervention et excluent les autres explications du résultat observé.

Mettre en évidence leur capacité d'exclure les autres explications, voilà l'idée de la méthode «contrefactuelle», c'est-à-dire la réponse à la question suivante: «Que se serait-il passé — en ce qui concerne les résultats présentant un intérêt — si l'intervention n'avait pas été mise en œuvre?». Pour bien comprendre le concept et l'expliquer très simplement, il est utile de prendre l'exemple d'une personne sans emploi qui participe à un programme de formation dont le but est d'inciter à l'emploi. Pour déterminer l'effet de la formation sur cette personne, la méthode contrefactuelle ambitionne deux résultats potentiels⁸. Le premier, c'est la situation au regard de l'emploi du participant au terme de la formation. Il s'agit du résultat observé sur la personne formée. Le deuxième, c'est la situation au regard de l'emploi de cette même personne si elle n'avait pas participé à la formation, toutes choses étant égales par ailleurs. Dans ce contexte, le deuxième résultat sera le résultat contrefactuel. L'impact de la formation sur la personne formée est déterminé par la différence entre le résultat observé et le résultat contrefactuel. Il s'agit de l'effet causal net ou de l'impact de la formation sur la personne. La seule différence dans les circonstances ou les conditions qui donnent lieu aux résultats observés et contrefactuels est la participation de cette personne à la formation. Ainsi, chaque différence entre les deux résultats représentera nécessairement l'impact de la formation sur la situation au regard de l'emploi de la personne.

*Méthode
contrefactuelle*

En réalité, nous n'observons pas, et ne pouvons le faire, les résultats contrefactuels sur des particuliers visés par une intervention. L'objectif premier des EIC consiste à donner des estimations convaincantes des résultats contrefactuels sur des groupes d'individus ou des entreprises concernés par les interventions cofinancées par le FSE. Ainsi, les impacts sont exprimés, par exemple, par la mesure de la différence entre les valeurs

*Définir les groupes
de traitement et...*

⁸ Un examen plus approfondi du modèle de causalité fondé sur les «résultats potentiels» est disponible dans [Holland P., 1986](#).

contrefactuelles moyennes observées et «estimées». Dans la plupart des cas, les EIC comparent les résultats d'une intervention (une mesure ou un instrument) sur les entités (personnes, PME, etc.) qui en ont bénéficié avec ceux d'un groupe qui n'a pas fait l'objet de l'intervention. Dans la terminologie des EIC, le groupe «traité» ou groupe «de traitement» (groupe des bénéficiaires d'une intervention) est à distinguer du groupe «témoin», qui doit être aussi semblable que possible au groupe de traitement, et ce à tous les égards (hormis les traitements reçus). Le groupe témoin fournit les estimations des résultats contrefactuels, une attention particulière étant portée aux différences (observées ou non) entre les caractéristiques des deux groupes. Il est également possible de comparer différents traitements en exposant les unités éligibles à un éventail de variantes de traitements (d'autres traitements ou interventions financés par le FSE ou par d'autres sources, par exemple), en créant plusieurs groupes de traitement et en comparant les résultats les uns avec les autres et/ou avec les résultats sur un groupe témoin non traité.

Lorsque le groupe témoin ne fait l'objet d'aucun traitement, l'évaluation répond à la question: «Quel est l'impact du bénéfice d'une intervention par rapport au fait de ne recevoir aucune aide ou aucun soutien?». Inversement, lorsque les résultats du bénéfice d'une intervention visée sont comparés aux résultats d'une autre intervention, l'évaluation répond à la question: «Quel est l'impact de l'intervention examinée par rapport à une autre intervention clairement définie?». Dans de nombreux cas, une EIC peut être élaborée pour répondre à ces questions fondamentales. Le choix de la question à laquelle répondre dépend des priorités des responsables politiques et des contraintes pratiques au niveau de la conception.

... les groupes témoins

Lorsqu'une comparaison est établie entre deux interventions différentes, celles-ci doivent être clairement contrastées, ce qui est important pour l'élaboration des politiques.

1.2. Pourquoi les évaluations contrefactuelles sont-elles importantes?

Les EIC apportent d'importantes informations sur les effets ou impacts finaux des interventions. Elles fournissent des estimations de l'ampleur et de la nature (positive ou négative) des impacts et des mesures statistiques de l'incertitude. Elles permettent de valider ou de rejeter le lien causal présumé entre l'intervention et les résultats, qui sous-tend la conception de l'intervention. Ces effets mesurés peuvent servir à atteindre un certain nombre d'objectifs: faire preuve de transparence et rendre des comptes quant à l'utilisation des ressources des contribuables; prendre des décisions stratégiques sur la base de preuves claires; tirer des enseignements des résultats obtenus dans les différentes régions d'un même pays ou dans les différents EM.

Les personnes responsables des interventions et soucieuses de continuer à bénéficier d'un financement de leurs programmes auront tout intérêt à promouvoir les EIC afin de montrer que leurs programmes offrent le meilleur rapport coût-avantage et produisent des bénéfices mesurables pour les participants et pour la société dans son ensemble. Les preuves fournies par les EIC seront particulièrement intéressantes pour les responsables de l'affectation des ressources. Les AG ne demanderont qu'à montrer que leurs

Transparence et obligation de rendre des comptes

programmes «fonctionnent» bel et bien. Pour réussir à convaincre, elles devront commander des EIC de haute qualité.

Grâce à ces caractéristiques, les EIC apportent d'importantes informations aux responsables politiques chargés d'allouer les ressources aux différentes interventions. Les décisions de financement d'interventions potentielles se prennent dans un contexte de restriction. Dès lors, les décideurs ont besoin de preuves fiables de l'impact des programmes et de leur rentabilité, de manière à utiliser au mieux les ressources disponibles. Dans l'évaluation de l'efficacité relative des interventions, les effets finaux estimés par les EIC peuvent être présentés plus en détail en étudiant le rapport coût-efficacité d'une intervention ou en procédant à une analyse coût-avantage complète.

Soutenir les décisions d'allocation des ressources

D'importantes connaissances supplémentaires peuvent également découler de la comparaison des effets nets d'opérations similaires mises en œuvre sur différents territoires. Si les EIC détectent des différences significatives dans leurs effets, cela signifie très probablement que la qualité des opérations diffère sensiblement ou que certaines opérations correspondent mieux à leur contexte socio-économique que d'autres. Ces comparaisons, généralement appelées «méta-évaluations» ou «méta-analyses», permettent une généralisation plus poussée des résultats des EIC et fournissent des tests importants pour les mesures politiques (formation, services intégrés, etc.)⁹. Pour effectuer ces comparaisons sur la base de critères scientifiques, un nombre adéquat d'EIC et une mise en œuvre systématique de ces dernières sont nécessaires. Par conséquent, une bonne coordination au niveau national ou de l'UE est une condition fondamentale pour augmenter la fréquence et l'utilité de ces comparaisons en vue d'éclairer les décisions stratégiques.

Comparaison des opérations de différentes régions ou de différents EM

1.3. Pourquoi les évaluations contrefactuelles sont-elles techniquement difficiles?

Il existe plusieurs approches qu'on pourrait qualifier de tentatives «non fiables» pour estimer les impacts d'une intervention. Elles sont examinées ici afin d'illustrer les complexités inhérentes aux EIC; aucune référence n'est faite à la pratique d'évaluation réelle.

Aller au-delà des approches simplistes

Un premier exemple est celui d'un responsable politique qui souhaite évaluer l'impact d'un programme de formation pour les chômeurs en comparant les revenus des participants à la formation au terme de celle-ci avec les revenus de tous ceux qui n'ont pas participé au programme. Il attribue alors au programme de formation les différences de revenus observées entre les participants et les autres.

Ce n'est pas une stratégie adéquate pour mettre en évidence l'impact de la formation sur les revenus, puisque les personnes non formées peuvent se distinguer à d'importants égards des participants à la formation, et ces différences peuvent influencer les résultats. Une telle approche ne permet donc pas d'exclure les autres explications des différences de revenus observées. Les personnes formées, par exemple, ont peut-être des compétences inhérentes supérieures aux personnes non formées. Autrement dit, des personnes sans emploi dotées de compétences supérieures se sont

⁹ Voir par exemple, [Card D., Kluve J. et Weber A., 2017.](#)

portées volontaires pour participer au programme de formation. Ainsi, les compétences influencent non seulement la décision de participer, mais également les résultats; les personnes sans emploi dotées de compétences inhérentes de niveau supérieur seront plus susceptibles de percevoir des revenus plus élevés que les personnes aux compétences inférieures. En conséquence, toute différence de revenus observée entre les groupes de traitement et les groupes témoins peut être due à la fois au programme de formation et/ou aux différences inhérentes de compétences et induire à «surestimer» l'impact de l'intervention.

Si les compétences ne peuvent être mesurées et si les différences de compétences inhérentes entre les deux groupes ne peuvent être prises en considération dans l'évaluation des impacts en comparant les résultats entre les deux groupes, l'impact évalué du programme de formation pêche par un **biais de sélection**. Pour contrer ce problème, les évaluateurs tentent de réunir le plus d'informations possible sur les facteurs majeurs qui influencent la décision de participer et les résultats produits. Ces données sont utilisées pour sélectionner un groupe témoin valide parmi les non-participants, c'est-à-dire un groupe qui ressemble le plus possible au groupe de traitement en ce qui concerne ces facteurs, et effectuer des analyses statistiques en tenant compte de ces facteurs. Ce faisant, les évaluateurs posent souvent pour hypothèse que la sélection dans le programme est dictée par ces facteurs observables. Cette «hypothèse d'identification» ne peut pas être facilement corroborée en général et l'évaluateur doit faire valoir de manière convaincante que l'hypothèse est plausible dans le contexte donné sur la base de ses connaissances des facteurs institutionnels et de la théorie comportementale.

Neutraliser le biais de sélection

Une deuxième approche «non fiable» serait celle où le responsable politique observerait les revenus des participants avant et après la formation et attribuerait le changement de revenus opéré au programme de formation. En substance, cette approche suppose qu'en l'absence d'intervention, les revenus moyens resteraient inchangés.

Avant et après le changement

À nouveau, dans la grande majorité des cas, cette stratégie n'est pas adéquate pour mesurer l'impact de la formation sur les revenus, sauf si l'hypothèse d'une stabilité temporelle peut vraisemblablement être invoquée. La raison en est que les revenus des participants changeront inévitablement au fil du temps, et ce d'une manière tout à fait indépendante de la formation. Par exemple, il est courant d'observer que les salaires des participants baissent avant la participation, en partie sous l'effet de facteurs transitoires. Dans de nombreux cas, une reprise s'amorcera indépendamment de l'intervention de la formation¹⁰. L'approche non fiable qui consiste à évaluer l'impact de la formation sur la base de la différence entre les salaires enregistrés immédiatement avant la formation et ceux enregistrés après, fait abstraction du fait que, dans de nombreux cas, les salaires auraient augmenté de toute façon.

Pour ajuster de telles approches, une mesure contrefactuelle est nécessaire afin de savoir dans quelle mesure les revenus des participants auraient changé sans l'intervention de la formation. Il est possible, par exemple, d'aboutir à un tel résultat contrefactuel en constituant soigneusement un

¹⁰ Ce mécanisme est connu sous le nom d'«Ashenfelter Dip», d'après le premier économiste à l'avoir commenté. Voir [Ashenfelter O., 1978](#).

groupe témoin n'ayant pas reçu la formation, et dont les revenus sont observés aux mêmes moments que ceux des participants. Il est alors souvent posé pour hypothèse que l'évolution des revenus des participants et du groupe témoin aurait été la même sans l'intervention.

Les limites de ces approches «non fiables» encouragent la recherche de méthodes d'évaluation plus convaincantes. Comme il a été dit précédemment, de telles méthodes sont toutefois techniquement plus difficiles à mettre en œuvre. La section suivante de ce chapitre décrit brièvement certaines des approches particulières des EIC qui se prêtent le mieux au contexte du FSE.

1.4. Aperçu des conceptions et des approches des EIC

Lorsque l'EIC se concentre sur les effets d'une intervention sur les participants, les résultats contrefactuels sont généralement estimés grâce aux informations recueillies auprès de non-participants qui s'apparentent aux participants à l'intervention évaluée. Le tableau 1 à la fin de ce chapitre donne un bref aperçu des approches, de certains de leurs avantages et limites et des principales données qu'elles requièrent.

La première distinction à faire pour une EIC concerne la conception de l'évaluation, qui peut être expérimentale ou quasi expérimentale. L'approche expérimentale est généralement désignée comme l'«essai aléatoire contrôlé» (randomised control trial, RCT) et parfois aussi comme l'«expérimentation sociale».

Parmi les méthodes d'EIC destinées à évaluer les effets des interventions sur des périodes relativement courtes, l'approche expérimentale est considérée comme la référence et, dans la plupart des cas, comme l'approche idéale. Une évaluation d'impact bien conçue doit viser à obtenir des estimations de résultats contrefactuels impartiales. Dans de nombreux cas, on considère qu'une approche expérimentale produit ces estimations impartiales. Pour examiner les approches de l'EIC, il est souvent préférable de commencer par dresser les grandes lignes de l'approche expérimentale. La raison en est que les méthodes quasi expérimentales cherchent pour la plupart à reproduire l'idéal expérimental.

L'approche aléatoire: la référence

Dans l'examen des modèles d'approche de l'EIC, les principales caractéristiques de chaque approche sont présentées le plus simplement possible afin de clarifier les principes sous-jacents. En réalité, l'utilisation de ces méthodes peut être infiniment plus complexe et des questions comme la non-conformité, c'est-à-dire le fait que des personnes supposées participer à l'intervention ne l'ont pas fait, peuvent accroître de manière significative les problèmes qui se posent.

Il existe une grande variété d'approches qui visent essentiellement à imiter la répartition aléatoire. Ce sont les approches dites **quasi expérimentales**. Il n'est pas possible de toutes les passer en revue dans le cadre du présent guide, ni de fournir une présentation technique complète et détaillée de chacune. Toutefois, les méthodes quasi expérimentales les plus susceptibles d'être appliquées dans le cadre du FSE sont: 1. l'appariement par score de propension; 2. les doubles différences; 3. la régression par discontinuité; 4. les variables instrumentales. Une synthèse des approches majeures et de leurs avantages respectifs figure dans le Tableau 5 du chapitre 3; leur présentation est grandement simplifiée afin de mettre en évidence les

Approches non aléatoires ou quasi expérimentales

principes essentiels de chaque approche. L'annexe 1 propose d'autres lectures relatives aux méthodes quasi expérimentales.

1.5. Comment intégrer une EIC dans un cadre d'évaluation plus large?

Les évaluations contrefactuelles abordent certains types de questions concernant les effets produits par les interventions. Ces approches sont limitées dans la mesure où elles pourraient aborder d'autres questions concernant une intervention. Il est nécessaire de distinguer les questions d'évaluation relatives à l'**explication causale** et celles qui concernent la **description causale**. Les EIC visent à **décrire** les conséquences d'une intervention. Ces méthodes sont moins aptes à **expliquer** les mécanismes et les contextes grâce auxquels les liens de causalité sont établis. Cette distinction est importante, car elle permet de clarifier le rôle distinctif de l'EIC¹¹.

Explication et description causales

Une EIC bien conçue montrera au responsable politique si une intervention a entraîné les changements de résultats qu'elle devait influencer. Elle fournira des informations concrètes sur l'ampleur d'un impact ou d'un effet, montrera au responsable politique si l'impact a été positif ou négatif et fournira également une mesure de l'incertitude. Toutefois, les évaluations d'impact contrefactuelles sont moins efficaces pour expliquer la manière et les raisons pour lesquelles les impacts mesurés par l'EIC se sont produits. Inversement, il est souvent difficile de déterminer, sur la base d'une EIC, la raison pour laquelle une intervention n'a eu aucun impact, si cela s'avère être le cas.

Ce que les EIC peuvent indiquer ou non aux responsables politiques

Dans la plupart des organismes d'élaboration de politiques, les parties prenantes s'interrogeant sur la description causale et les questions d'explication causale ont tendance à viser des perspectives et des intérêts différents. Les gestionnaires de programme et les praticiens ont tendance à mettre l'accent sur des questions d'explication causale. Les responsables de l'affectation des ressources et les responsables politiques de haut niveau chargés de l'établissement du budget ont tendance à mettre l'accent sur des questions de description causale. Dans la pratique, la distinction entre l'explication causale et la description causale peut être floue. Dans certaines circonstances, les EIC peuvent expliquer les raisons pour lesquelles certains impacts se sont produits, en explorant par exemple l'impact d'interventions sur des sous-groupes importants. Cependant, il est essentiel d'examiner soigneusement les types de questions que les parties prenantes se posent au sujet d'une intervention, et de choisir l'approche appropriée pour y répondre. Dans les cas où la question principale est de savoir si une intervention fonctionne, une évaluation d'impact contrefactuelle est appropriée en de nombreuses circonstances. Dans les cas où la question principale est de savoir comment fonctionne une intervention, les méthodes d'évaluation fondées sur la théorie et les processus sont plus appropriées.

Ces différents niveaux de questions et d'objectifs sont résumés dans la figure ci-dessous.

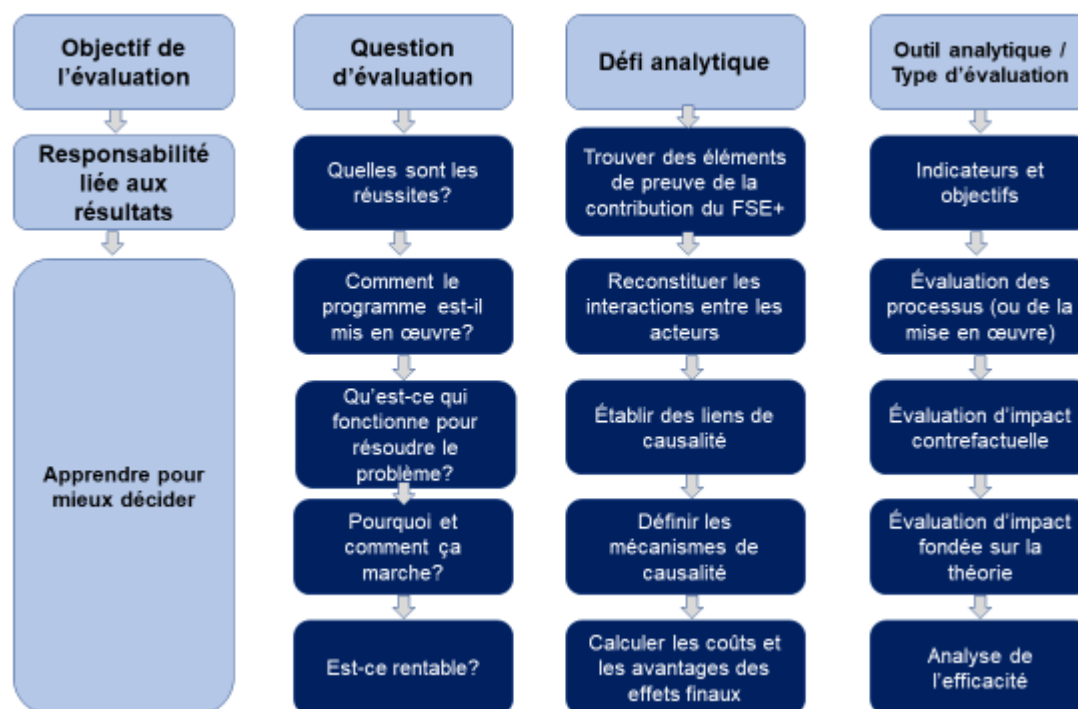
Cette discussion mène à la conclusion que les EIC doivent être élaborées dans le plan d'évaluation. Ce dernier doit comporter différentes formes

¹¹ Voir Shadish W.R., Cook T.D. et Campbell D.T., 2002, et [Stern E. et al., 2012](#).

d'évaluation qui visent à répondre à des questions différentes, pour des acteurs politiques différents. Dans la pratique, un plan d'évaluation intégrera rarement, voire jamais, une EIC sans comprendre une évaluation des processus.

Un large éventail d'approches sont adoptées au nom de l'évaluation, et poursuivent une série d'objectifs différents. La question essentielle est de savoir comment ces approches peuvent être combinées de manière efficace afin de promouvoir l'apprentissage politique. Combiner différents types d'évaluation de manière appropriée — avec des objectifs différents dans le cycle de programmation — est le véritable défi dans ce domaine. Comme examiné plus haut et indiqué dans la figure suivante, l'EIC, l'évaluation des processus et les approches fondées sur la théorie se complètent.

Figure 1 Différentes tâches et types d'évaluation



Source: adapté de [Martini A., 2009.](#)

Une stratégie d'évaluation solide devrait comprendre les éléments suivants:

- une évaluation d'impact fondée sur la théorie;
- une évaluation des processus (ou de la mise en œuvre);
- une évaluation d'impact contrefactuelle (EIC);
- une analyse de l'efficacité.

Combiner différents types d'évaluations

Dans ce guide, seules les approches contrefactuelles de l'évaluation d'impact sont examinées en détail. Dans le cadre d'une EIC, les approches fondées sur la théorie sont des moyens de comprendre la raison pour laquelle une intervention a été conçue.

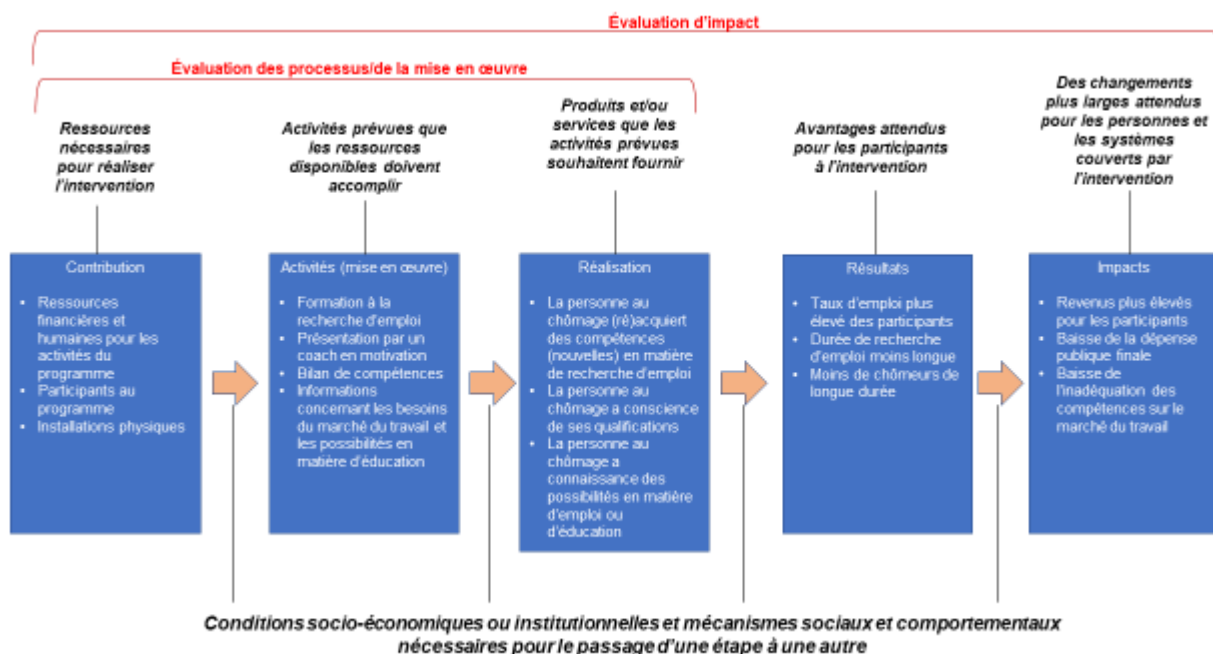
Les **évaluations fondées sur la théorie** sont utilisées dans certaines circonstances pour décrire l'utilisation prévue de l'intervention et pour vérifier si les changements de résultats prévus par la théorie d'intervention peuvent être constatés. En ce sens, les approches fondées sur la théorie peuvent être

L'évaluation fondée sur la théorie renvoie à une «théorie du changement»

utilisées pour évaluer l'impact en répondant à la question de savoir «comment» un effet a été produit, et pour examiner l'impact d'une intervention lorsque les EIC ne sont pas faisables. La figure suivante montre un exemple schématisé de théorie du changement pour une intervention visant à améliorer les services de l'emploi. Une évaluation fondée sur la théorie examine si les éléments probants confirment la séquence des étapes causales, de la contribution à l'impact, comme le suggère la théorie du changement, et dans quelles conditions et par quels mécanismes sociaux cela a été possible. Une explication détaillée de l'utilisation des approches fondées sur la théorie en vue de déterminer l'impact dépasse le propos du présent document.

Dans le cadre de l'EIC, l'évaluation fondée sur la théorie examine la manière dont une intervention est prévue et conçue, ainsi que la manière dont elle est censée fonctionner. L'approche consiste essentiellement à coopérer avec les parties prenantes d'une intervention afin d'élaborer un exposé commun de la «théorie du changement» à la base d'une intervention, telle que simplifiée dans la figure suivante. Toutes les interventions intègrent une logique de programme qui relie les moyens et les activités aux réalisations, aux résultats intermédiaires et ensuite aux résultats à long terme. Par conséquent, la formulation d'une théorie du changement constitue également un soutien important pour les EIC; cela facilite la reconstruction du processus de mise en œuvre et son influence éventuelle sur les réalisations et les résultats, ainsi que la détermination des variables de résultat les plus représentatives à vérifier dans l'analyse contrefactuelle.

Figure 2 Illustration de l'approche du modèle logique ou de la «théorie du changement»



Source: adapté de [W.K. Kellogg Foundation, 2004](#) et de [Bredgaard T., 2015](#).

L'évaluation fondée sur la théorie peut être associée efficacement à des évaluations d'impact contrefactuelles de plusieurs manières. Une théorie du changement (ou logique d'intervention) clairement formulée peut faciliter la conception d'une EIC. Entre autres aspects, une théorie du changement bien

... renforçant les EIC

définie peut indiquer à la personne responsable de la conception d'une évaluation de l'impact les éléments suivants:

- Quels résultats sont importants et doivent être mesurés?
- Quels pourraient être le signe et la taille probables des impacts de l'intervention?
- Quel est le groupe cible visé et comment un groupe témoin peut-il être sélectionné?
- En combien de temps les effets et les résultats du programme pourraient-ils se concrétiser?
- Quelles données pourraient être nécessaires pour mesurer la participation à l'intervention?
- Quel est le niveau de plausibilité du groupe témoin en tant que mesure contrefactuelle?

L'élaboration d'une théorie du changement peut également permettre de détecter les effets potentiels imprévus qui peuvent être pris en considération lors de la conception d'une EIC. Dans une certaine mesure, une théorie du changement clairement formulée peut aussi aider l'évaluateur à interpréter les résultats d'une EIC. Toutefois, en termes d'interprétation, une **évaluation des processus** peut également s'avérer très instructive.

Évaluation des processus

Une **évaluation des processus** effectuée dans le cadre d'une EIC poursuit deux objectifs. Le premier consiste à évaluer la «fidélité», le second consiste à évaluer l'écart qui existe entre les expériences des participants traités et des participants témoins.

L'évaluation de la «fidélité» examine la mesure dans laquelle une intervention, une fois achevée, est fidèle à sa conception. Une évaluation des processus examine les services qui ont été effectivement fournis aux participants de l'intervention. Sont-ils conformes à ce qui est prévu dans la théorie de l'intervention? Comment expliquer la variation de la prestation entre les différents sites, le cas échéant? La plupart des interventions disposent d'une logique de gestion et d'effet.

Évaluation de la fidélité

- La logique de gestion concerne la manière dont les organes d'exécution sont censés réagir aux règles et mesures incitatives du programme.
- La logique d'effet porte sur la manière dont les personnes ciblées par l'intervention sont censées réagir, compte tenu des services effectivement fournis.

L'aspect «fidélité» de l'analyse des processus fournit des informations sur ce qui a été effectivement accompli lors d'une intervention et, par conséquent, sur les éléments qui contribuent réellement aux effets observés. Il fournit également des retours d'informations essentiels pour la gestion de projet.

L'évaluation de l'écart est particulièrement importante dans le cadre d'une évaluation contrefactuelle. Il est fréquent de mettre l'accent sur les impacts de l'intervention, comme cela a été fait pour une grande partie du présent guide. Mais avant l'impact sur les résultats vient l'impact sur les moyens, l'écart en termes de possibilités entre les groupes traités et témoins qu'une intervention entraîne réellement. En principe, chaque EIC peut être «inversée» et le groupe de traitement peut être utilisé comme groupe témoin

Écart entre les groupes traités et témoins

afin d'évaluer le résultat pour les personnes qui faisaient partie, avant l'inversion, du groupe témoin. Cela implique d'en savoir autant sur la situation du groupe témoin que sur celle du groupe de traitement, parce que c'est en fonction de l'écart entre les deux groupes en termes de moyens que l'EIC attribue une causalité aux écarts dans les résultats.

En ce qui concerne à nouveau le programme de formation, on peut imaginer deux circonstances initiales tout à fait différentes. Dans l'une, le programme de formation est fourni dans un contexte général dans lequel rien de tel n'existe. Les contrôles s'en passent tout simplement. Mais il se peut aussi qu'il existe des substituts. La formation professionnelle peut être fournie, par exemple, par des entreprises spécialisées. Si tel est le cas, l'analyse des processus doit inclure, dans la mesure du possible, l'évaluation de l'écart de participation aux formations entre le traitement et le contrôle, et ne pas simplement supposer que toutes les dimensions du traitement sont hors de portée du groupe témoin.

Alors que les évaluations des processus peuvent être commandées de manière totalement indépendante d'autres formes d'évaluation, leur importance tant pour la gestion que pour l'EIC impose la planification conjointe de l'évaluation des processus et de l'évaluation d'impact.

L'EIC nécessite une évaluation des processus

Une bonne analyse des processus peut favoriser la fidélité, et des évaluations des processus fournissent une explication causale d'une intervention. Sans une évaluation des processus bien conçue, il est souvent difficile d'interpréter pleinement les résultats d'une EIC ou de mesurer les coûts nécessaires pour l'évaluation de rentabilité, une fois que les estimations de l'impact sont disponibles.

Comme il est noté ci-dessus, une évaluation des processus peut également contribuer à l'interprétation des résultats des évaluations d'impact en rendant compte du contexte dans lequel une intervention a été réalisée. Il importe de comprendre le contexte, car il fournit les conditions nécessaires à la réussite de l'intervention et donne une idée de la mesure dans laquelle elle peut produire des effets similaires si elle est réalisée ailleurs, dans des zones géographiques différentes ou à des moments différents. Ce point est particulièrement important pour discuter de la transférabilité des approches stratégiques et pour souligner les bonnes pratiques dans l'apprentissage et l'échange transnational. L'analyse des processus contribue à favoriser la confiance dans ce que l'on appelle la **validité externe** des résultats de l'évaluation.

Dans la plupart des cas, l'analyse de l'efficacité implique soit une évaluation du rapport coût-efficacité, soit une analyse de rentabilité complète.

Détermination des rapports coût-efficacité

Une **analyse d'efficacité** suppose de comparer les coûts de l'intervention à ses effets ou impacts qui ont été déterminés par une EIC. En clair, le rapport coût-efficacité est obtenu en divisant l'impact d'une intervention — exprimé en unités de mesure ou en unités standardisées — par le coût net de l'intervention par unité traitée.

Le rapport coût-efficacité d'un programme de formation qui vise à aider les chômeurs à trouver du travail pourrait révéler les fonds nécessaires par participant pour placer un participant sur le marché du travail.

Comparaison des avantages et des coûts nets au moyen de l'analyse de rentabilité

L'analyse d'efficacité est un instrument important pour les responsables de l'allocation des ressources aux divers programmes. Les résultats obtenus

dans une série d'interventions différentes permettent à ces responsables de déterminer de manière relative les interventions qui optimisent le mieux les ressources.

Au lieu d'exprimer les effets du programme dans leur unité de mesure ou en unités standardisées, une **analyse de rentabilité** tente de monétiser les estimations d'impact obtenues grâce à une EIC et de les comparer aux coûts nets d'une intervention. L'objectif d'une telle analyse est de déterminer si les avantages monétisés d'un programme dépassent ses coûts nets. L'analyse du rapport coût-efficacité d'un programme de formation typique au titre du FSE comparerait les avantages de l'intervention pour ses participants, et plus largement les pouvoirs publics et la société, aux coûts nets de l'intervention. Pour les participants, les avantages du programme (généralement, une amélioration de l'employabilité et une augmentation des salaires) sont obtenus grâce à une EIC. On en déduira les montants payés par les participants et d'autres coûts salariaux afin d'obtenir un avantage net. Pour les pouvoirs publics, bien que devant assumer la majorité des coûts de l'intervention, leurs avantages de l'intervention découleront de recettes fiscales supplémentaires et de la réduction des prestations d'aide sociale. Les coûts pour l'ensemble de la société sont obtenus en additionnant les avantages pour les participants et les pouvoirs publics, et en soustrayant la somme des coûts pour les participants et les pouvoirs publics.

L'analyse d'efficacité et l'analyse de rentabilité ne sont pas encore très répandues dans les évaluations du FSE. Toutefois, ces analyses sont très utiles pour décider si une intervention devrait être de nouveau financée ou pour repérer l'intervention la plus efficace parmi une série d'interventions similaires (voir l'exemple dans l'encadré suivant).

Les estimations d'impact issues d'une EIC sont un élément essentiel des analyses d'efficacité ou de rentabilité. Dans le premier cas, elles fournissent les mesures de l'efficacité, tandis que dans le second, elles constituent un élément essentiel pour estimer les avantages monétisés. Il est également clair que les études d'efficacité et les analyses de rentabilité nécessitent la collecte de données précises sur les coûts à partir desquelles les coûts nets peuvent être calculés. Ces activités constituent ce que l'on appelle généralement une étude de coûts. Dans certaines évaluations complexes faisant appel à plusieurs méthodes, les études de coûts sont souvent intégrées à l'évaluation des processus, dans laquelle les instruments de recherche peuvent être adaptés afin de recueillir des données importantes sur les coûts.

Encadré 1 Exemple d'une EIC visant à comparer les avantages et les coûts

Un exemple de l'application d'une méthode contrefactuelle pour estimer les avantages et les coûts figure dans [Bazzoli M. et al., 2018](#). L'étude se concentre sur les programmes de formation professionnelle menés dans la province autonome italienne de Trente entre 2010 et 2011, proposant plus de 300 heures d'activités de formation. Deux groupes de cours ont été évalués: ceux financés par les ressources provinciales (RP) et ceux financés par le Fonds social européen (FSE), qui ont été suivis respectivement par 954 et 205 participants.

Les principales étapes de la mise en œuvre de l'analyse de rentabilité étaient les suivantes: 1) les auteurs ont évalué l'impact des cours de formation sur la probabilité que les participants trouvent un emploi au cours de la période de trois ans suivant le début de leur cours, en appliquant la méthode de l'appariement

par score de propension¹²; 2) l'impact des cours sur les salaires bruts jusqu'à fin 2013 a été estimé; 3) les auteurs ont estimé le montant des recettes fiscales supplémentaires (découlant des effets sur les salaires) et l'argent public économisé du fait de la réduction du nombre de bénéficiaires d'allocations de chômage; 4) les avantages et les coûts des cours ont été comparés.

Plusieurs ensembles de données administratives ont été utilisés: a) les données de suivi relatives aux participants et à leurs caractéristiques; b) les données provenant des registres du chômage des SPE (Centri per l'impiego ou agences pour l'emploi) afin de définir le groupe témoin; c) les données issues de la base de données administratives sur les contrats de travail activés dans la région (Comunicazioni Obbligatorie - COB), les archives des notifications obligatoires de contrat de travail effectuées par les sociétés auprès des services publics de l'emploi utilisées pour déterminer le statut professionnel des membres des groupes traité et témoin, avant et après leur participation au cours de formation; d) les données provenant des archives des recettes fiscales, afin de calculer les salaires des personnes et e) les données de l'Institut italien de la sécurité sociale (INPS) afin d'obtenir des informations sur les allocations de chômage perçues par les personnes.

Après 36 mois, la probabilité d'emploi dans le groupe traité pour les cours RP était supérieure d'environ 5 pp à celle du groupe témoin, tandis que l'impact des cours FSE était beaucoup plus important et s'élevait à environ 28 pp. Dans les trois années qui ont suivi l'intervention, les personnes ayant participé aux cours RP gagnaient en moyenne 2 250 EUR par an de plus que le groupe témoin, tandis que les personnes participant aux cours FSE gagnaient 4 106 EUR par an de plus que le groupe témoin. Les données ont également permis aux auteurs d'estimer les avantages pour les administrations publiques en matière d'augmentation des recettes fiscales et de diminution des dépenses consacrées aux prestations sociales. Les participants au groupe de cours RP ont payé 126 EUR d'impôt sur les revenus de plus que le groupe témoin pour chaque année considérée, tandis que le montant payé par les personnes participant au groupe de cours FSE a été estimé à 318 EUR par an. En ce qui concerne la diminution des allocations de chômage versées par l'administration publique, l'impact des cours a été négligeable, très probablement parce que la valeur monétaire des allocations de chômage dépendait de la durée de l'emploi avant la perte de celui-ci, et que de nombreux participants étaient des jeunes avec une expérience professionnelle limitée.

Le coût des interventions s'élevait à 4 800 EUR par participant pour les cours RP et à 14 500 EUR pour les cours FSE. Les auteurs ont également comparé les coûts et les avantages au niveau individuel pour les deux types de cours; en moyenne, il a été constaté que les coûts étaient supérieurs aux avantages pour ce qui est de la période 2010-2013¹³.

¹² Plus précisément, les auteurs ont appliqué le blocage à l'aide d'un estimateur d'ajustement par régression.

¹³ Un autre exercice «similaire» figure dans [Lammers M. and Kok L., 2021](#).

Chapitre 2. Considérations pratiques pour l'élaboration d'une EIC

Ce chapitre examine les questions pratiques à prendre en considération lors de l'élaboration d'une évaluation. Il convient de le consulter au moment de la planification des activités d'évaluation, de la détermination des interventions à soumettre à une approche EIC et du recensement des questions clés à aborder lors de la conception d'une EIC.

Élaboration d'une EIC

La situation de départ est supposée être celle dans laquelle un gestionnaire de programme d'une AG [ou un gestionnaire d'un organisme intermédiaire (OI) responsable de la mise en œuvre d'une intervention du FSE] détermine les interventions à évaluer, et les stratégies appropriées à intégrer dans une EIC. Il est également supposé que les responsables au sein d'une AG n'effectueront pas d'évaluations eux-mêmes, mais sous-traiteront ou commanderont plutôt des services d'évaluation auprès d'experts externes. Même si l'EIC est effectuée par un sous-traitant, l'AG (ou l'OI) devra planifier et préparer une évaluation d'impact préalablement à la commande.

La stratégie d'évaluation comprenant les différents types d'évaluations, telle que décrite dans le précédent chapitre, doit être exposée dans le plan d'évaluation.

Les plans d'évaluation sont obligatoires pour tous les programmes et doivent être approuvés par le comité de suivi au plus tard un an après l'approbation du programme¹⁴. Les plans doivent être fixés au début de la période de programmation et inclure des dispositions relatives au processus d'évaluation (la gouvernance des évaluations et le lien entre l'évaluation et le suivi), des activités d'évaluation réelles (par exemple, une liste indicative des évaluations à réaliser, la portée de chaque évaluation, les questions principales, les données nécessaires, l'utilisation potentielle, un calendrier indicatif, la structure de gestion), la durée des évaluations, le budget global et le renforcement des capacités d'évaluation.

Les plans d'évaluation sont plutôt d'ordre général, tandis que la planification d'une EIC nécessite une élaboration plus approfondie. Idéalement, celle-ci devrait avoir lieu au moment de la rédaction du plan d'évaluation, certains détails pouvant être ajoutés à un stade ultérieur. Cependant, les AG/OI doivent savoir qu'il est rarement aisé d'établir des liens entre les parties prenantes, ainsi que d'autres dispositifs nécessaires pour une collecte de données liée à une intervention, et que cette procédure doit être planifiée bien à l'avance.

Le présent guide met l'accent sur les manières d'élaborer un système d'évaluation pour des interventions spécifiques qui sont susceptibles d'être soumises à une EIC. Un tel système devrait faire partie du plan d'évaluation ou bien pourrait être mis en place comme une étape opérationnelle dans le suivi d'un plan d'évaluation.

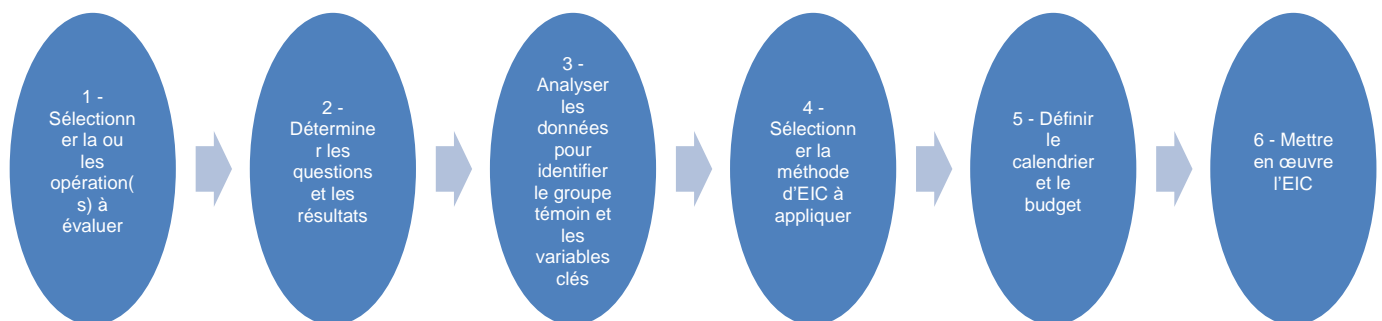
Élaboration d'un système d'évaluation pour des interventions spécifiques

¹⁴ Les plans d'évaluation sont requis conformément à l'article 44 (paragraphe 5 et 6) du règlement (UE) 2021/1060 du Parlement européen et du Conseil. Un plan peut inclure les évaluations prévues de plusieurs programmes, mais tous les programmes doivent être couverts par un plan d'évaluation.

Toutes les interventions financées par le FSE ne peuvent pas faire l'objet d'une évaluation contrefactuelle. Les responsables politiques doivent choisir sur lesquelles concentrer leur attention. Un processus de sélection des interventions qui feront l'objet d'une évaluation d'impact sera nécessaire. Le présent guide propose certains aspects dont les AG devront tenir compte lors de la sélection des interventions appropriées. En outre, l'objectif principal de ce guide est d'aider les responsables de la commande des EIC à examiner en détail certaines des difficultés rencontrées lors de la réalisation d'une évaluation d'impact fructueuse et, ce faisant, à élaborer des systèmes d'évaluation pour les différentes EIC qu'ils envisagent.

Le présent guide considère qu'après avoir sélectionné les interventions qui feront l'objet de l'EIC, les AG devront élaborer un système d'évaluation pour chaque intervention choisie. Le terme «système» est utilisé ici pour distinguer cette activité des «plans» d'évaluation officiels exigés par le règlement (UE) 2021/1060 portant dispositions communes pour la période de programmation 2021-2027. En particulier, le terme «système» désigne l'ensemble des activités normalisées nécessaires à la définition et à la mise en œuvre d'une EIC, qui doivent être préparées avant son lancement, comme illustré dans la figure suivante.

Figure 3 Principale séquence d'activités d'une EIC



Ces systèmes serviront de base pour commander des EIC et poseront les jalons qui permettront aux sous-traitants d'entreprendre une étude rigoureuse et de grande valeur. Le reste du chapitre se penche sur les questions qui doivent être abordées lors de la planification de l'évaluation. Les systèmes d'évaluation doivent être adaptés aux circonstances particulières dans lesquelles l'intervention est réalisée. Il est impossible de prévoir ces circonstances particulières. En conséquence, le présent guide examine les questions qui a) devraient être abordées dans les systèmes, ou b) devraient stimuler la réflexion sur les difficultés que les systèmes devront surmonter.

Questions à aborder lors de la planification de l'évaluation

Après avoir passé en revue quelques-unes des questions qui doivent être abordées lors de la sélection des interventions qui pourraient faire l'objet d'une EIC et déterminer s'il est même possible de procéder à une EIC en fonction des types de données disponibles, l'attention se tourne maintenant vers quelques-unes des questions clés qui doivent être prises en considération lors de l'élaboration d'un système d'évaluation. Ce dernier doit être rédigé avant la commande d'une EIC – ou d'une étude d'évaluation plus large – afin d'être en mesure de préparer le cahier des charges et de désigner

un contractant. Le tableau ci-dessous donne un aperçu du contenu principal d'un tel système d'évaluation.

Tableau 1 Contenu recommandé d'un système d'évaluation

Principales étapes de la préparation et de la mise en œuvre d'une EIC	Contenu
1. Sélection de l'opération ou des opérations à évaluer	<ul style="list-style-type: none"> - Les opérations du FSE, ou les types d'opérations (si possible, faisant partie d'une typologie prédéfinie), qui feront l'objet de l'EIC - Description succincte de la logique de fonctionnement des opérations sélectionnées (objectifs, principaux critères d'admissibilité et population cible, types d'aide, modèle de mise en œuvre, dates approximatives d'activation et d'achèvement, budget indicatif)
2. Détermination des questions d'évaluation et des variables de résultat	<ul style="list-style-type: none"> - Questions posées dans l'évaluation - Fonctionnement attendu des opérations («théorie du changement») et variables de résultat (par exemple, statut professionnel, évolution des salaires, état de pauvreté, note moyenne aux examens, etc.)
3. Analyse des données pour identifier le groupe témoin et mesurer les variables de résultat	<ul style="list-style-type: none"> - Données administratives (par exemple, registres du chômage, registres fiscaux, données administratives en matière d'assurance, registre des établissements scolaires ou base de données étudiante, etc.) ou autres données (enquêtes, mégadonnées, etc.) à utiliser - Variables de la base de données ou de l'enquête à utiliser dans l'EIC - Séries chronologiques des variables nécessaires - Principales règles et questions relatives à l'accès aux données (accessibilité directe par l'AG, nécessité d'un accord avec d'autres administrations, règles et contraintes en matière de protection de la vie privée)
4. Sélection de la méthode d'EIC	<ul style="list-style-type: none"> - Méthode d'EIC possible à adopter (elle peut être détaillée ultérieurement, mais le recours à des essais aléatoires contrôlés ou à d'autres méthodes devrait être indiqué à l'avance afin de favoriser un processus d'évaluation cohérent)
5. Définition du calendrier et du budget	<ul style="list-style-type: none"> - Calendrier des EIC, principales étapes clés: a) choix entre une approche expérimentale ou une approche quasi expérimentale, b) questions d'évaluation détaillées, c) préparation du cahier des charges, d) choix de l'évaluateur, e) préparation des données, f) analyse des données, g) rapports intermédiaires et finaux, h) validation des résultats, i) diffusion des résultats et des enseignements. - Le budget maximal disponible pour l'EIC
6. Mise en œuvre de l'évaluation	<ul style="list-style-type: none"> - Sélection de l'évaluateur - Supervision de la mise en œuvre de l'EIC - Rapports sur l'EIC - Diffusion des résultats (principales parties prenantes à associer, principaux outils)

Encadré 2 EIC intégrée dans un cadre plus large

Un grand nombre d'EIC des interventions financées par le FSE menées dans les États membres sont intégrées dans des cadres d'évaluation plus larges:

- en Allemagne, dans le cadre du PO Bund FSE 2014-2020, l'évaluation contrefactuelle du programme pour l'intégration des chômeurs de longue durée sur le marché du travail financé au titre de la priorité d'investissement (PI) 9.i s'inscrit dans le cadre d'une stratégie d'évaluation plus large prévoyant des rapports intermédiaires annuels qui examinent également des thèmes liés à la mise en œuvre des interventions. Entre 2017 et 2021, quatre rapports ont été produits¹⁵. Il en va de même pour l'évaluation

¹⁵ Voir [Boockmann B. et al., 2017](#) - [Boockmann B. et al., 2018](#) - [Boockmann B. et al., 2019](#) et [Boockmann B. et al., 2021](#). En 2019

des mesures du FSE soutenant l'intégration des chômeurs de longue durée à Bade-Wurtemberg, dans laquelle l'analyse contrefactuelle venait après une analyse plus qualitative axée sur l'évaluation des interventions par les participants¹⁶.

- Dans le Piémont (Italie), l'évaluation contrefactuelle des effets sur l'emploi des chèques-emploi destinés aux personnes vulnérables financés au titre des PI 8.i et 9.i faisait partie d'une évaluation continue plus complète. Deux rapports initiaux examinaient les questions de mise en œuvre et les perceptions subjectives des participants (avril 2018 et février 2019), tandis que deux autres rapports datant de juillet 2019 et de fin 2020 se concentraient sur les effets sur l'emploi en appliquant une méthode contrefactuelle¹⁷.
- Dans les Marches (Italie), l'EIC examinant les effets des interventions pour les chômeurs de longue durée en 2020 a consisté à effectuer une analyse d'impact articulée autour d'un axe thématique, à la suite d'une analyse d'impact plus générale en 2019 des opérations du FSE ciblant les chômeurs (rapport sur le placement)¹⁸.
- En Pologne, le PO Connaissance Éducation Développement pour la période 2014-2020 a commandé un certain nombre d'évaluations (8 rapports¹⁹) entre fin 2015 et mai 2020 afin d'analyser l'aide apportée aux jeunes par le FSE et l'IEJ sous différents angles et en appliquant diverses méthodes d'analyse (par exemple, analyses qualitatives au moyen d'enquêtes et d'entretiens, modèles macro, activités sur le terrain et une approche contrefactuelle).

et 2021, une EIC a été réalisée.

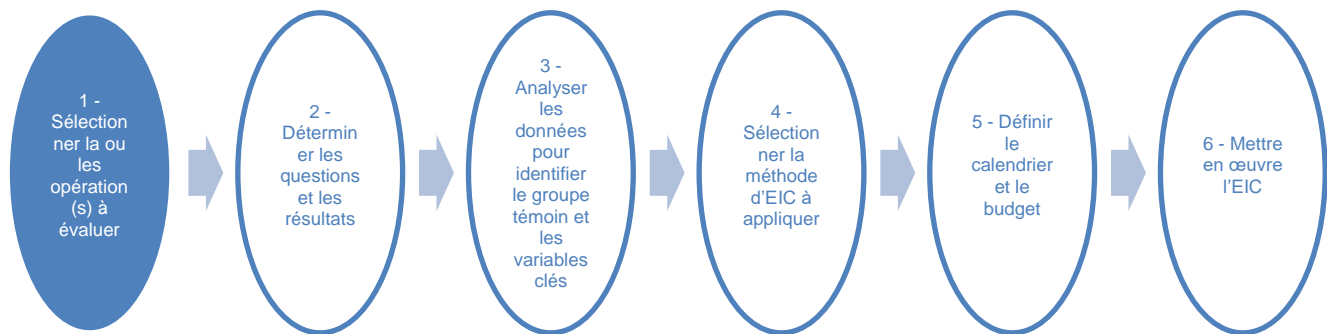
¹⁶ Voir [Hunger K. and Sattler K., 2017](#) et [Scheller F. and Seidel K., 2020](#). En 2020, une EIC a été réalisée.

¹⁷ Voir [Pomatto G., 2017](#) – [Pomatto G., 2019](#) – [Poy S., 2019](#); [Poy S., 2020](#). En 2019 et 2020, une EIC a été réalisée.

¹⁸ Voir [Pompili M., Giorgetti I., 2020](#) - [Pompili M., Giorgetti I., 2020a](#)

¹⁹ Voir [Instytut Badań Strukturalnych - Imapp - IQS, 2015](#) – [Baran J. et al., 2016](#) – [Baran J. et al., 2017](#) – [Baran J. et al., 2018](#) – [Baran J. et al., 2018a](#) – [Palczyńska M. et al., 2019](#) – [Kalinowski H., 2020](#) – [Kalinowski H. et al., 2020](#). En 2017 et 2020, une EIC a été réalisée.

2.1. Sélectionner les interventions pour l'évaluation d'impact



La sélection des interventions pour l'évaluation d'impact comporte trois étapes principales:

Critères de sélection des interventions

1. les questions stratégiques doivent être identifiées;
2. une fois les priorités stratégiques clairement définies, les interventions individuelles doivent être évaluées pour déterminer si elles sont conformes aux exigences de base d'une méthode contrefactuelle, et dans quelle mesure elles sont innovantes et/ou apporteraient une contribution importante à la base de connaissances;
3. il convient de préciser clairement la disponibilité réelle ou potentielle des types de données nécessaires à la réalisation d'une EIC. Cette troisième question s'est avérée jusqu'à présent être un obstacle majeur à la réalisation d'évaluations contrefactuelles des interventions du FSE et mérite donc une attention particulière.

Encadré 3 Questions posées lors de la sélection des interventions pour une EIC

L'EIC n'est pas appropriée pour toutes les interventions et il n'est généralement pas rentable de réaliser des EIC pour l'ensemble d'entre elles. Les autorités de gestion doivent décider de la manière d'allouer les ressources afin d'obtenir le plus grand bénéfice. Le plan d'évaluation devrait refléter ces choix et, lors de la planification des EIC, trois éléments principaux devraient être pris en considération: a) les priorités stratégiques, b) la faisabilité d'une EIC et c) la disponibilité des données nécessaires.

La **stratégie d'évaluation** est influencée par l'échelle, l'élaboration des politiques, les zones d'incertitude et le besoin de connaissances. Les AG devraient poser les questions suivantes:

- Les quantités importantes de fonds alloués à cette intervention rendent-elles particulièrement important le fait de justifier les dépenses? Ces interventions sont relativement aisées à repérer, car elles reçoivent la majeure partie des fonds alloués à chaque objectif spécifique [tel que défini à l'article 2 du règlement (UE) 2021/1057 relatif au FSE+].
- La mesure est-elle l'axe central d'un processus de réforme et les résultats de l'évaluation sont-ils susceptibles de contribuer à un examen critique de l'effort? Ces interventions sont liées aux récentes réformes de la politique du travail et des politiques éducationnelle et sociale; elles ne reçoivent peut-être pas d'importantes sommes d'argent, mais elles sont néanmoins essentielles au succès de la réforme.
- L'intervention est-elle innovante et testée au cours d'une phase pilote ou expérimentale avant d'être étendue? Ces interventions pourraient ne pas recevoir de ressources considérables, mais elles nécessitent une évaluation précoce afin de décider s'il y a lieu de poursuivre et d'étendre l'expérience ou d'y mettre un terme.
- L'intervention se concentre-t-elle sur des domaines qui nécessitent une preuve supplémentaire de

l'efficacité? Ce groupe comprend les interventions qui n'ont pas été évaluées dans le passé ou dont les dernières évaluations sont tellement anciennes qu'elles nécessitent une mise à jour. Les programmes du FSE+ comptent de nombreuses interventions de ce type, car les EIC ne se sont généralisées que récemment et les «effets nets» sont inconnus.

La **faisabilité** concerne à la fois les caractéristiques des interventions et les circonstances dans lesquelles elles sont introduites. Les planificateurs devraient être en mesure de répondre par l'affirmative aux questions suivantes:

- Le traitement appliqué par l'intervention est-il distinct, caractéristique et suffisamment homogène?
- La comparaison entre les groupes traités et les groupes témoins est-elle suffisamment pertinente pour mesurer l'impact?
- La population cible est-elle importante et bien définie?
- La théorie qui lie l'intervention aux résultats escomptés est-elle logique et cohérente?
- Le groupe de traitement peut-il être clairement identifié au sein de la population cible?
- La taille du groupe de traitement sera-t-elle suffisante?
- Un groupe témoin crédible peut-il être repéré?
- L'écart entre l'expérience du traitement et du contrôle peut-il être maintenu pendant une période suffisamment longue pour mesurer l'impact?

Les **données** sont essentielles. L'essence de l'EIC est la mesure, et la mesure a besoin d'informations quantitatives, à la fois sur les groupes traités et témoins, et sur le contexte dans lequel l'évaluation est effectuée. La théorie de l'intervention et la stratégie utilisée pour établir la méthode contrefactuelle déterminent généralement les types précis de données qui sont nécessaires. Lors du choix des interventions qui feront l'objet de l'EIC, l'AG planifiant une EIC doit se poser les questions suivantes:

- Qu'est-il essentiel de savoir sur les membres des groupes cibles et des groupes témoins?
- Qu'est-il essentiel de savoir sur la nature de l'intervention réellement fournie au groupe de traitement?
- Le groupe témoin ne reçoit-il aucun traitement ou reçoit-il un autre traitement? Des données sont-elles disponibles à ce sujet?
- Quelles sont les données disponibles auprès de sources administratives et d'autres sources?
- Les données disponibles décrivent-elles des carrières individuelles?
- Les données individualisées provenant de diverses sources peuvent-elles être liées?

De plus amples informations sur ces questions sont disponibles ci-après dans ce chapitre.

2.1.1. Hiérarchisation des interventions pour l'évaluation d'impact

Avant de donner la priorité à des interventions spécifiques pour l'EIC, des questions stratégiques plus vastes doivent être examinées. Il convient de mettre l'accent sur la sélection des interventions pour lesquelles les évaluations d'impact assurent le meilleur rendement en termes d'apprentissage sur ce qui fonctionne. Les avantages qui découlent d'évaluations rigoureuses et bien conçues profitent non seulement aux AG et aux EM qui les ont commandées, mais également à d'autres EM et leur AG, à d'autres parties prenantes et à la Commission.

Contribution à la justification des dépenses

Étant donné que l'EIC met l'accent sur les questions qui sont essentielles pour les responsables politiques, en particulier ceux qui sont responsables des décisions d'allocation des ressources, il est logique de concentrer les efforts d'évaluation d'impact sur les programmes et les interventions qui sont

Accent mis sur les interventions gourmandes en ressources

particulièrement gourmandes en ressources. Plus un programme ou une intervention spécifique mobilise de temps et d'autres ressources, plus il est important de comprendre s'il ou elle fonctionne, et donc de savoir si les avantages générés sont supérieurs aux coûts encourus. Il peut être nécessaire de reconsidérer des interventions coûteuses qui ne produisent pas de valeur sociale ou économique, tandis qu'une attention et un financement accrus peuvent être accordés à d'autres interventions qui présentent des signes de valeur ajoutée.

Les résultats des évaluations récentes des interventions financées par le FSE au cours de la période de programmation 2014-2020 ont montré qu'il est nécessaire de procéder à des ajustements stratégiques et de mettre davantage l'accent sur les principaux objectifs politiques. Les interventions en faveur de l'emploi et de la mobilité de la main-d'œuvre se sont révélées moins efficaces pour les personnes âgées et les personnes plus éloignées du marché du travail; ces groupes nécessitent des instruments innovants et plus efficaces²⁰. Bien que les interventions sociales varient considérablement du point de vue du coût par participant et du type d'opération, d'une manière générale, aucune analyse de rentabilité systématique n'est réalisée²¹. Les interventions axées sur l'éducation et la formation ont produit des résultats positifs, mais le nombre limité d'études d'impact empêche d'évaluer de manière exhaustive leurs effets à long terme²². Les EIC sont susceptibles de permettre le passage au crible des interventions dans le but de définir les approches les plus efficaces et efficaces pour des groupes cibles donnés, optimisant ainsi le rapport coût-avantage des nouveaux programmes du FSE+.

Contribution d'une intervention à un processus de réforme

Les interventions qui constituent un élément clé d'un programme de réforme plus large bénéficieront souvent d'un financement considérable. Le fait qu'une intervention du FSE est au cœur d'une stratégie d'inclusion sociale, ou constitue une caractéristique essentielle d'un plan de politique active du marché du travail, suscitera naturellement une attention accrue sur cette intervention.

Interventions contribuant à l'innovation politique

Contribution innovante et exploratoire

Les interventions pilotes qui sont nouvelles et innovantes sont des candidates évidentes pour les EIC. Tester les effets des interventions au moyen d'une expérimentation ou d'un essai nécessite à l'évidence une évaluation rigoureuse. L'évaluation par la mise en œuvre d'une EIC bien conçue est d'autant plus importante lorsqu'il existe une volonté claire d'étendre ou de déployer plus largement l'intervention, si elle devait être perçue comme étant fructueuse.

Contribution à l'apprentissage

L'importance de concentrer l'attention et les ressources sur des domaines spécifiques de programmes — et des interventions spécifiques dans ces domaines — est renforcée lorsqu'il existe peu ou pas de preuves de ce qui fonctionne dans le domaine concerné; par exemple, lorsqu'il existe une réelle incertitude quant à la stratégie à appliquer à l'avenir et un risque de

Production de preuves

²⁰ [Fondazione G. Brodolini, Metis GmbH, Applica, Ockham IPS \(2020\).](#)

²¹ [ICF, Cambridge Econometrics and Eurocentre \(2020\).](#)

²² [Ecorys, Ismeri Europa \(2020\).](#)

dépendance excessive à l'égard de preuves qui ne sont peut-être pas directement pertinentes (par exemple, des preuves provenant d'autres pays).

Des évaluations de haute qualité peuvent être considérées comme un bien public. Les avantages qu'elles produisent en termes d'apprentissage s'étendent à des parties prenantes autres que celles d'une AG spécifique. En conséquence, il est important de déterminer les parties prenantes qui devraient bénéficier de l'évaluation d'impact proposée. Ces parties prenantes peuvent être des organismes intermédiaires ou des organes s'occupant d'interventions dans le cadre du même programme, d'autres AG ou OI dans l'État membre concerné, ou des organes et des institutions s'occupant de fonds nationaux ou régionaux. La Commission européenne est une autre partie prenante externe évidente dont il faut tenir compte, et il existe également des parties prenantes dans d'autres EM qui peuvent tirer parti d'une évaluation. La prise en considération des besoins des personnes qui ne sont pas des parties prenantes immédiates est une contribution importante que les responsables politiques et les gestionnaires de programme peuvent apporter à l'apprentissage mutuel.

Lors de la sélection des domaines auxquels il convient de donner la priorité dans l'élaboration des EIC, une dernière réflexion stratégique consiste à tenir compte des interventions qui pourraient mettre en évidence les avantages des EIC et servir de modèles.

Promouvoir les méthodes d'EIC

2.1.2. Sélectionner des interventions qui sont susceptibles d'être soumises à une méthode contrefactuelle

Après avoir examiné les préoccupations stratégiques plus larges susceptibles de motiver la sélection d'interventions particulières pour l'EIC, cette section se penche sur les caractéristiques spécifiques des interventions, qui pourraient rendre celles-ci appropriées à une approche contrefactuelle. Ces caractéristiques sont nombreuses et variées. Certaines fonctionnalités d'une intervention pourraient se prêter à une EIC dans certaines circonstances, mais dans d'autres, elles pourraient entraver les tentatives de mise en œuvre de celle-ci. En conséquence, il n'est pas possible de fournir une liste exhaustive des caractéristiques. Toutefois, les caractéristiques des interventions qui sont plus susceptibles de conduire à une EIC fructueuse méritent d'être mentionnées.

Encadré 4 Types d'interventions et de groupes cibles les plus courants choisis pour les EIC du FSE

La majorité des EIC des interventions du FSE se concentrent sur des politiques actives du marché du travail visant les chômeurs et les sous-groupes parmi eux qui souffrent d'un désavantage spécifique. Le grand nombre d'analyses d'impact des interventions en faveur des jeunes reflète la réglementation relative à l'IEJ, qui impose la réalisation de telles évaluations à des intervalles spécifiques.

Environ la moitié des EIC recensées dans le cadre du projet du Helpdesk Évaluation²³ depuis 2015 concernent l'objectif thématique (OT) 8 «Promouvoir un emploi durable et de qualité et soutenir la mobilité de la main-d'œuvre». De même, dans la plupart des EIC qui examinent les interventions financées au titre de l'OT 9, l'accent est mis sur l'efficacité des interventions en ce qui concerne l'intégration des chômeurs vulnérables sur le marché du travail [par exemple, l'intervention allemande ciblant les chômeurs de longue durée financée au titre du PO Bund FSE pour la période 2014-2020].

Les formes de soutien fourni aux chômeurs les plus fréquemment analysées sont la formation, les stages

²³ Voir https://ec.europa.eu/regional_policy/en/policy/evaluations/member-states/

ou d'autres formes d'expérience professionnelle et les emplois subventionnés. C'est ce qui ressort clairement des exemples analysés dans le présent guide [voir les exemples de l'évaluation italienne dans les Marches, de l'évaluation des cours de formation dispensés aux migrants en Allemagne et de l'évaluation de la formation professionnelle à l'intention des jeunes sans emploi, ne suivant ni enseignement ni formation (NEET) en Lettonie]. L'analyse des services de conseil et de placement est moins courante (comme dans le cas de l'exemple suédois présenté dans ce guide, dans lequel une action pilote visant à renforcer le soutien fourni aux chômeurs par les SPE est évaluée au moyen d'une approche aléatoire). Les interventions du FSE soutenant le travail indépendant et la création d'emploi ne sont pas souvent évaluées, comme il ressort des données du Helpdesk²⁴.

Dans certains cas, les EIC analysent différents types d'interventions de manière groupée, risquant ainsi de mélanger différentes logiques d'intervention et de réduire la fiabilité de la comparaison entre les groupes traités et les groupes témoins.

La tentative d'évaluer les effets des interventions sur les personnes vulnérables mérite d'être soulignée, non seulement en ce qui concerne les résultats en matière d'emploi, mais aussi les résultats difficilement quantifiables: une EIC menée en Allemagne s'est concentrée sur cet aspect, en évaluant l'impact des programmes de création d'emplois pour les chômeurs de longue durée sur les mesures de santé perçues, la satisfaction à l'égard de la vie, le sentiment d'appartenance et le statut social.

Les interventions du FSE dans le domaine de l'éducation sont moins souvent évaluées selon une méthode contrefactuelle. Un nombre plus restreint d'EIC se rapportent à l'objectif thématique 10 plutôt qu'à l'OT 8 ou 9, et l'accent est souvent mis sur les interventions financées au titre du PI 10.IV, en ce qui concerne la formation professionnelle et ses effets en matière d'intégration sur le marché du travail. Ceci s'explique pour deux raisons: l'accessibilité plus limitée aux ensembles de données et des règles plus strictes en matière de protection de la vie privée²⁵. Néanmoins, il y a eu des tentatives en ce sens: en Espagne (Asturies), une intervention dans les écoles secondaires visant à décourager le décrochage scolaire a été évaluée; en Pologne (Podalskie), les EIC ont examiné les effets d'un projet visant à promouvoir la formation professionnelle chez les jeunes étudiants; au Portugal, les bourses en faveur des étudiants de l'enseignement supérieur ont été analysées.

Les EIC peuvent être réalisées à différents niveaux politiques (c'est-à-dire un ou plusieurs axes prioritaires, objectifs spécifiques ou opérations²⁶ dans un programme) et couvrir des programmes nationaux ou régionaux du FSE+, viser des groupes cibles (hommes ou femmes, populations jeunes ou vulnérables, chômeurs de longue durée, etc.) ou des types d'intervention homogènes [par exemple, formation, services pour l'inclusion sociale ou lutte contre le décrochage scolaire (voir encadré précédent)].

Les exemples provenant des États membres indiquent que plusieurs instruments utilisés dans le cadre du FSE+ sont appropriés pour l'EIC, notamment la formation, les incitations à l'emploi et les services liés au marché du travail (par exemple, conseils en matière d'emploi, coaching). D'autre part, les interventions axées sur la rotation des emplois et le partage d'emplois, les incitations au démarrage ou le soutien en faveur de systèmes et de structures, ainsi que les interventions dans les domaines de l'éducation et de l'inclusion sociale, posent davantage de difficultés pour la réalisation d'une EIC.

²⁴ Parmi les exemples figurent: [Borik V. et al., 2015](#) – [Ires Piemonte, 2019](#) – [Openfield, 2019](#).

²⁵ Voir par exemple [Ismeri Europa – Ecorys – Institute for Employment Studies, 2019](#).

²⁶ Conformément à l'article 2, paragraphe 4, du règlement (UE) 2021/1060 du Parlement européen et du Conseil, on entend par «opération»: a) un projet, un contrat, une action ou un groupe de projets sélectionné au titre des programmes concernés; b) dans le contexte d'instruments financiers, une contribution d'un programme à un instrument financier et le soutien financier ultérieur apporté aux bénéficiaires finaux par ledit instrument.

Il est instructif d'examiner quelles interventions sont les plus prometteuses du point de vue d'une EIC en se posant les questions suivantes:

L'intervention est-elle distincte, caractéristique et suffisamment homogène?

Le ou les traitement(s) fourni(s) par une intervention doi(ven)t se distinguer de ceux provenant d'autres interventions. En outre, il doit y avoir un contraste significatif entre les avantages que reçoivent les participants d'une intervention et les avantages dont bénéficient d'autres groupes similaires de personnes. Si les traitements sont flous au point qu'il n'est pas possible de définir un groupe distinct de bénéficiaires, les méthodes contrefactuelles ne sont pas faisables ou pas souhaitables.

*Traitement
clairement
distinguable*

Il devient très compliqué et difficile d'appliquer des méthodes d'EIC si l'état du traitement d'une unité donnée (une entreprise ou un particulier) affecte le résultat potentiel d'autres unités (par ce que l'on appelle des «effets d'équilibre général» plus larges). Dans les programmes de formation, cela peut se produire lorsque des diplômés du programme empêchent d'autres non-stagiaires de trouver un emploi à court terme. Lorsque cela est considéré comme un problème majeur (par exemple dans le cas d'interventions à grande échelle), une analyse macroéconomique peut être nécessaire afin d'évaluer l'ampleur des effets de substitution et de déplacement. Les AG devraient demander des conseils auprès d'experts lorsque ces effets sont susceptibles de se produire.

L'intervention elle-même devrait être relativement homogène. Cela signifie que l'ensemble des participants à une intervention devrait recevoir ou être exposé, de manière générale, au même paquet de mesures. Une trop grande diversité dans les mesures délivrées aux participants d'une même intervention aura plusieurs conséquences pour l'EIC. Premièrement, il serait logique, en réalité, de ne pas parler d'une intervention cohérente, mais plutôt d'interventions avec des processus causals séparés. Deuxièmement, il sera difficile d'interpréter les impacts déclarés en tant qu'effets finaux moyens pour un groupe d'interventions disparates. Troisièmement, l'analyse de sous-groupes pourrait être justifiée, mais si ceux-ci sont trop nombreux au sein d'un groupe de traitement, les limitations dues à la taille de l'échantillon peuvent restreindre la capacité à présenter des résultats exploitables.

*Interventions
homogènes*

Le traitement est-il comparé à la situation sans traitement ou existe-t-il d'autres formes pertinentes de traitement?

Le FSE cofinance des politiques régionales et nationales en matière de participation au marché du travail et d'inclusion sociale. Ainsi, tout système d'EIC doit soigneusement examiner si l'intervention est clairement identifiable et si les particuliers ont la possibilité de recevoir des services d'autres sources de financement et programmes (nationaux ou régionaux). Il est important que les traitements évalués modifient réellement les opportunités ou les ressources disponibles pour les participants par rapport à ce qui est disponible pour les groupes témoins, et que les différences puissent être mesurées et contrôlées.

Les questions de «traitement complexe» ont tendance à dépendre du contexte. Elles compliquent la conception et la mise en œuvre des EIC. Leur présence souligne qu'il est important de planifier minutieusement une

*Traitements
complexes*

évaluation – en élaborant le système d'évaluation – avant de la mettre en œuvre.

Existe-il un groupe cible de grande taille et bien défini?

Les EIC nécessitent des échantillons de grande taille par rapport à d'autres formes d'évaluation. Les groupes cibles composés d'individus en nombre suffisant sont essentiels. Il doit également être possible de repérer des groupes témoins de taille suffisante. Cette question est examinée plus en détail ci-dessous.

Échantillons de grande taille

Il est important que l'intervention envisagée pour l'EIC s'adresse à un groupe bien défini. Si l'on ne connaît pas clairement l'identité des groupes cibles, il est difficile de repérer un groupe témoin significatif. Certaines interventions cherchent délibérément à recruter des personnes pour le traitement au moyen de mécanismes informels, encourageant ainsi des processus qui ne sont pas prédéfinis ou qui sont trop normatifs (par exemple, des cibles sociales difficiles, telles que les NEET ou les groupes défavorisés, peuvent participer au moyen de procédures occasionnelles et au cas par cas); de ce fait, il peut être difficile d'identifier précisément les personnes traitées et le groupe témoin correspondant.

Détermination de l'identité du groupe cible

Existe-t-il un mécanisme de causalité clair?

Comme mentionné précédemment, lors de la présentation des principales méthodes d'évaluation et de la «théorie du changement», il est souvent utile qu'une évaluation fondée sur la théorie ait été réalisée au préalable ou en combinaison avec une EIC. L'élaboration d'une théorie du changement, ou d'une logique détaillée de l'intervention, peut aider les responsables de la conception d'une EIC de nombreuses manières; cela leur permet surtout de déterminer si une intervention dispose d'un mécanisme de causalité cohérent qui la sous-tend. Les interventions ne disposant pas d'un mécanisme de causalité clair et convaincant ne sont pas susceptibles de produire des impacts d'une ampleur suffisante pour être statistiquement recensés grâce à une EIC.

Mécanisme stratégique distinct

Les résultats peuvent-ils être définis quantitativement?

Il est nécessaire d'obtenir des mesures quantifiables des résultats. Ces données et indicateurs peuvent être obtenus à partir de sources administratives ou d'enquêtes ciblées.

Nécessité de mesurer les résultats

Dans certaines circonstances, les interventions peuvent présenter des résultats escomptés qui ont besoin de dispositions spécifiques afin d'être mesurés quantitativement. Par exemple, une intervention pourrait être concernée par un changement d'attitudes, de convictions ou d'opinions. Dans de tels cas, les enquêtes doivent être gérées de manière à mesurer ces changements. Certaines interventions présentent des résultats assez vagues ou mal définis. Encore une fois, l'élaboration d'une logique d'intervention peut contribuer à comprendre les objectifs poursuivis par une intervention et la manière dont elle entend apporter des changements dans les résultats présentant un intérêt.

L'intervention est-elle introduite de manière à pouvoir trouver un groupe témoin significatif?

Mécanisme de sélection du groupe de traitement

Afin d'identifier un groupe témoin pertinent, il importe d'examiner la méthode de sélection des unités traitées (personnes ou entreprises) pour une intervention ou les raisons pour lesquelles elles décident de participer; si la même source de données — par exemple, le même instrument d'enquête — peut être utilisée pour l'échantillon témoin et pour le groupe de traitement; et enfin, s'il est nécessaire de sélectionner des échantillons témoins soumis aux mêmes conditions du marché du travail que le groupe de traitement. Quelques exemples sont mis en évidence dans l'encadré ci-dessous.

Si une intervention est obligatoire et livrée à l'ensemble de la population cible plus ou moins en même temps, il pourrait s'avérer difficile de déceler une partie non traitée de la population cible qui pourra servir de témoin.

Encadré 5 Définir les groupes témoins

Dans les exemples d'EIC examinés dans le guide, la sélection des groupes témoins s'appuyait sur les caractéristiques des interventions (par exemple, les critères d'admissibilité) et sur la disponibilité des données appropriées.

Par rapport à l'expérience des EIC au cours de la précédente période de programmation, l'identification et la sélection du groupe témoin se fondent plus souvent sur des données administratives, en particulier les registres du chômage. La stratégie la plus courante consiste à identifier les personnes témoins potentielles présentant des caractéristiques similaires, inscrites en tant que chômeurs auprès des SPE pendant une certaine période, comme l'exigent les critères d'admissibilité. Cela vaut pour certaines évaluations effectuées en Italie (Marches, province de Trente et Piémont), en Pologne (Lubelskie et Podlaksie), ainsi que pour les évaluations allemandes des cours de formation destinés aux migrants, des programmes de création d'emplois et des mesures d'intégration pour les chômeurs dans le Bade-Wurtemberg.

Dans d'autres évaluations, bien qu'elles ne soient pas couvertes par les exemples présentés en détail dans le guide, les personnes qui se sont inscrites, mais n'ont pas été sélectionnées pour le traitement, ont ensuite été sélectionnées pour constituer le groupe témoin. C'est le cas de l'évaluation italienne de Youth Guarantee, de l'évaluation allemande du projet pilote «Citizen labour funded in 2007-2013» et de l'évaluation du programme PIPOL mis en œuvre en Frioul-Vénétie julienne²⁷. Toutefois, cette stratégie n'est pas souvent adoptée, étant donné que, dans la plupart des pays et des régions, les systèmes d'information de suivi ne contiennent pas d'informations sur les personnes qui ont introduit une demande, mais n'ont pas participé.

L'approche allemande pour l'évaluation des interventions à destination des chômeurs de longue durée²⁸ était différente, dans la mesure où l'évaluation portait sur l'«intention de traiter» et non les «effets moyens du traitement sur les traités». Dans ce cas, le groupe traité était composé de personnes potentiellement admissibles tout au long de la période de mise en œuvre du programme, qu'elles aient ou non été effectivement traitées (à partir de 2015), et le groupe témoin était composé de personnes présentant les mêmes critères d'admissibilité, mais qui étaient des chômeurs de longue durée avant la mise en œuvre du programme (2010-2012).

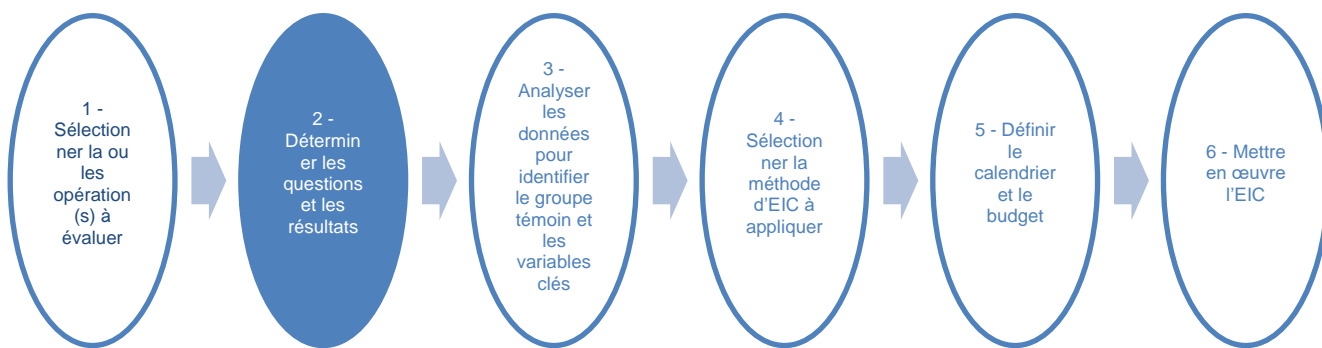
Ce n'est que dans l'exemple suédois, qui a appliqué une approche aléatoire, que le groupe témoin a été identifié de manière aléatoire: par sa conception, le traitement (soutien renforcé fourni par les SPE) a été délivré de manière aléatoire à un groupe de jeunes, tandis que le groupe témoin a reçu le soutien ordinaire offert par les SPE.

Dans les quelques évaluations ciblant les mesures du FSE à destination des entreprises, une distinction entre groupe traité et groupe témoin a été effectuée entre les entreprises bénéficiaires d'un financement et celles qui ne l'étaient pas, comme dans certaines évaluations danoises, qui comparent la performance des entreprises bénéficiaires d'un financement au titre du FSE à celle d'un échantillon d'entreprises ayant des caractéristiques similaires, mais qui n'ont pas reçu le financement.

²⁷ Voir respectivement [Isfol, 2016](#) – [IAW Institut für Angewandte Wirtschaftsforschung – ISG Institut für Sozialforschung und Gesellschaftspolitik, GmbH, 2015](#) – [Ismeri Europa, 2018](#).

²⁸ Voir [Boockmann B. et al., 2019](#).

2.2. Questions d'évaluation et variables de résultat



2.2.1. Quels sont les buts et les objectifs de l'intervention?

Lors de l'élaboration d'un système d'évaluation, il est avant tout conseillé de décrire les buts, les objectifs et les principales caractéristiques de l'intervention.

Dans de nombreux cas, les documents qui définissent les buts et les objectifs de l'intervention existeront déjà. Cependant, dans le cas d'une EIC, il importe d'être précis quant aux résultats et aux changements que l'intervention cherche à atteindre et donc quant aux effets escomptés.

Il est souvent bénéfique de formuler la théorie du changement de l'intervention, qui définit la manière dont ses différents moyens et activités sont censés se rapporter aux réalisations, aux résultats et donc aux impacts.

Combiner la conception des EIC et les enseignements tirés de la logique d'intervention

2.2.2. Quel est l'objectif de l'évaluation?

Lors de l'élaboration d'un système d'évaluation pour une EIC, il importe de considérer avec soin la finalité de l'évaluation. Si l'on ne comprend pas clairement les raisons pour lesquelles l'évaluation est nécessaire, il est peu probable que celle-ci fournira les informations requises. Dans le cadre d'évaluations d'interventions financées par le FSE, une série de questions doit être abordée :

- Quels sont l'objectif et la nature de l'évaluation dans le contexte des exigences réglementaires et des orientations de la CE?
- Qui sont les principales parties prenantes de l'évaluation?
- À quelles fins seront utilisés les résultats de l'évaluation?
- Quelles questions spécifiques l'évaluation devra-t-elle aborder?

Quels sont les buts et la nature de l'évaluation?

Tout d'abord, il convient de définir les **motifs** de la réalisation de l'évaluation. Conformément au règlement (UE) 2021/1060, «l'État membre ou l'autorité de gestion réalise des évaluations des programmes au regard d'un ou de plusieurs des critères suivants: l'efficacité, l'efficience, la pertinence, la cohérence et la valeur ajoutée de l'Union, dans le but d'améliorer la qualité de conception et de mise en œuvre des programmes. Les évaluations peuvent également porter sur d'autres critères pertinents, tels que le caractère inclusif, la non-discrimination et la visibilité, et couvrir plusieurs programmes»²⁹. Comme indiqué ci-dessus, les conclusions d'une EIC se rapportent généralement à l'efficacité (la mesure dans laquelle les résultats

Le but et la nature de l'évaluation

²⁹ Article 44, paragraphe 1, du RPDC.

attendus ont été atteints) et à l'efficacité (le rapport coût-efficacité ou coût-avantage de l'intervention).

Plus généralement, la CE encourage les États membres à suivre des approches orientées sur les résultats dans l'élaboration de leurs politiques et à effectuer des évaluations qui répondent à leurs exigences internes en matière de portée, de conception et de calendrier. À cet égard, dans les programmes du FSE+, les EIC peuvent également être mis en œuvre pour répondre à des questions d'évaluation spécifiques ou conformément à la politique d'évaluation nationale.

Deuxièmement, la **nature de l'évaluation** doit être déterminée.³⁰

- Les évaluations **d'impact** examinent les effets d'un programme ou d'un groupe de programmes au regard des priorités de l'UE et nationales (elles peuvent estimer l'impact macroéconomique du FSE ou se concentrer sur des politiques et des thèmes spécifiques ou des priorités horizontales telles que l'enfance et l'égalité des chances).
- Les évaluations **des processus (ou de la mise en œuvre)** soutiennent la mise en œuvre d'un programme, en analysant les progrès réalisés et les méthodes de mise en œuvre et en fournissant des recommandations en vue d'améliorer le programme.

En principe, l'approche contrefactuelle peut être appliquée aux évaluations d'impact, tandis que les évaluations des processus nécessitent d'autres méthodes (voir également la

Figure 1 ci-dessus).

Le RPDC n'impose pas un nombre spécifique d'évaluations d'impact pour la période de programmation 2021-2027, contrairement à celle de 2014-2020. Il demande à la place une stratégie d'évaluation permettant d'évaluer la manière dont le soutien accordé par les Fonds européens a contribué à la réalisation des objectifs du programme en ce qui concerne tous les profils stratégiques principaux³¹. En outre, le RPDC ne détermine pas les priorités ou interventions qui devraient être soumises à l'évaluation, mais laisse cette décision à l'AG et au plan d'évaluation. Cela signifie que la stratégie d'évaluation de chaque programme doit définir la combinaison d'évaluations d'impact et des processus, ainsi que les priorités et interventions sur lesquelles elle doit concentrer ses principaux efforts. Une évaluation finale et générale d'impact doit être produite d'ici juin 2029³², mais aucune autre restriction ne limite le calendrier des autres évaluations qui seront décrites dans le plan d'évaluation.

Qui sont les destinataires principaux?

Il convient de définir les destinataires de l'évaluation. Selon la nature de l'évaluation, il peut s'agir de responsables politiques, d'AG et de gestionnaires de programmes, d'autres AG ou organismes d'exécution dans l'État membre, ou d'autorités nationales ou régionales qui gèrent des programmes similaires. Lorsque les données sont fournies par des institutions qui ne relèvent pas de la gestion du programme, il y a lieu de considérer ces organismes (les «propriétaires des données») comme des parties prenantes. Il est important

Détermination des parties prenantes

³⁰ Voir Commission européenne, 2007.

³¹ Article 44, paragraphe 1, du RPDC.

³² Article 44, paragraphe 2, du RPDC.

d'associer toutes les principales parties prenantes au sein d'un comité de pilotage de l'évaluation afin d'établir la coresponsabilité du processus de conception et de réalisation de l'évaluation, ainsi que certains experts en évaluation provenant d'institutions universitaires ou publiques, afin qu'ils fournissent des conseils techniques à l'AG.

À quelles fins seront utilisés les résultats de l'évaluation?

Après avoir recensé les destinataires de l'évaluation, il est possible de déterminer l'utilisation qui sera faite des résultats. En pratique, cette étape peut être réalisée en associant le comité de pilotage à l'élaboration des questions d'évaluation et aux discussions relatives au cahier des charges.

Les résultats de l'EIC contribuent fréquemment à la prise de deux décisions essentielles, à savoir:

- si une intervention existante doit être maintenue, et
- si un nouveau type d'intervention doit être mis en œuvre plus largement (c'est-à-dire étendu).

Dans le premier cas, une EIC peut tenter d'évaluer l'efficacité d'un programme existant ou en cours dans lequel les budgets sont sous pression et où d'autres utilisations des ressources concernées sont possibles. Dans cette situation, il est peu probable que l'intervention ait été évaluée préalablement au moyen de scénarios contrefactuels.

Dans le deuxième cas, des interventions peuvent être soumises à des contraintes de mise en œuvre. Par exemple, il se peut qu'une intervention soit appliquée dans une région ou une zone spécifique d'un EM, ou uniquement pour une période limitée. Dans de tels cas, les résultats d'une EIC peuvent servir à déterminer si l'intervention concernée est efficace et peut donc être réalisée de manière utile ailleurs. Les interventions dans de telles situations sont considérées comme étant mises à l'essai, ou testées, avant un déploiement à plus grande échelle.

À quelles questions convient-il de répondre?

Après avoir déterminé les objectifs de l'intervention, ainsi que le but et l'utilisation finale de l'évaluation, et avoir recensé clairement les destinataires, il devrait être possible d'énoncer de manière précise les questions dont l'EIC devra traiter. Dans bien des cas, plusieurs destinataires et parties prenantes voudront que l'EIC explore des questions d'ordre causal. L'AG ou l'évaluateur devrait rassembler ces questions au moyen d'une vaste consultation, en tenant compte des différents points de vue et suggestions des personnes intervenant à différents stades de l'intervention. Par la suite, l'AG et/ou l'évaluateur doi(ven)t hiérarchiser les questions et axer l'EIC sur celles qui sont plus importantes et appropriées.

Voici certains aspects qui pourraient être abordés lors de la finalisation d'une liste de questions de recherche clés pour une EIC:

- L'intervention a-t-elle produit ou favorisé les résultats escomptés à court, moyen et long terme? Les effets à court terme diffèrent-ils de manière significative de ceux à long terme? La priorité doit être accordée aux questions qui traitent de ces aspects.
- Est-il possible de mesurer quantitativement les résultats? L'EIC doit s'appuyer sur un ensemble adéquat de données (données administratives

*Questions de
recherche clés*

ou enquête directe envoyée aux participants) indépendant d'une source privilégiée.

- Dans quelle mesure les changements dans la situation des participants, ou dans le contexte socio-économique, peuvent-ils être attribués aux interventions? Ce type de question nécessite une mesure des effets finaux de l'intervention à des fins de comparaison avec les participants témoins et les indicateurs contextuels.
- Les effets de l'intervention étaient-ils les mêmes pour tous les membres du groupe cible? Par exemple, l'impact d'une intervention visant les chômeurs de longue durée était-il le même pour les hommes et les femmes? Et pour les personnes âgées de moins de 25 ans ou de plus de 50 ans? La capacité des EIC à examiner les effets dans différents sous-groupes est renforcée si le nombre de personnes dans les groupes traités et témoins est suffisamment élevé.
- Le rapport coût-efficacité de l'intervention est-il satisfaisant (par rapport à d'autres interventions)? Et quel est son rapport coût-avantage? Cette analyse de l'efficacité nécessite qu'une EIC fournisse une mesure précise des effets, de même que des informations précises sur les coûts et avantages directs et indirects.
- Dispose-t-on d'un grand nombre d'informations et de connaissances sur les effets d'interventions similaires? Dans quelle mesure ces connaissances sont-elles applicables à l'intervention examinée? Lorsqu'une politique est bien connue et que ses effets ont fait l'objet d'une analyse approfondie, il peut être utile de concentrer les questions d'évaluation sur des aspects spécifiques de la politique et d'éviter ainsi de répéter d'autres analyses. Un examen détaillé de la littérature existante peut éclairer les décisions à cet égard et constituer, en tout état de cause, un soutien utile lors de la conception de l'EIC.

Il convient de se faire une idée claire de l'éventail de questions de recherche qu'une EIC doit aborder avant d'en effectuer la commande. L'examen des questions que l'évaluation abordera est un élément clé de tout système d'évaluation.

Il est essentiel de hiérarchiser les questions et de ne pas céder à la propension courante de surcharger une évaluation avec trop de questions. Il est difficile de trouver le juste équilibre entre garantir que l'évaluation est pertinente pour un éventail de parties prenantes ayant des intérêts divergents et rendre l'évaluation souple. Une évaluation devant traiter d'un trop grand nombre de questions de recherche peut perdre de vue ses objectifs et finir par répondre de manière peu optimale à un grand nombre de préoccupations. Il s'agit souvent de «faire mieux avec moins» – la hiérarchisation est une phase essentielle dans le processus de planification de l'évaluation.

Hiérarchisation des questions

Pour hiérarchiser les questions d'évaluation, il est nécessaire d'exclure les doublons et d'attribuer une note à chaque question selon des principes pertinents. Il peut s'agir: de l'importance et de l'engagement réel des parties prenantes qui ont formulé la question, de l'adéquation et de la congruence de la question avec la théorie du changement du programme, de la pertinence de la question pour l'objectif général de l'évaluation, de la faisabilité de la question par rapport aux données, au temps et aux ressources disponibles³³.

³³ Voir par exemple, [centre de prévention et de contrôle des maladies CDC, 2013](#).

Le classement ainsi obtenu permettra de trier les questions en fonction de leur importance et de sélectionner les questions les plus pertinentes.

Dans certains cas, il est possible de regrouper des questions d'évaluation sous une question plus générale; par exemple, «quel a été l'effet net de l'intervention?». D'autres aspects, tels que les effets sur différents groupes de participants, peut-être même à des périodes différentes, pourraient être imbriqués. Toutefois, ce processus d'imbrication doit toujours donner lieu à une série de questions «gérables» et faisables.

Quels critères d'évaluation peuvent être associés à l'évaluation?

La relation entre les critères et les questions d'évaluation a été mentionnée ci-dessus, mais elle mérite d'être clarifiée davantage. Des critères d'évaluation (efficacité, efficacité, valeur ajoutée européenne, cohérence, etc.) sont nécessaires pour attribuer une valeur à tout ensemble de données probantes et pour parvenir à évaluer une politique (efficacité, etc.); tandis que des questions d'évaluation sont nécessaires pour rendre explicite la demande de la personne qui commande l'évaluation et se concentrer sur les principales questions stratégiques en jeu. Toutefois, les critères d'évaluation et les questions d'évaluation sont liés. Chaque question renvoie généralement à un critère spécifique et cette relation est importante tant pour les AG que pour les évaluateurs, car elle relie la conception de l'évaluation, nécessairement fondée sur les questions, aux exigences énoncées dans le RPDC pour la période 2021-2027, qui sont liées de manière différente aux critères d'évaluation.

*Critères
d'évaluation et
questions
d'évaluation*

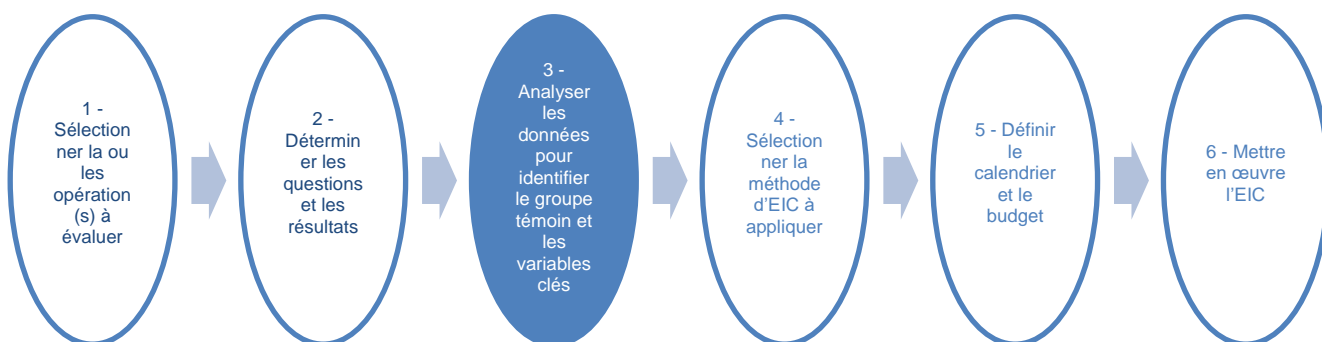
Quelques exemples de questions types d'évaluation regroupées avec leurs critères correspondants sont présentés ci-dessous. Ils ont été sélectionnés et adaptés à partir de la «boîte à outils pour une meilleure réglementation»³⁴ (Outil n° 47 – *Critères et questions d'évaluation*) de la Commission européenne afin de présenter des exemples cohérents avec des questions qui nécessitent une analyse d'impact et, si possible, une EIC:

- *Exemples typiques de questions relatives à l'efficacité*
 - Quels ont été les effets quantitatifs de l'intervention?
 - Dans quelle mesure ces effets/changements peuvent-ils être imputés à l'intervention?
 - Dans quelle mesure les facteurs influençant les réalisations observées peuvent-ils être liés à l'intervention de l'UE?
- *Exemples typiques de questions relatives à l'efficacité*
 - Quel est le rapport coût-efficacité de l'intervention?
 - Dans quelle mesure les coûts de l'intervention sont-ils justifiés eu égard aux changements/effets qui en résultent?
 - S'il existe des différences importantes de coûts (ou d'avantages) entre les territoires, quelles en sont les causes? Comment ces différences sont-elles liées à l'intervention?
- *Questions typiques relatives à la valeur ajoutée européenne*
 - Quelle est la valeur supplémentaire résultant de l'intervention ou des interventions du FSE+ par rapport à ce qui est produit par des interventions nationales et/ou régionales similaires?

³⁴ Voir [Commission européenne, 2017](#).

- Quelles seraient les conséquences les plus probables de l'interruption ou de la suppression de l'intervention actuelle du FSE+?

2.3. Données permettant d'identifier le groupe témoin et de mesurer les variables de résultat



2.3.1. Les données adéquates sont-elles disponibles ou peuvent-elles être mises à disposition?

Les discussions qui ont eu lieu avec les AG et les experts en évaluation de toute l'UE suggèrent que l'accès aux données adéquates est l'une des principales difficultés rencontrées dans la réalisation des EIC; une considération pratique essentielle est de savoir si les types de données requis sont disponibles. Cette section présente une catégorisation simplifiée des types de données nécessaires et les sources à partir desquelles ces données peuvent être obtenues, ou les types de collectes de données primaires qui peuvent s'avérer nécessaires. La question cruciale de la protection des données est également abordée.

Auparavant, il convient d'aborder un point important en ce qui concerne la planification appropriée. Dans une certaine mesure, des tentatives visant à mettre en œuvre des EIC ont été contrariées dans le passé par un manque de données, car des plans adéquats n'avaient pas été mis en place assez tôt. En ce qui concerne les interventions existantes, il importe d'identifier les membres des groupes traités et non traités et de mettre en place des mécanismes de collecte de données auprès de ces membres, étant donné qu'ils seront au centre de l'évaluation. Pour les nouvelles interventions, il y a lieu de prendre des mesures à un stade précoce afin de garantir la collecte des types de données adéquats à des moments opportuns.

Planification de la collecte des données

Quels sont les types de données nécessaires?

De manière générale, trois types de données sont nécessaires pour procéder à une EIC. Dans certains cas, une seule source de données peut contenir un ou plusieurs de ces types de données. Il s'agit: des documents relatifs aux groupes traités et témoins, des enregistrements des résultats et des registres de données contextuelles.

Principales sources de données

- **Documents relatifs aux groupes traités et témoins:** il est nécessaire de disposer de sources de données qui permettent aux évaluateurs d'identifier chaque unité constitutive des groupes traités et témoins (entreprises, personnes ou, éventuellement, zones géographiques).

- **Enregistrements des résultats:** comme illustré dans les figures 6 et 7 du chapitre 3 du présent guide, les EIC nécessitent que les résultats soient mesurés tant pour les groupes traités que pour les groupes témoins. Idéalement, il convient de rassembler des données sur les résultats des deux groupes en utilisant les mêmes méthodes de collecte et d'effectuer les mesures des résultats aux mêmes moments.
- **Registres de données contextuelles:** des données sont nécessaires pour permettre de sélectionner un groupe témoin correspondant de manière adéquate au groupe traité et de contrôler toute différence qui subsiste entre les deux groupes dans l'analyse. Il importe de recueillir autant de données que possible sur les facteurs et les caractéristiques unitaires qui peuvent être liés au choix de participer à une intervention et aux résultats potentiels, en particulier les indicateurs de résultats mesurés avant l'intervention. Les données contextuelles peuvent également inclure celles qui décrivent des marchés locaux du travail (par exemple, les taux de chômage locaux ou des mesures de resserrement du marché du travail) et celles qui permettront une analyse par sous-groupe.

Le tableau 2 ci-dessous présente ces trois types de données et suggère des sources à partir desquelles elles pourraient être collectées. Des exemples de données utilisées pour les EIC du FSE sont présentés dans le premier encadré ci-dessous, tandis que le deuxième encadré fournit des exemples d'ensembles de données intégrés utilisés dans l'analyse de questions relatives au marché du travail et dans les évaluations de politiques du marché du travail.

Tableau 2 Types de données et sources

Types de données	Sources
Documents relatifs au groupe traité	<ul style="list-style-type: none"> - Registres de participation à des interventions (généralement conservés par les bénéficiaires) - Données de suivi du FSE+ (caractéristiques des interventions, dates de début et d'achèvement, registres de données relatives à l'orientation des sujets, registres de demande)
Documents relatifs au groupe témoin	<ul style="list-style-type: none"> - Données administratives, telles que les registres de sécurité sociale, de l'éducation et d'allocation chômage - Registres de participation (ceux qui étaient admissibles pour participer, mais qui ne l'ont pas fait pour des raisons autres que l'admissibilité³⁵) - Enquêtes nationales existantes, telles que l'enquête sur les forces de travail
Enregistrements des résultats (nécessaires pour les groupes traités et témoins)	<ul style="list-style-type: none"> - Données administratives: par exemple, les registres de la sécurité sociale et du chômage peuvent également être utilisés pour réunir des mesures de résultats (perception des allocations/des prestations de sécurité sociale), les registres nationaux fiscaux et d'assurance (résultats en matière de salaires et d'emploi) - Dossiers administratifs de l'enseignement (tests normalisés sur les compétences acquises, taux d'obtention des diplômes, inscriptions et présences) - Recensement officiel des entreprises ou registres fiscaux (productivité ou chiffre d'affaires avant et après la formation interne ou de nouvelles embauches) - Données relatives à l'emploi ou à la production issues des statistiques officielles (dans les analyses contrefactuelles territoriales pour mesurer les niveaux d'emploi et de PIB) - Enquêtes sur mesure des groupes traités et témoins
Données contextuelles/variables de contrôle (nécessaires pour les groupes de traitement et de contrôle)	<ul style="list-style-type: none"> - Systèmes administratifs: par exemple, les registres d'allocation fournissant des informations sur les antécédents en matière d'allocations avant l'intervention; les registres nationaux fiscaux et d'assurance, les salaires historiques et les registres d'emploi - Statistiques officielles sur le marché du travail ou l'éducation (par exemple, enquête sur les forces de travail, données de base au niveau régional ou national qui fournissent également des microdonnées au niveau individuel pour une élaboration spécifique) - Enquêtes auprès des groupes témoins et traités. Lorsque les règles de l'intervention sont claires, les groupes témoins peuvent être définis préalablement et des données de référence peuvent être recueillies - Systèmes de suivi: dans certains cas, des systèmes de suivi peuvent être utilisés pour collecter les mesures à partir des groupes traités et des groupes témoins, par exemple les procédures de candidature lorsque les candidats ayant échoué peuvent être utilisés comme groupes témoins.

Encadré 6 Exemples de données utilisées pour les EIC

Parmi les exemples examinés en détail dans le présent guide, les évaluations allemandes sont probablement celles qui reposent sur les ensembles de données administratifs les plus riches. Les biographies intégrées sur l'emploi (BIE) intègrent un certain nombre de sources administratives

³⁵ Ces participants sont considérés comme «admissibles, mais non admis»; la non-admission dépend généralement de la finalité du financement ou de certaines autres causes externes (par exemple, maladie des enseignants, panne des transports, etc.). Ces sujets peuvent représenter un groupe témoin préférentiel, car ils sont admissibles et veulent participer au même titre que les participants réels; toutefois, ils ne sont parfois pas assez nombreux pour constituer un groupe témoin, ou leurs informations ne sont pas enregistrées.

d'information gérées par l'agence fédérale pour l'emploi et contiennent des informations sur l'emploi (à l'exception du travail indépendant) et les épisodes de chômage d'une personne, sur les caractéristiques sociodémographiques, sur les paiements de transfert (allocations de chômage) et sur la participation aux politiques actives du marché du travail. Les BIE ont une longue histoire et ont nécessité des efforts et des investissements continus pour développer et gérer l'ensemble de données. De plus amples détails sur les BIE sont fournis dans l'encadré 7³⁶. Dans l'évaluation allemande des programmes de création d'emploi, ces données ont été complétées par une enquête structurée en trois vagues afin de recueillir des informations sur les résultats non quantifiables (tels que perçu par les personnes) à différents stades au cours des interventions.

Dans les évaluations italiennes examinées en détail dans le guide, les données utilisées étaient similaires. Les dossiers des chômeurs inscrits auprès des SPE ont été utilisés pour identifier les groupes témoins potentiels, tandis que les données administratives sur les changements d'emploi (Comunicazioni obbligatorie, COB), qui enregistrent tous les contrats de travail des entreprises et des employeurs publics (à l'exclusion du travail indépendant), ont été utilisées pour mesurer les variables de résultat. Compte tenu du fait que les COB sont gérées au niveau régional, la qualité des informations varie. Toutefois, des normes nationales ont été introduites ces dernières années et la qualité s'est améliorée. Dans une évaluation (pour la province de Trente), l'évaluateur a également été en mesure de fusionner ces données avec les données de déclaration fiscale fournies par l'INPS (l'institut national de sécurité sociale), afin de mesurer l'impact en matière de salaires individuels.

En Pologne, les évaluations ont utilisé des données administratives issues des registres du chômage des SPE pour définir les groupes témoins. Toutefois, ces données ne contiennent pas d'informations sur le statut professionnel ou les antécédents d'emploi des personnes, et les évaluations ont dû s'appuyer sur une approximation (en annulant le statut de chômeur dans les SPE) ou collecter des informations au moyen d'enquêtes auprès d'un échantillon de personnes traitées et non traitées.

En Lettonie, l'évaluateur est parvenu à fusionner deux principaux ensembles de données administratives. Les données de l'agence nationale lettone pour l'emploi (SEA) ont fourni des informations sur les participants et les non-participants inscrits en tant que chômeurs à des dates précises, et les données du service national des impôts (SRS) ont fourni des informations sur les conditions d'emploi à différentes dates ainsi que sur les revenus de particuliers. Cela a permis à l'évaluateur d'évaluer les effets sur le plan tant de la probabilité d'être employé à des moments différents que des revenus.

Le seul cas de randomisation, l'exemple suédois, a utilisé à la fois des données administratives pour mesurer les variables de résultat et une enquête auprès des bureaux et des intermédiaires de l'emploi pour mesurer l'intensité du soutien et les types de soutien fournis.

Les évaluations dans le domaine de l'éducation ont utilisé des données administratives pour identifier le groupe témoin et pour évaluer les variables de résultat, à l'exception de la Pologne, qui a dû s'appuyer sur une enquête pour mesurer les variables utilisées comme résultats.

Dans l'ensemble, les exemples montrent l'importance de données administratives appropriées pour les EIC, idéalement à la fois pour la détermination des groupes témoins et la mesure des variables de résultat. De ce point de vue, les AG qui planifient les EIC devraient s'assurer préalablement que les données administratives nécessaires à la réalisation des évaluations sont disponibles, en mettant en place des mesures appropriées pour résoudre les problèmes potentiels d'accessibilité, d'intégration des données ou d'autres aspects problématiques.

Encadré 7 Exemples de bases de données intégrées pour les EIC

L'ensemble de données longitudinales sur les demandeurs d'emploi (JLD) en Irlande

L'ensemble de données longitudinales sur les demandeurs d'emploi (JLD) est une base de données administrative gérée par le ministère de la Protection sociale (DSP).

³⁶ Une expérience similaire est représentée par la base de données irlandaise «Jobseekers Longitudinal Dataset» (ensemble de données longitudinales sur les demandeurs d'emploi, JLD), qui a intégré les informations sur les prestations d'assistance sociale et d'assurance sociale, les programmes du marché du travail et les antécédents d'emploi dans une base de données unique. Le JLD a été utilisé, entre autres, dans l'évaluation du programme d'activation JobBridge financé au cours de la période de programmation 2007-2013. Voir [Indecon, 2016](#).

Le JLD est une tentative ambitieuse d'adapter les données administratives à des fins de recherche. L'élaboration de la base de données a débuté il y a 10 ans, après que le DSP a commandé auprès de l'université de Dublin une étude préliminaire sur la gestion du Live Register³⁷ et, plus généralement, des données relatives au marché du travail. L'étude a formulé plusieurs suggestions pour améliorer la collecte des données et a révélé certains problèmes (par exemple, la duplication des données dans divers systèmes de données, des informations manquantes, etc.).

Le JLD intègre plusieurs sources d'information: les données sur les paiements et les données administratives du DSP, par exemple les prestations d'assistance sociale et d'assurance sociale versées à la population active contenues dans le Live Register et les données relatives aux programmes de politique active du marché du travail gérés par le DSP; les données relatives aux programmes du marché du travail gérés par SOLAS, l'organisme national d'éducation et de formation; les données collectées par les autorités fiscales (direction des impôts). En ce qui concerne les interventions du FSE, le JLD couvre partiellement les interventions financées par le Fonds, mais plus particulièrement celles financées par le DSP et SOLAS³⁸.

Les données provenant des sources susmentionnées sont réorganisées sous la forme d'une série d'épisodes, un épisode commençant lorsque la personne entame une période de chômage et s'achevant lorsqu'elle trouve un emploi, une autre activité ou un programme de formation; ensuite, lorsque le statut de la personne change à nouveau, un nouvel épisode commence. Le début d'une période de chômage coïncide avec le moment où la personne dépose une demande en vue de recevoir l'indemnité de chômage et l'allocation de demandeur d'emploi³⁹. Dans le JLD, l'un des avantages de cette structure par épisodes est que les périodes contiguës de perception de l'indemnité de chômage et de l'allocation de demandeur d'emploi peuvent être reliées et représentées sous forme d'un épisode de chômage. Il est à noter que des épisodes, comme indiqué ci-dessus, peuvent se chevaucher et que le chercheur est confronté à la difficulté de donner la priorité à un épisode plutôt qu'un autre.

Le JLD suit les demandes d'assistance sociale, les antécédents d'activation, de formation et d'emploi depuis 2004 et comprend environ 13 millions d'épisodes individuels d'assistance et d'emploi concernant environ 2 millions de personnes. Chaque épisode comporte une date de début et de fin, ainsi qu'un code opérationnel⁴⁰ qui permet la désignation de la situation d'une personne au cours de chaque période. Le JLD couvre un large ensemble de «variables»: le genre, l'âge, l'état civil, la nationalité, le niveau d'instruction, la profession antérieure, les antécédents d'emploi et de chômage (durée et nombre d'épisodes) et les caractéristiques des emplois (secteur par exemple), les antécédents de formation en situation de chômage (type, durée et nombre d'épisodes), le type de prestation (JA, JB), le nombre d'enfants, les personnes à charge, le type de prestations familiales (c'est-à-dire les allocations pour adulte et enfant à charge, pour adulte uniquement, etc.), les salaires, l'imposition et la situation géographique⁴¹.

En outre, les identifiants individuels relient le JLD à d'autres données administratives; par exemple, pour l'évaluation de JobsPlus, le JLD a été mis en correspondance avec une base de données de suivi distincte contenant des informations détaillées sur les dates de début et de fin de JobsPlus, le nombre de jours dans le Live Register au début de JobsPlus et le type de traitement.

Il existe deux principaux canaux pour accéder aux données du JLD: la recherche sous contrat, lorsque le DSP lance une procédure d'appel d'offres; ou à la demande de chercheurs possédant des références adéquates. Dans ce dernier cas, les demandes d'accès sont examinées au cas par cas et, en cas

³⁷ Le Live Register contient des informations sur les personnes qui s'inscrivent auprès des bureaux locaux du DSP pour bénéficier de l'indemnité de chômage (JB), de l'allocation de demandeur d'emploi (JA) ou de divers autres droits légaux.

³⁸ Les programmes dans des domaines tels que la justice (initiatives visant à donner aux jeunes de nouveaux centres d'intérêt, projets d'insertion des jeunes, projets pour l'intégration et l'emploi des migrants), l'éducation (accès au troisième cycle et alphabétisation des adultes), la communauté (centres de formation communautaires) et la défense (programme de soutien à l'emploi des forces de défense) ne sont pas compris dans le JLD.

³⁹ L'indemnité de chômage est versée toutes les semaines par le DSP aux personnes sans emploi, se trouvant en chômage complet ou qui travaillent à temps partiel parce que leur employeur a réduit leurs heures. Elle s'applique aux demandeurs d'emploi qui ont versé des cotisations d'assurance sociale (PRSI) au taux approprié, un nombre suffisant de cotisations dans l'année d'imposition concernée et au total au moins 104 cotisations. L'allocation de demandeur d'emploi est versée sur la base des moyens aux demandeurs d'emploi qui sont au chômage et qui ne peuvent pas bénéficier de l'indemnité de chômage, ou dont les droits à cette indemnité ont expiré.

⁴⁰ Ces variables représentent une combinaison d'activités d'activation/de formation, de demandes d'assistance et de périodes d'emploi.

⁴¹ Il convient d'attirer l'attention sur le fait que certaines données renvoient à des périodes différentes, par exemple les salaires sont enregistrés sur une base annuelle, tandis que les prestations de protection sociale sont mises à jour chaque semaine.

d'acceptation, un accord juridiquement contraignant relatif au partage des données entre les chercheurs/instituts et le DSP doit être signé. Toutes les données sont pseudonymisées et seuls les champs de données pertinents pour le thème de recherche sont transmis. L'accès aux données par le chercheur ou l'institut de recherche est limité et les données doivent être supprimées à la fin du projet de recherche ou d'évaluation. Un manuel de codes est fourni aux chercheurs avec les données du JLD, bien qu'il soit nécessaire d'améliorer sensiblement la documentation.

Le JLD a été utilisé pour diverses activités de recherche et évaluations au fil des ans: Back to Education Allowance (2015), le programme d'activation JobBridge (2016), Back to Work Enterprise Allowance (2017), JobPath (2019) et Jobplus (2020) font partie des initiatives qui ont été sélectionnées à des fins d'évaluation.

La structure et le contenu du JLD se sont améliorés depuis les phases initiales. Toutefois, sa tenue à jour, son amélioration et la correction de ses lacunes nécessitent une activité continue, par exemple en ajoutant des informations plus détaillées sur l'éducation, en intégrant davantage de données spécifiques aux programmes (contenu, achèvement, etc.) et en améliorant les données sur les salaires. Un prochain projet devrait apporter des améliorations au JLD en ce qui concerne: sa mise à jour régulière (par exemple, tous les mois) avec automatisation et test du circuit de données, et l'ajout ou le remplacement des sources de données afin de garantir que le soutien à l'emploi et à la formation, les autres programmes de protection sociale et les données en temps réel sur les salaires sont pleinement couverts⁴².

L'institut de recherche sur l'emploi (IAB) en Allemagne et les biographies intégrées sur l'emploi (BIE)

En Allemagne, l'accès aux données à des fins scientifiques a été amélioré à la suite des réformes du marché du travail au cours de la période 2003-2005, en mettant l'accent sur l'évaluation des politiques et les recommandations de la Commission concernant l'«amélioration de l'infrastructure d'information entre la science et la statistique» en vue d'établir un centre de données de recherche au sein de chaque producteur public de microdonnées. En conséquence, l'agence fédérale pour l'emploi a créé un centre de données de recherche au sein de l'Institut de recherche sur l'emploi (für Arbeitsmarkt- und Berufsforschung, IAB) en 2004. L'IAB est chargé d'extraire les données des processus administratifs afin d'élaborer des ensembles de données utiles pour la recherche empirique.

Plus de 15 ensembles de données sont gérés par l'IAB et mis à la disposition de la communauté scientifique. Les données proviennent des données administratives du système de sécurité sociale, de processus internes de l'agence fédérale pour l'emploi et d'enquêtes menées par l'IAB⁴³.

En ce qui concerne les systèmes de sécurité sociale, tous les employeurs sont tenus de déclarer un certain nombre d'éléments et de caractéristiques de leur personnel, et ces données fournissent un large ensemble d'informations sur les antécédents d'emploi des personnes. Les données administratives relatives aux procédures internes de l'agence fédérale pour l'emploi comprennent l'assurance chômage obligatoire, les allocations de chômage et les périodes d'inscription correspondantes, les réunions de consultation avec les chômeurs, les offres de placement et les mesures actives du marché du travail. L'IAB combine les données dans un ensemble de données complet et unique, appelé «biographies intégrées sur l'emploi» (BIE). La collecte de ces données administratives a débuté en 1975, bien que toutes les variables ne soient pas disponibles pour l'ensemble de la période d'observation. Les BIE peuvent être intégrées à plusieurs données issues d'enquêtes mises en œuvre par l'IAB, telles que: le panel d'établissements de l'IAB, l'enquête sur les vacances d'emploi en Allemagne de l'IAB, l'enquête sur les pratiques allemandes de gestion et organisationnelles (GMOP); l'étude de panel «Marché du travail et sécurité sociale» (PASS); l'étude «Travail et apprentissage dans un monde en mutation»; l'enquête auprès du personnel «Primes, augmentations de salaire et équité» (BLoG); les données liées relatives aux employeurs et employés de l'IAB (LIAB); le panel «WeLL» – enquête auprès du personnel pour le projet «Formation complémentaire dans le cadre de l'apprentissage tout au long de la vie».

L'IAB met régulièrement à jour ses produits de données et propose différents échantillons de ces sources abondantes de données administratives à des fins de recherche. Pour chaque produit de données, l'IAB

⁴² Nous tenons à remercier Frank Humphreys, Ciaran Judge, Saidhbhín Hardiman et Krzysztof Gigon du DSP pour les informations qu'ils ont fournies sur le JLD.

⁴³ En 2011, le centre de rapprochement des données, un projet commun avec l'Université de Duisburg-Essen financé par la fondation allemande pour la recherche, a été créé au sein de l'IAB; le centre vise à simplifier le rapprochement d'ensembles de données sans identifiant spécifique.

met à disposition une documentation détaillée en allemand et en anglais. La base juridique de l'accès aux données figure dans le livre du code social allemand (plusieurs versions). Plus précisément, quatre types d'accès aux données pour la communauté scientifique sont envisagés: les dossiers de campus (entièrement anonymisés et utiles uniquement pour l'enseignement); les dossiers à usage scientifique sont des microdonnées anonymisées transmises à des institutions scientifiques en Allemagne et dans les États membres de l'UE pour des projets de recherche dans le domaine de la recherche sur le marché du travail et non à des fins d'enseignement ou de recherche commerciale. La sécurité des données doit être garantie par l'institution scientifique qui demande les données; les données peu anonymisées contenant des informations plus détaillées ne sont accessibles que sur place. L'IAB fournit des lieux de travail distincts dans un environnement informatique sécurisé à Nuremberg et dans différentes régions en Allemagne, aux États-Unis et au Royaume-Uni. Les chercheurs bénéficient d'un accès direct aux données, mais ils ne peuvent obtenir les résultats de leurs programmes qu'après l'examen des divulgations par le personnel de l'IAB; l'exécution à distance signifie que les chercheurs préparent leurs programmes à l'aide de données artificielles et les téléchargent dans l'application de soumission de travaux (JoSuA). Dans le cadre de ce processus, les chercheurs ne voient jamais les données originales et ne reçoivent que leurs résultats. Des formulaires de demande normalisés sont disponibles pour tous les accès aux données. Après l'approbation d'une demande, un contrat régissant l'utilisation des données pour un projet spécifique au cours d'une période donnée est conclu entre l'institution du chercheur et l'IAB. Le contrat précise les règles en matière de protection des données et des sanctions sévères s'appliquent en cas de violation. Certains des ensembles de données gérés par l'IAB ne sont disponibles que pour une utilisation sur place (par exemple, les ensembles de données liés).

Sur la base des données produites par l'IAB, les politiques actives du marché du travail sont régulièrement évaluées en appliquant les méthodes empiriques les plus récentes et, dans certains cas, les résultats de ces évaluations ont conduit au remaniement des politiques⁴⁴. L'IAB permet aux chercheurs d'accéder à ses ensembles de données non seulement en Allemagne, mais aussi dans d'autres pays, et le nombre d'utilisateurs est en augmentation constante (par exemple, en 2016, près d'un tiers des accords relatifs à l'utilisation des données provenaient d'un site non allemand).

En ce qui concerne l'utilisation des données administratives pour les EIC et les questions pratiques associées, voir également le guide européen élaboré par les chercheurs du JRC, [European Commission, 2020](#).

Quels sont les éventuels problèmes en matière de protection des données?

Il peut s'avérer difficile d'obtenir des données qui identifient les personnes ou les entreprises ayant participé à des interventions financées par le FSE+⁴⁵. Les EIC nécessitent des microdonnées — à savoir des données qui comportent des observations sur chaque unité constitutive dans les groupes traités et témoins.

Le règlement (UE) 2021/1057 relatif au FSE+ (annexe I) requiert des données sur les participants, avec une ventilation par genre, statut sur le marché du travail, groupe d'âge, niveau d'instruction et groupe vulnérable (migrants, minorités, handicapés, autres personnes défavorisées). Le RPDC et le règlement relatif au FSE+ pour la période 2021-2027 imposent légalement aux AG de recueillir et de traiter les données à caractère personnel sous la forme de registres individuels des participants. En outre, l'article 17, paragraphe 6, du règlement (UE) 2021/1057 relatif au FSE+, pour ce qui est du suivi et des indicateurs, précise que: «*[L]orsque des données sont disponibles dans des registres ou des sources équivalentes, les États*

Règlement relatif au FSE+ en ce qui concerne la gestion des données

⁴⁴ On peut citer à titre d'exemple une évaluation de l'accord d'intégration obligatoire entre le demandeur d'emploi et le gestionnaire de dossiers. À l'aide d'une expérience aléatoire sur le terrain et en suivant les biographies sur le marché du travail des personnes participant à l'expérience, l'IAB a été en mesure de démontrer que pour certains groupes de chômeurs, la réglementation obligatoire était contreproductive et devrait être remplacée par une application plus souple de l'instrument (van den Berg et al. 2016).

⁴⁵ Voir [Isméri Europa – Ecorys – Institute for Employment Studies, 2019](#).

membres peuvent permettre aux autorités de gestion et aux autres organismes chargés de la collecte des données nécessaires au suivi et à l'évaluation du soutien général au titre du volet FSE+ relevant de la gestion partagée d'obtenir des données à partir de ces registres ou sources équivalentes, conformément à l'article 6, paragraphe 1, points c) et e), du règlement (UE) 2016/679».

Ces règles, énoncées dans les dispositions communes et les règlements relatifs au FSE+, facilitent la consultation et l'utilisation des données à caractère personnel nécessaires au suivi du FSE+ et, dans le cas d'une EIC, à la définition du groupe traité. Toutefois, l'accès aux données à caractère personnel nécessaires à la formation des groupes témoins et le traitement des données relatives aux groupes traités et témoins doivent être conformes au règlement (UE) 2016/679 (règlement général sur la protection des données, RGPD), qui couvre le transfert général et l'utilisation des données à caractère personnel, y compris les catégories particulières de données⁴⁶, au sein de l'Union européenne. L'encadré suivant fournit une vue d'ensemble des principaux contenus et obligations régis par le RGPD.

Encadré 8 Cadre réglementaire de l'UE relatif au traitement des données à caractère personnel

La législation pertinente relative au traitement des données à caractère personnel en Europe se compose principalement du règlement (UE) 2016/679 et des lignes directrices et mesures adoptées par les autorités compétentes telles que le comité européen de la protection des données (CEPD). Cette législation fixe de nombreuses conditions et limitations en ce qui concerne le traitement des données à caractère personnel afin de protéger les droits et les libertés des personnes concernées. En tout état de cause, le besoin de protection et de garantie qui découle des dispositions et des règlements doit trouver un équilibre avec la nécessité de ne pas entraver la recherche scientifique et, en réalité, devenir un atout pour son évolution. C'est pour cette raison, et dans ce cadre juridique, que l'Union européenne a prévu, sous certaines conditions, des exceptions afin de permettre la conduite des activités de recherche et la diffusion des résultats à condition que le droit fondamental des personnes concernées, à savoir le droit au respect de la vie privée, soit garanti.⁴⁷

RÈGLEMENT (UE) 2016/679 (RÈGLEMENT GÉNÉRAL SUR LA PROTECTION DES DONNÉES)

Le règlement général sur la protection des données (RGPD) est entré en vigueur le 24 mai 2016 et est devenu pleinement applicable dans tous les États membres le 25 mai 2018. Le RGPD s'applique «*au traitement de données à caractère personnel, automatisé en tout ou en partie, ainsi qu'au traitement non automatisé de données à caractère personnel contenues ou appelées à figurer dans un fichier*»⁴⁸.

S'agissant du champ d'application territorial, le règlement s'applique tant au traitement de données à caractère personnel effectué par des personnes situées sur le territoire de l'Union européenne qu'aux cas dans lesquels le traitement porte sur des personnes concernées se trouvant dans l'Union européenne, même lorsque le responsable du traitement ou le sous-traitant est situé en dehors de l'UE.

Du point de vue du fond, et par rapport à la législation antérieure, le règlement renforce les droits des personnes concernées et impose une série d'obligations aux responsables du traitement selon une logique fondée sur l'analyse des risques et le principe de responsabilité. En outre, le règlement prévoit une série

⁴⁶ Le RGPD a cessé d'employer le terme «données sensibles» et fait désormais référence à des «catégories particulières de données». Conformément à l'article 9, ces dernières comprennent: «[l]es données [...] qui révèle[nt] l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que [...] [l]es données génétiques, [l]es données biométriques [traitées] aux fins d'identifier une personne physique de manière unique, [l]es données concernant la santé ou [l]es données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique».

⁴⁷ Les autres textes de la législation européenne qui contribuent au cadre juridique du traitement des données à caractère personnel incluent: le règlement (UE) n° 557/2013 relatif aux statistiques européennes en ce qui concerne l'accès aux données confidentielles à des fins scientifiques; et le règlement (UE) 2018/1725 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel par les institutions, organes et organismes de l'Union et à la libre circulation de ces données.

⁴⁸ Article 2 du règlement (UE) 2016/679 du Parlement européen et du Conseil.

d'exigences à remplir par le responsable du traitement, lesquelles ne figuraient pas dans la législation antérieure: parmi ces exigences, l'analyse d'impact relative à la protection des données (articles 35 à 36), la minimisation des opérations de traitement selon les critères de protection des données dès la conception et de protection des données par défaut (article 25), l'adoption du registre des activités de traitement (article 30) et la désignation du délégué à la protection des données (DPD) (articles 37 à 39).

Il importe de souligner les grands principes énoncés à l'article 5 du règlement, auxquels doivent se conformer les personnes qui traitent des données à caractère personnel, en particulier les principes suivants:

- a) licéité, loyauté et transparence: les données à caractère personnel sont traitées de manière licite, loyale et transparente;
- b) limitation de la finalité: les données sont collectées pour des finalités déterminées, explicites et légitimes, et ne peuvent pas être traitées ultérieurement d'une manière incompatible avec ces finalités. En cas de traitement à des fins statistiques ou de recherche scientifique, le responsable du traitement adopte les mesures de garantie et de protection nécessaires.
- c) minimisation des données: seules les données strictement nécessaires pour atteindre des finalités spécifiques devraient être collectées;
- d) exactitude: les données doivent être exactes et, si nécessaire, mises à jour;
- e) limitation de la conservation: les données à caractère personnel sont conservées sous une forme permettant l'identification des personnes concernées pendant une durée n'excédant pas celle nécessaire à la réalisation des finalités pour lesquelles elles sont collectées. À cette fin, les données à caractère personnel peuvent être conservées pour des durées plus longues dans la mesure où elles sont traitées uniquement à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques, conformément à l'article 89 du RGPD. Dans de tels cas, les données font l'objet de mesures techniques et organisationnelles appropriées mises en œuvre pour garantir les droits et les libertés des personnes concernées;
- f) intégrité et confidentialité: les données à caractère personnel sont traitées de façon à garantir une sécurité appropriée de ces données, y compris la protection contre le traitement non autorisé ou illicite et contre la perte, la destruction ou les dégâts d'origine accidentelle, à l'aide de mesures techniques ou organisationnelles appropriées;
- g) responsabilité: il s'agit de l'un des principes les plus importants énoncés dans le règlement. Il précise que le responsable du traitement est responsable du traitement des données et peut démontrer la mise en œuvre de toutes les mesures requises.

Cet ensemble de principes constitue la structure principale sur laquelle repose le RGPD et détermine une série d'obligations à remplir par le responsable du traitement et le sous-traitant.

TRAITEMENT DES DONNÉES À CARACTÈRE PERSONNEL DANS LE CADRE DE L'EIC: CONDITIONS, LIMITES ET PRINCIPALES QUESTIONS

La réalisation d'une évaluation d'impact contrefactuelle implique le traitement de grandes quantités de données, y compris de données à caractère personnel. En ce qui concerne le cadre réglementaire décrit ci-dessus, il est important de connaître les conditions, les limitations et les principales questions auxquelles les EM et les AG sont confrontés lorsqu'ils effectuent des EIC.

Base juridique et finalités du traitement

Pour procéder au traitement de données à caractère personnel, au moins l'une des conditions de licéité énoncées à l'article 6 du RGPD doit être remplie⁴⁹. Si, en revanche, les données à traiter constituent des «catégories particulières de données», il convient également de se référer à l'article 9 du RGPD.

Dans le cadre d'une EIC, l'**intérêt public** au sens de l'article 6, paragraphe 1, point e), du RGPD («l'exécution d'une mission d'intérêt public ou relevant de l'exercice de l'autorité publique dont est investi

⁴⁹ Ces conditions sont les suivantes: a) la personne concernée a consenti au traitement de ses données à caractère personnel pour une ou plusieurs finalités spécifiques; b) le traitement est nécessaire à l'exécution d'un contrat auquel la personne concernée est partie ou à l'exécution de mesures précontractuelles prises à la demande de celle-ci; c) le traitement est nécessaire au respect d'une obligation légale à laquelle le responsable du traitement est soumis; d) le traitement est nécessaire à la sauvegarde des intérêts vitaux de la personne concernée ou d'une autre personne physique; e) le traitement est nécessaire à l'exécution d'une mission d'intérêt public ou relevant de l'exercice de l'autorité publique dont est investi le responsable du traitement; f) le traitement est nécessaire aux fins des intérêts légitimes poursuivis par le responsable du traitement ou par un tiers, à moins que ne prévalent les intérêts ou les libertés et droits fondamentaux de la personne concernée qui exigent une protection des données à caractère personnel, notamment lorsque la personne concernée est un enfant.

le responsable du traitement») semble être la base juridique la plus appropriée pour le traitement des données. Cette base juridique doit être définie par le droit de l'Union ou le droit national, comme précisé à l'article 6, paragraphe 3, du RGPD. L'«intérêt public» représente à l'évidence les obligations incombant aux autorités de gestion définies dans l'article 17, paragraphe 6, du règlement (UE) 2021/1057 relatif au FSE+ en ce qui concerne l'utilisation de données «disponibles dans des registres ou des sources équivalentes». En outre, le RPDC dispose dans son article 4 que «[l]es États membres et la Commission ne sont autorisés à traiter des données à caractère personnel que lorsque cela est nécessaire pour remplir les obligations qui leur incombent au titre du présent règlement, en particulier pour le suivi, l'établissement de rapports, la communication, la publication, l'évaluation [...]». Les législations nationales peuvent également conférer aux responsables du traitement des données des pouvoirs similaires aux fins de la gestion et du traitement des données dans l'intérêt public.

D'autres bases juridiques d'une EIC peuvent être mentionnées dans d'autres conditions énumérées à l'article 6, paragraphe 1, en particulier:

- le **consentement** de la personne concernée (article 6, paragraphe 1, point a), du RGPD). Le consentement, par exemple, peut constituer une base juridique appropriée lorsque les données destinées à l'EIC sont collectées au moyen d'une enquête et que les personnes concernées peuvent facilement donner leur consentement au traitement (voir l'article 7 du RGPD sur le consentement). En général, le consentement est plus complexe s'il n'est pas planifié suffisamment à l'avance; il peut être considéré comme une base juridique «résiduelle» de l'EIC lorsque d'autres bases juridiques ne sont pas applicables;
- le traitement, lorsque celui-ci est nécessaire au **respect d'une obligation légale** à laquelle le responsable du traitement est soumis (article 6, paragraphe 1, point c), du RGPD, en référence à l'article 17 du règlement relatif au FSE+). Cette base juridique doit être établie par le droit et peut concerner des entités privées ou publiques; il se pourrait, par exemple, qu'en raison de prescriptions légales spécifiques, une entité privée ou publique responsable d'un ensemble de données soit tenue de collaborer avec l'autorité de gestion dans le cadre de l'EIC.

En outre, le responsable du traitement **peut utiliser les données collectées pour d'autres finalités si celles-ci sont compatibles avec les finalités initiales**. À cet égard, le responsable du traitement devra évaluer les conditions énoncées à l'article 6, paragraphe 4, et à l'article 5, paragraphe 1, point b), du RGPD. Ces dispositions sont particulièrement pertinentes en cas de traitement des données à des fins de recherche scientifique ou à des fins statistiques. Elles autorisent l'utilisation de données administratives pour des finalités différentes des finalités initiales et qui ne nécessitent pas de consentement spécifique à la nouvelle utilisation, sous réserve du respect des règles en matière de protection fixées à l'article 89, paragraphe 1, principalement de pseudonymisation (voir ci-dessous). Ces dispositions, par exemple, peuvent être pertinentes dans le cas de l'utilisation de données issues du registre du chômage dans une EIC.

En raison de la spécificité des EIC, le responsable du traitement qui collecte les données et l'entité qui effectue les recherches peuvent être différents. Dans ce cas, il doit exister une condition qui justifie le transfert des données et permet au destinataire de réaliser l'EIC.

Anonymisation et pseudonymisation: traitement à des fins statistiques

En vertu de l'article 89 du RGPD, le traitement des données effectué à des fins d'intérêt public, dans le cadre de recherches scientifiques ou à des fins statistiques, prévoit des garanties appropriées pour les droits et libertés des personnes concernées et respecte en particulier le principe de «minimisation». Autrement dit, en recourant à des techniques de pseudonymisation⁵⁰.

Lorsque les finalités peuvent être atteintes par des opérations de traitement ultérieures qui ne permettent pas ou plus l'identification des personnes concernées, ces finalités sont atteintes en anonymisant les données, puis en les traitant sous une forme agrégée. Dans les EIC, par définition, les résultats sont agrégés et ce risque est absent, à moins que les ensembles de données initiaux ne soient publiés pour des raisons scientifiques. Dans ce cas, les ensembles de données doivent être anonymisés (voir l'exemple

⁵⁰ La pseudonymisation signifie que les données à caractère personnel sont traitées de manière à ne plus pouvoir être attribuées à une personne spécifique sans le recours à des informations supplémentaires. L'anonymisation désigne le fait de traiter les données à caractère personnel de façon à rendre impossible l'identification des personnes à partir de ces données. Un aperçu des techniques de pseudonymisation figure à l'adresse <https://www.enisa.europa.eu/publications/pseudonymisation-techniques-and-best-practices>.

dans l'encadré ci-dessous).

Conservation et traitement sécurisé des données

L'un des principaux aspects du traitement des données est la conservation des données. La législation ne précise pas comment les données doivent être conservées, mais les principes susmentionnés exigent que la conservation et le traitement soient toujours liés à la finalité de la recherche. Lorsque la finalité du traitement est atteinte et que la conservation des données n'est plus nécessaire, elle prend fin. Cette règle générale doit être précisée dans la déclaration de confidentialité qui est fournie aux personnes concernées lorsqu'elles sont enregistrées dans des ensembles de données administratives ou, compte tenu de l'impossibilité probable d'informer personnellement toutes les personnes concernées, d'autres moyens de communiquer les informations peuvent être trouvés (par exemple, en publiant une page d'information contenant la politique de confidentialité relative aux activités de recherche sur le site web de l'autorité de gestion).

En particulier, une fois l'EIC réalisée, les résultats mis à disposition et agrégés à des fins statistiques, la finalité est considérée comme atteinte. À ce stade, les données doivent être anonymisées (si elles doivent être réutilisées à un stade ultérieur et à des fins différentes) ou supprimées. Lorsque la législation ou la réglementation l'autorise, la conservation peut être illimitée dans le temps, mais uniquement dans le cadre d'une référence réglementaire explicite.

L'autre condition requise par le RGPD est que le traitement soit assorti de mesures de sécurité adéquates et appropriées. L'article 32 du règlement prévoit que quiconque effectue le traitement (responsable du traitement ou sous-traitant) met en œuvre les mesures techniques et organisationnelles appropriées pour assurer un niveau de sécurité adapté au risque. Cela signifie qu'aucune mesure de base n'est prescrite et que les mesures doivent être définies au cas par cas en fonction des risques que présente chaque opération de traitement.

OBLIGATIONS LÉGALES

Information des personnes concernées

La première obligation du responsable du traitement est d'informer les personnes concernées. Cette obligation est énoncée à l'article 13 du RGPD pour les données qui sont collectées auprès de la personne concernée, mais lorsque ce n'est pas le cas, la référence est l'article 14. La personne concernée doit être informée des finalités et des méthodes du traitement, de la base juridique, des durées de conservation et des droits qu'elle peut exercer. Dans le cas d'une EIC, ces informations sont fournies au moment de la collecte des données qui seront ultérieurement utilisées à des fins de recherche statistique et d'évaluation, par exemple dans les ensembles de données de suivi du FSE ou des registres du chômage. Lorsque cette communication est pratiquement impossible, comme indiqué ci-dessus, et en particulier dans le cas de personnes inscrites dans des registres utilisés pour l'EIC, un site web ou d'autres outils d'information généraux peuvent être utilisés.

Gouvernance de la relation entre les différentes entités participant au traitement

En général, les activités de recherche supposent l'intervention de plusieurs organismes et, dans de tels cas, les relations entre ceux-ci doivent être réglementées au moyen d'accords spécifiques appelés «accords relatifs au traitement des données»⁵¹. Ces relations doivent être définies au cas par cas en tenant compte des contributions des différentes parties prenantes. Voici des exemples de relations possibles:

- la relation *responsable du traitement – responsable du traitement* survient lorsque les entités collaborent à la mise en œuvre d'un projet, bien qu'elles le fassent dans des conditions différentes et dans le cadre de missions différentes, chacun en poursuivant sa propre finalité de traitement distincte;
- la relation *responsable du traitement – sous-traitant* survient lorsqu'une entité (le responsable du traitement) détermine les moyens et les finalités du traitement et fait appel à une autre entité (le sous-traitant) pour réaliser certaines activités de traitement. Dans ce cas, la référence et les obligations figurent à l'article 28 du RGPD;
- *responsables conjoints du traitement*: ce type de relation est défini à l'article 26 du RGPD, qui dispose ce qui suit: «[L]orsque deux responsables du traitement ou plus déterminent conjointement les finalités et les moyens du traitement, ils sont les responsables conjoints du traitement». Dans ce cas, les parties

⁵¹ Par exemple, c'est le cas lorsque le responsable du traitement et les personnes effectuant la recherche sont différents; dans ce cas, il est nécessaire de conclure un accord qui régit les relations entre les parties pour l'accès aux données et la conduite de la recherche.

doivent déterminer leurs responsabilités respectives en matière de respect de la réglementation de manière transparente, notamment en ce qui concerne les droits des personnes concernées.

Analyse d'impact relative à la protection des données

L'article 35 du RGPD dispose que, lorsque le traitement des données à caractère personnel est susceptible d'engendrer un risque élevé pour les droits et libertés des personnes physiques, le responsable du traitement doit effectuer, avant de procéder au traitement, une analyse d'impact⁵². Une analyse d'impact dans les EIC ne peut être importante que dans le cas du traitement à grande échelle de catégories particulières de données relatives à des personnes concernées. Dans ce cas, le responsable du traitement au sein de l'administration peut effectuer l'analyse d'impact conformément aux règles et aux outils mis à disposition par l'autorité nationale chargée de la protection des données.

PERSPECTIVES: L'ACTE SUR LA GOUVERNANCE DES DONNÉES

Le cadre décrit ci-dessus reflète l'état actuel de la technique et les règles en vigueur dans l'Union européenne. Le RGPD a été introduit spécifiquement pour permettre une avancée en matière de protection des données à caractère personnel et pour créer un système commun dans l'ensemble de l'UE, conformément aux évolutions technologiques actuelles et à la société actuelle des données. D'autres réformes sont en cours d'élaboration et pourraient avoir une incidence significative sur la collecte de données à caractère personnel dans les années à venir.

L'acte sur la gouvernance des données est actuellement toujours au stade de proposition au sein de la Commission européenne⁵³. L'objectif explicite de cet acte est de promouvoir la disponibilité des données utilisables en renforçant la confiance dans les intermédiaires de données et en améliorant les mécanismes de partage des données dans l'ensemble de l'UE. Les données à caractère personnel sont susceptibles d'être soumises à la nouvelle norme et leur utilisation à des fins statistiques et de recherche pourrait être davantage étendue grâce à l'introduction d'un nouvel acteur: l'intermédiaire de partage de données à caractère personnel. Une telle mesure pourrait contribuer à créer un environnement plus favorable pour les EIC, ce qui faciliterait l'accès à grande échelle à de grandes quantités d'informations.

Il est évident que les règles du RGPD n'excluent pas la mise en œuvre d'une EIC; en conséquence, certaines étapes procédurales et opérationnelles de base doivent être respectées dans tous les EM afin de sécuriser l'EIC. Par exemple, dans un cas normal où les données de suivi (personnes traitées) sont combinées avec des données provenant d'un registre public (groupe témoin) et où un évaluateur externe effectue l'EIC, les étapes les plus importantes sont les suivantes:

*Principales étapes
pour mieux se
conformer au
RGPD*

1. l'AG conclut un accord avec les administrations responsables des données (par exemple, registre du chômage, registre fiscal, etc.) nécessaires à l'identification du groupe témoin et à l'analyse des groupes traité et témoin. L'AG vérifie que l'utilisation de ces données est conforme à l'article 6, paragraphe 1, point e), ou à l'article 6, paragraphe 4, et à l'article 5, paragraphe 1, point b), du RGPD, si un consentement explicite n'a pas été recueilli;
2. l'AG conclut des accords avec les autres entités (les propriétaires de données et l'évaluateur) en vue de réglementer les flux d'informations et les responsabilités mutuelles conformément aux règles du RGPD. Le contrat de service entre l'évaluateur et l'AG doit comporter une clause spécifique sur la protection des données; dans le cas d'autres administrations, un protocole d'accord ou des procédures nationales

⁵² Pour plus d'informations, veuillez consulter les lignes directrices publiées par le CEPD à l'adresse <https://ec.europa.eu/newsroom/article29/items/611236>.

⁵³ La proposition relative à l'acte sur la gouvernance des données a été adoptée par la Commission à la fin de l'année 2020. L'état d'avancement de l'acte sur la gouvernance des données peut être consulté à l'adresse suivante: <https://eur-lex.europa.eu/legal-content/FR/HIS/?uri=CELEX%3A52020PC0767>.

spécifiques peuvent réglementer la protection des données lors de la mise en œuvre des EIC.

3. Sur la base du contrat susmentionné, l'AG reçoit les données sous forme pseudonymisée du propriétaire des données et les transfère à l'évaluateur aux fins du traitement des groupes traités et témoins, conformément au RGPD⁵⁴.
4. La conservation des données à caractère personnel des groupes traités et témoins n'est conforme aux règles en matière de conservation des données que si les données sont conservées pour la durée et les finalités de la recherche et dans le respect des règles fondamentales de sécurité, comme indiqué dans le RGPD. Ces règles sont respectées par toutes les entités participant à l'EIC.

Les pratiques nationales sont généralement conformes au RGPD, mais elles diffèrent d'un EM à l'autre. Selon les évaluateurs, les règles nationales en matière de protection des données continuent de poser de sérieux obstacles à l'utilisation des microdonnées. Cela peut s'expliquer par le temps nécessaire pour adapter les règles et les habitudes nationales à la dernière version du RGPD, par des divergences d'interprétation du RGPD au sein de différentes administrations nationales, par la surréglementation, qui rend le RGPD plus strict dans certains pays, ou par d'autres erreurs d'interprétation. Dans certains pays, une initiative spécifique des AG ou des autorités nationales serait nécessaire pour surmonter ces obstacles dans l'esprit du RGPD et du règlement relatif au FSE+. L'encadré ci-dessous explique comment la région de Vénétie a résolu certaines des difficultés les plus courantes. Une bonne pratique consiste à prendre contact avec l'autorité nationale chargée de la protection des données afin de discuter des accords proposés avant de les conclure.

Encadré 9 Protection et échange des données

Un bon exemple de l'accès à des données à caractère personnel anonymisées dans un délai relativement court figure dans une évaluation contrefactuelle présentée dans l'étude «Pilot and feasibility study on the sustainability and effectiveness of results for European Social Fund participants using counterfactual impact evaluations»⁵⁵ réalisée par la DG EMPL en 2019.

L'évaluation, l'un des quatre cas présentés dans l'étude, s'est concentrée sur les interventions du FSE (PI 9.i) réalisées en Vénétie en 2015/2016 à destination des chômeurs de longue durée. Les données de suivi sur les participants et les types de mesures ont été combinées avec des données sur les périodes d'emploi avant et après les interventions du FSE (appelées «Comunicazioni obbligatorie»). Le processus a duré environ deux mois (de fin juillet à fin septembre) à compter de la première réunion concernant les exigences en matière de données entre l'autorité de gestion et l'institution détenant les données, l'institut régional Veneto Lavoro. Veneto Lavoro était chargé d'anonymiser les données et tous les ensembles de données transmis ont pu être facilement reliés entre eux⁵⁶ au moyen d'un identifiant commun unique fourni par Veneto Lavoro.

Bien que les organisations participantes ne commandaient pas directement l'évaluation, l'expérience a été positive en raison de deux facteurs principaux: une coopération solide (bien qu'informelle), entre les bureaux régionaux de l'autorité de gestion et l'institut régional Veneto Lavoro, nécessaire en vue d'aborder et de résoudre les questions relatives à la protection des données, et l'existence d'une base de données

⁵⁴ Si les groupes traités et témoins sont analysés au moyen d'une enquête, le processus doit inclure le consentement à l'utilisation des données des personnes questionnées à des fins de recherche.

⁵⁵ [Isméri Europa – Ecorys – Institute for Employment Studies, 2019.](#)

⁵⁶ Un ensemble de données supplémentaires contenant les dates de début et de fin de chaque déclaration de disponibilité immédiate au travail signée par l'ensemble des personnes traitées et non traitées a été fourni.

(Mercurio) qui est gérée par Veneto Lavoro et qui est mis à disposition sous forme de fichier à usage public à des fins de recherche⁵⁷.

Mercurio est une base de données statistiques contenant toutes les informations enregistrées par les bureaux de placement de la région de Vénétie concernant les périodes d'emploi et de chômage dans la population. Elle contient également des informations détaillées concernant l'ensemble des travailleurs et entreprises enregistrés. À la date de l'analyse, Mercurio contenait des informations sur plus de 4 millions de travailleurs, 17,6 millions d'événements d'emploi et 4 millions d'événements de chômage⁵⁸.

L'ensemble de données est périodiquement «mis à jour et nettoyé», ce qui a rendu relativement simple sa mise en correspondance avec les données de suivi de l'AG. Une limite potentielle réside dans le fait que le fichier à usage public n'est pas assez souvent mis à jour (par exemple, la dernière version actualisée actuellement disponible date de décembre 2020), étant donné que cela demande des efforts considérables en vue de garantir la bonne qualité des données⁵⁹. L'expérience montre que l'AG devrait prendre des dispositions à l'avance, tout en préparant des plans d'évaluation pour: déterminer les formes de coopération potentielles avec des organismes externes gérant les données nécessaires à un exercice contrefactuel; trouver des solutions aux obstacles juridiques potentiels; préciser les moyens d'accéder aux données administratives, de les gérer et de les fournir sous une forme anonymisée.

2.3.2. Comment le groupe «traité» est-il défini?

Pour mener à bien une EIC, il est essentiel de définir clairement ce que l'on entend par «traitement» et par «participation à l'intervention». En outre, après avoir compris clairement les circonstances dans lesquelles une personne ou une entreprise est considérée comme ayant été traitée, il est important de pouvoir les identifier. Les principales questions relatives à la définition des groupes traités et témoins sont présentées dans cette section. Le chapitre 3 fournit des explications plus détaillées sur leurs caractéristiques méthodologiques.

La définition de la participation peut sembler simple à première vue. Cependant, il existe un certain nombre d'éléments qui peuvent ne pas être immédiatement évidents, mais qui sont essentiels et qui nécessitent une réflexion approfondie. Par exemple, les stagiaires d'un programme de formation qui quittent l'intervention sont-ils considérés comme ayant été traités? À combien de séances d'un programme de formation les stagiaires doivent-ils participer pour être considérés comme des participants? Il faut également tenir compte des effets anticipés. Dans l'attente d'être soumis à une intervention, certains demandeurs de prestations sociales peuvent quitter les listes de bénéficiaires afin d'éviter des mesures d'activation. Ces personnes sont-elles traitées même si, par exemple, elles ne se sont jamais présentées physiquement à des rendez-vous organisés à leur attention par un bureau de SPE?

Définition du groupe de traitement

Il convient également de distinguer l'«intention de traiter» et le «traitement sur les traités» lors de la définition du «groupe traité». Du point de vue stratégique, toute la question est généralement de savoir si l'intérêt réside dans les effets liés au fait de se voir offrir la possibilité de participer à une

Intention de traiter ou traitement sur les traités

⁵⁷ Une description de Mercurio et des règles régissant l'accès à cette base de données est disponible à l'adresse suivante: <https://www.venetolavoro.it/public-use-file>.

⁵⁸ Les variables suivantes présentant un intérêt ont été extraites des données administratives: les dates de début et de fin de chaque période d'emploi et de chômage; le type de contrat pour chaque période d'emploi (permanent, temporaire, stage); la durée prévue des contrats temporaires; les informations sociodémographiques (genre, date de naissance, études, nationalité).

⁵⁹ Cela signifie que lorsqu'il est nécessaire de travailler sur des informations plus récentes que celles contenues dans la dernière version de Mercurio disponible, un travail supplémentaire spécifique est requis pour intégrer et nettoyer les informations les plus récentes relatives aux périodes d'emploi.

intervention, ou dans les effets d'une participation effective. Dans le premier cas, les personnes auxquelles on offre une intervention peuvent y participer ou non. Dans le second cas, où l'intérêt réside dans l'effet du traitement sur les traités, le groupe traité est uniquement composé de personnes qui participent⁶⁰.

À première vue, les responsables politiques estiment souvent qu'il est intéressant de déterminer les effets finaux du traitement sur ceux qui y participent. Cependant, après mûre réflexion, ces aspects peuvent s'avérer moins clairs. S'il est possible de recenser les personnes qui se voient offrir un traitement, il peut être plus utile, d'un point de vue stratégique, de les définir comme le groupe «traité». C'est le cas en particulier lorsque la participation à une intervention n'est pas obligatoire. Les responsables politiques ne peuvent forcer ceux qui se voient proposer une intervention à participer, de sorte que la question pertinente à poser est la suivante: quel est l'impact sur les futurs emplois et salaires de ceux qui se sont vu offrir la possibilité de participer à un programme de formation?

Offre ou traitement effectif

Pour estimer les effets de l'offre d'un traitement sur une série de résultats, les destinataires de l'offre doivent être identifiables, ce qui peut s'avérer difficile dans de nombreux cas.

Où trouver des données appropriées?

Après avoir défini les personnes traitées et ce qui constitue un traitement, il convient d'examiner la manière dont les personnes traitées seront identifiées aux fins de l'évaluation. Cela signifie invariablement qu'il faut trouver une source de données à partir de laquelle les unités traitées, qu'il s'agisse de personnes ou d'entreprises, peuvent être entièrement dénombrées ou échantillonnées. Ces registres sont généralement ceux issus des systèmes de suivi du FSE et, le cas échéant, d'autres enregistrements de données établis pour l'intervention concernée.

Trouver des sources de données pour les personnes traitées

En raison des obligations de suivi et de notification au titre du FSE (et du FSE+), les organisations bénéficiaires doivent consigner le nombre de personnes ou entreprises qui bénéficient de services dans le cadre d'une intervention, ainsi que certaines de leurs caractéristiques personnelles. Aux fins d'une EIC, les interventions devront aller plus loin et fournir des microdonnées sur leurs participants. Les évaluateurs auront besoin d'un enregistrement pour chaque unité traitée (entreprise ou personne) contenant des données sur leurs principales caractéristiques (genre, âge, niveau d'éducation, etc.). Ces données peuvent être anonymisées/pseudonymisées à des fins de confidentialité, mais lorsqu'une enquête est requise pour réaliser une EIC, il sera nécessaire d'établir si les conditions imposées par le RGPD autorisent la transmission de l'identité de ces unités (noms, adresses, numéros de téléphone, etc.) afin qu'elles puissent être échantillonnées. Des identifiants uniques pour chaque unité sont également nécessaires pour faciliter la mise en relation des enregistrements entre toutes les sources de données.

⁶⁰ Lorsque la participation à une intervention est obligatoire, il n'y a fondamentalement aucune différence entre ces deux statuts; toutes les personnes auxquelles le traitement est offert doivent y participer. Toutefois, dans la plupart des cas, les interventions ne sont pas obligatoires (comme cela a été supposé tout au long du présent guide).

2.3.3. Facteurs à prendre en considération lors de l'identification d'un groupe témoin

Afin d'obtenir une estimation du scénario contrefactuel, il conviendra généralement de définir un groupe témoin. Le choix d'un groupe témoin sera généralement limité par le fait que l'intervention est obligatoire ou non pour les participants et qu'elle est mise en œuvre dans une juridiction universelle, ou limitée à une région particulière ou à une certaine période. Le choix d'un groupe témoin approprié revêt trois aspects: 1) analytique; 2) stratégique; et 3) pratique.

Trois aspects à prendre en considération lors de la définition d'un groupe témoin

Définir un groupe témoin d'un point de vue analytique

L'objectif d'une EIC est d'obtenir des estimations non biaisées de l'incidence d'une intervention sur une série de résultats. Pour atteindre cet objectif, des estimations de résultats contrefactuels sont nécessaires. Les estimations de résultats contrefactuels sont obtenues à partir d'un groupe témoin (voir section 1.1). Comme le montrent les figures 6 et 7, un impact est estimé en déduisant une estimation du résultat contrefactuel d'un résultat observé pour le groupe «traité». La mesure dans laquelle un impact est biaisé dépend de la mesure dans laquelle le résultat contrefactuel calculé pour le groupe témoin représente le résultat qui serait apparu pour le groupe traité s'il n'avait pas été traité, toutes choses restant égales par ailleurs.

En l'absence de randomisation, le groupe témoin à trouver doit être équivalent en regard de tous les aspects importants, aussi bien observables que non observables. Cela vaut pour toutes les approches quasi expérimentales et, par conséquent, il s'agit d'une condition nécessaire pour toutes les EIC qui n'adoptent pas une approche aléatoire.

Étant donné que presque toutes les interventions du FSE sont a) volontaires (le groupe cible n'est pas obligé de participer à une intervention), et/ou b) limitées d'une autre manière (il s'agit d'interventions ou d'instruments pilotes limités à une région ou à une juridiction), les évaluateurs seront confrontés à une réserve d'unités qui pourraient être choisies pour servir de groupe témoin. Il sera nécessaire de passer au crible cette réserve potentielle afin d'affiner le choix final des groupes témoins, de sorte qu'ils soient bien appariés aux participants (le groupe traité). Dans de nombreux cas, quatre options sont potentiellement disponibles:⁶¹

Options pour la sélection des groupes témoins

- **le lieu:** des groupes témoins semblables à ceux qui participent à une intervention, mais qui sont situés dans des régions de l'État membre où l'intervention n'est pas disponible (le cas échéant). La double différence est souvent l'approche privilégiée lorsque de tels groupes témoins et les données adéquates sont disponibles. Les populations de lieux différents peuvent être très semblables et ces groupes n'auront pas eu la possibilité de participer à l'intervention ou de refuser l'invitation à participer, et par conséquent, cette source importante de biais éventuel sera absente. Cependant, les populations de lieux différents seront soumises à différentes conditions du marché du travail. Les techniques de double différence pour de telles variations fonctionnent assez bien, étant donné que les différences dans les conditions locales du marché du travail ont tendance à rester raisonnablement stables au fil du temps. Cependant, il est moins judicieux de prélever des échantillons témoins de différents

⁶¹ La présente section s'appuie sur [Card, D., Ibararan, P. et Villa, J.M., 2011.](#)

marchés locaux du travail dans les cas où l'appariement est utilisé pour estimer les impacts. Il a été démontré que les biais associés à la sélection d'échantillons témoins de différents marchés du travail peuvent être plus importants que les biais de sélection⁶²;

- **le moment:** des groupes témoins qui sont semblables aux participants, mais qui sont observés à différents moments, soit avant ou après l'intervention. Les groupes témoins sélectionnés de cette manière sont souvent nécessaires lorsqu'une intervention est universelle et obligatoire — en d'autres termes, lorsque tous les membres du groupe cible sont obligés de participer et que le programme est mis en œuvre dans une juridiction entière. Les groupes témoins formés de cette manière ont un désavantage considérable, à savoir que leurs résultats seront mesurés à des moments différents de ceux du groupe traité et qu'ils seront donc sensibles aux fluctuations cycliques, aux changements de composition et à l'évolution des tendances macroéconomiques qui peuvent entraver la reconnaissance d'un résultat contrefactuel non biaisé. Ces groupes témoins ne doivent être envisagés que lorsque la variation des résultats est limitée au fil du temps et lorsqu'un groupe témoin contemporain n'est pas disponible;
- **l'admissibilité:** dans ce cas, les groupes témoins sont choisis parmi des groupes situés au même endroit et au même moment, mais parmi des candidats qui, pour une raison ou une autre, n'étaient pas admis à participer. Ces groupes témoins sont souvent recherchés lorsqu'une intervention est universelle, que les taux de participation sont élevés, ou que la participation est obligatoire, et lorsque des règles d'admissibilité sont clairement établies, afin, par exemple, que ceux qui sont «juste inadmissibles» constituent une source éventuelle de groupes témoins. L'objectif est de trouver des groupes qui sont semblables à ceux traités mais qui, pour des raisons bien connues et fixes, (qui peuvent être quantifiées dans les données), n'étaient pas admissibles au traitement. L'accès aux interventions au titre du FSE+ est souvent fondé sur des règles d'admissibilité distinctes qui peuvent être facilement mesurées et ne sont pas sujettes à manipulation (âge spécifique des participants, durée du statut de chômeur, etc.); par conséquent, la sélection des groupes témoins peut utiliser ces seuils comme facteurs discriminatoires autour desquels les groupes traités et témoins sont répartis.
- **le choix/la sensibilisation:** en substance, les groupes traités et témoins (plutôt que le groupe traité seul) sont soumis à un processus de sélection reposant sur des choix motivés par des facteurs potentiellement non observés⁶³. Les groupes témoins peuvent être choisis parmi ceux qui étaient admissibles, mais qui n'ont pas participé. L'avantage de ce type de sélection est que les groupes témoins sont généralement issus du même marché du travail que ceux qui ont été traités. Il convient donc d'examiner ces groupes témoins avec soin, en particulier lorsqu'une approche d'appariement est utilisée pour l'EIC et lorsque des données abondantes peuvent être mises à profit pour éclairer la décision de sélection. Dans d'autres circonstances, notamment lorsque la double différence est

⁶² Voir [Heckman, J.J., Ichimura, H., Smith, J. et Todd, P., 1998.](#)

⁶³ C'est ce que [Card, D., Ibarra, P. et Villa, J.M., 2011](#) appellent les biais de «sélection à deux côtés».

appliquée, la sélection de groupes témoins de choix/de sensibilisation sera moins intéressante.

Un autre point mérite d'être souligné. Lorsque des mesures de résultats préalables au traitement sont disponibles pour les groupes traités et témoins, il est important d'analyser les tendances avant l'intervention dans les mesures de résultats pour ces deux groupes. La vérification de l'hypothèse de la «tendance commune» aborde le problème des diminutions temporaires avant l'intervention dans les niveaux d'emploi et de salaires qui auront touché certaines personnes ou entreprises admissibles aux PAMT (sinon, ils ne seraient pas admissibles à l'aide – le problème appelé «fléchissement d'Ashenfelter»). L'évaluateur recherche des tendances temporelles similaires dans les mesures de résultats pour les groupes traités et témoins afin que la reprise après une perte d'emploi à court terme ou de salaire ne soit pas confondue avec les gains relatifs à long terme que l'EIC tente de détecter.

Analyse des tendances préalables à l'intervention

La sélection de groupes témoins appropriés est un exercice complexe sur le plan technique et méthodologique. Lors de l'élaboration des systèmes d'évaluation, il est recommandé que les responsables se familiarisent avec les principaux concepts et prennent rapidement des mesures pour recenser des groupes témoins potentiels. Il importe que les personnes qui commandent une évaluation engagent des experts au début du processus de conception afin d'obtenir un soutien et des conseils.

Quelles sont les considérations stratégiques pertinentes?

La sélection d'un groupe témoin approprié n'est pas seulement un processus technique ou analytique. Bien que les aspects analytiques de l'établissement de groupes témoins appropriés soient essentiels, il est également important qu'un groupe témoin représente une alternative pertinente à l'intervention lorsque celle-ci est envisagée du point de vue de l'élaboration des politiques.

Définir une alternative à une intervention

Les EIC peuvent prendre plusieurs formes: par exemple, elles peuvent comparer les résultats d'un groupe traité ou de plusieurs groupes traités à un groupe témoin ne recevant aucun traitement; ou elles peuvent comparer un traitement à un autre sans avoir recours à un groupe témoin non traité. Le choix du groupe témoin dépendra du type de comparaison qui est le plus pertinent au regard de la stratégie et de l'existence même d'une possibilité de trouver un groupe témoin «non traité». L'encadré 10 ci-dessous fournit un exemple de comparaison entre deux traitements sans groupe témoin «non traité», l'objectif étant d'évaluer s'il y a lieu de poursuivre une intervention plutôt que l'autre. Il convient également de noter que la comparaison d'un programme avec un autre, sans l'avantage de disposer d'un groupe témoin non traité, peut donner lieu à une ambiguïté (ce point est examiné plus en détail à l'encadré 11).

Comparer le traitement à l'absence de traitement ou à un autre traitement

Encadré 10 Questions stratégiques liées à un programme de formation

Prenons un exemple où le responsable politique a l'intention d'introduire une nouvelle intervention sous forme de formation qui doit être financée par le FSE – appelons-la «intervention A». Supposons en outre que l'EM dispose déjà d'un programme de formation axé sur les mêmes personnes, mais financé par des fonds nationaux (intervention B). Dans un tel cas, une question stratégique pourrait être la suivante: les niveaux de l'emploi et des salaires des participants à l'intervention A sont-ils plus élevés que ceux des participants à l'intervention B après la participation? Et par extension, l'intervention A optimise-t-elle mieux les ressources? Si les salaires sont plus élevés pour les participants à l'intervention A, alors la réponse stratégique logique consiste à interrompre l'intervention B en privilégiant l'intervention A, si sa prestation

s'avère également efficace sur le plan des coûts.

Encadré 11 Interprétation des effets finaux

Il peut arriver qu'une étude ne constate aucune différence entre les salaires des participants à l'intervention A et ceux des participants à l'intervention B. Il se peut que la réponse stratégique à cette information ne soit pas claire, si, par exemple, l'intervention B a été très efficace par rapport à la situation sans traitement. Cela signifierait que les deux interventions sont très efficaces. Toutefois, dans certains cas, il se peut qu'aucune information ne soit disponible sur l'efficacité de l'intervention B par rapport à la situation sans traitement. Dans d'autres cas, les interventions A et B peuvent être toutes deux inefficaces, même si une intervention peut sembler relativement plus efficace que l'autre. Dans les cas où certains groupes de la population peuvent être ciblés par plus d'une intervention, il pourrait être encore plus utile de tenter de trouver un groupe approprié d'unités non traitées qui servira de point de comparaison.

Il convient de noter que la double différence ne peut pas être utilisée pour comparer des traitements multiples en l'absence d'un groupe témoin non traité.

Quelles sont les considérations pratiques nécessaires pour sélectionner un groupe témoin?

Parallèlement aux considérations analytiques et stratégiques, il est nécessaire de prendre en considération les aspects pratiques de la sélection des groupes témoins. La sélection ou l'échantillonnage d'unités (personnes ou entreprises) pour servir de groupe témoin requiert la définition d'un cadre d'échantillonnage adapté. En outre, les cadres d'échantillonnage doivent contenir des unités individuelles qui sont conformes aux exigences analytiques et stratégiques. La manière idéale de s'y prendre variera d'une évaluation à l'autre en fonction du contexte spécifique de l'intervention évaluée.

Dans de nombreux cas, deux sources de données sont souvent utilisées pour déceler des groupes témoins appropriés. Toutes deux exigent que l'identité du groupe traité soit connue.

Divers types de registres de la population peuvent être utilisés pour constituer des groupes témoins. Par exemple, si une intervention active en faveur du marché du travail s'adresse aux personnes âgées de 18 à 24 ans percevant des allocations de chômage, leurs dossiers d'allocation peuvent être utilisés pour définir la population cible. En outre, si le groupe traité est connu et peut être relié aux données relatives aux allocations, il est possible de repérer les 18-24 ans qui ne sont pas traités et qui constituent donc un groupe témoin potentiel. Autre exemple: supposons qu'une intervention soit destinée aux petites et moyennes entreprises. Les registres nationaux du commerce (s'ils sont disponibles) peuvent être utilisés pour définir la population cible, et grâce aux informations disponibles sur les entreprises qui sont traitées, des groupes témoins potentiels peuvent être définis.

Registres de la population et registres du commerce

Les registres de candidatures peuvent être utilisés lorsque l'application de l'intervention n'est pas universelle; par exemple, lorsque ceux qui ont demandé à bénéficier du programme de formation ne sont pas tous acceptés (un groupe témoin de choix/de sensibilisation). De même, les entreprises qui demandent un financement ne connaîtront pas toutes le succès et celles qui ne seront pas acceptées pour une formation ou un financement peuvent, dans certains cas, être utilisées comme groupe témoin (veuillez vous référer à la

Registres de demandes

discussion précédente de cette section concernant la prudence dont il faut faire preuve lors de la sélection de groupes témoins dans ces circonstances).

2.3.4. Quels genres de considérations en matière de données interviennent dans le système d'évaluation?

Quels types de données sont nécessaires et comment les collecter?

Comme il a été observé, les EIC nécessitent généralement d'avoir accès à des quantités considérables de microdonnées (dans certains cas, des données groupées peuvent être utilisées, comme des données régionales). Ces données doivent être collectées, rassemblées et enregistrées; les données provenant de différentes sources doivent être rapprochées en fonction des domaines d'identification communs; elles doivent être stockées et transférées en toute sécurité entre les responsables de la gestion et de l'exécution des EIC; enfin, des ensembles de données analytiques doivent être formés à partir de ces sources de données afin de faciliter l'estimation des impacts.

Gestion des sources de données

Pour concevoir un système d'évaluation, il est important de prendre en considération les questions relatives aux données ci-après:

Une liste de contrôle de base pour la gestion des données

Disponibilité

- Quelles sources peuvent servir à obtenir différents types de données groupées/microdonnées?
- Les données individuelles nécessaires sont-elles disponibles? Cela s'applique-t-il également aux «catégories particulières» de données (si nécessaire)?
- Lorsqu'une enquête est requise, les groupes cibles et témoins sont-ils définis de manière à pouvoir les suivre à travers des entretiens d'enquête – les coordonnées sont-elles disponibles et à jour?

Cohérence

- Existe-t-il une source de données unique ou faut-il établir un lien entre les sources de données (par exemple, les statistiques sur le chômage, les prestations sociales, la sécurité sociale, les données d'entreprise/d'établissement, etc.)?
- Les sources sont-elles cohérentes les unes avec les autres? Les personnes peuvent-elles être reconnues de manière cohérente à partir des différentes sources?

Accessibilité

- Est-il possible d'accéder à des sources de données nationales sur des carrières, des salaires ou des prestations sociales individuels pour comparer des participants au FSE à un groupe témoin potentiel?

Accords

- Des accords spécifiques relatifs à la disponibilité des données entre l'AG et d'autres propriétaires de données sont-ils en vigueur? Quels obstacles juridiques ou organisationnels doivent être examinés?
- Qui se chargera de négocier l'accès aux données et d'obtenir l'autorisation de les utiliser?

Conservation

- Où les données seront-elles conservées? Quels systèmes et infrastructures informatiques seront nécessaires?
- Quelles démarches seront entreprises pour garantir que les données sont conservées en toute sécurité et que leur accès est réservé à ceux qui en ont besoin aux fins de l'évaluation?
- Comment les données sont-elles anonymisées? Est-il possible de suivre des personnes dans le temps et en reliant les sources de données?

Respect du droit à la protection des données

- Les solutions précédentes sont-elles conformes aux règles fondamentales du RGPD et du règlement relatif au FSE+? Les obstacles éventuels à l'accessibilité des données sont-ils justifiés par le recours à différentes bases juridiques conformément à l'article 6, points a), c) et e), du RGPD et à l'article 17 du règlement relatif au FSE+ et par l'utilisation éventuelle de données anonymisées?

Comment les données seront-elles traitées?

Dans de nombreux cas, les EIC nécessiteront des microdonnées — c'est-à-dire, des données qui contiennent des observations sur différentes unités (généralement, différentes personnes ou entreprises) dans les groupes traités comme dans les groupes témoins (occasionnellement, des données groupées peuvent être utilisées, par exemple des données régionales ou provenant des bureaux de SPE). Nous avons distingué trois grands types de données nécessaires: a) les données relatives aux groupes traités et témoins, b) les données relatives aux résultats, et les données contextuelles (les données utilisées pour contrôler les importantes différences potentielles entre les groupes traités et témoins). Ces données peuvent provenir de sources identiques ou distinctes. Les sources doivent être structurées de manière à former des ensembles analytiques (ou des échantillons analytiques) qui servent à estimer les impacts. Dans de nombreux cas, cet exercice de structure nécessitera d'établir des liens entre les données relatives aux différentes personnes ou entreprises à partir des différentes sources. Ce rapprochement nécessite soit des identifiants personnels (par exemple, les numéros personnels d'identification à la sécurité sociale) qui permettent que les données sur une personne (issues par exemple des registres fiscaux) soient alignées avec les données sur la participation, soit des données en suffisance pour relier les données issues de différentes sources (par exemple, le nom et le lieu de naissance doivent être disponibles parmi les sources). Il est important d'examiner quelles sources seront exploitées dans le cadre de l'EIC planifiée, mais également la possibilité de relier les données issues de différentes sources.

Relier les microdonnées issues de différentes sources

2.3.5. Quelles sont les principales contraintes de l'analyse des données et des résultats?

Comme il a été dit, les impacts relevés dans les EIC sont généralement déterminés en comparant les résultats du groupe traité avec les résultats du groupe témoin. Cette comparaison fondamentale fait partie, en substance, de l'approche EIC. La différence entre les groupes traités et témoins est appelée l'impact ou l'effet net de l'intervention. Toutefois, la manière exacte dont les impacts sont estimés dépendra de la méthode de recherche adoptée.

Au moment de planifier une EIC, il importe d'examiner si l'intervention a une portée suffisante et pourrait produire des impacts pouvant être repérés

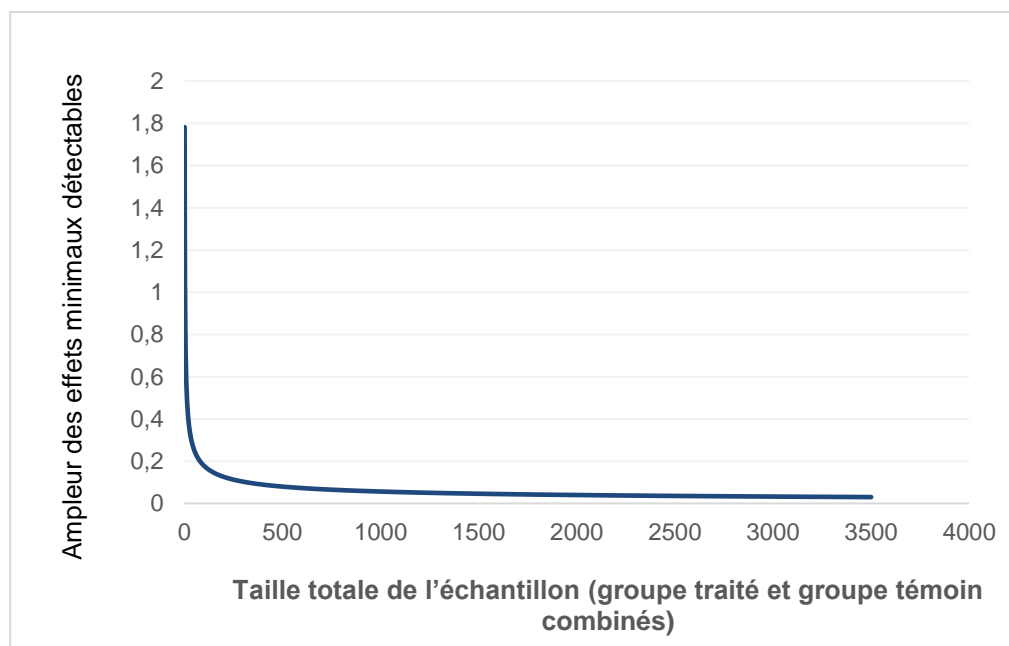
statistiquement. En réalité, les méthodes d'EIC reposent sur le principe selon lequel les groupes cibles et les groupes témoins constituent un échantillon de l'ensemble des populations potentiellement traitées et témoins et, pour produire des résultats significatifs, le nombre de personnes appartenant à ces deux groupes doit être statistiquement pertinent.

Lorsqu'il est question de déterminer si la taille d'un échantillon est suffisante pour l'analyse; un principe utile est celui de l'«effet minimal détectable»⁶⁴. Autrement dit, un effet minimal détectable est le plus petit impact réel que la taille d'un échantillon permet de détecter à des niveaux standard de confiance et de puissance statistique⁶⁵.

Évaluation de la taille de l'échantillon et de l'ampleur de l'effet

La figure 4 ci-dessous montre comment l'ampleur de l'effet minimal détectable varie selon la taille totale de l'échantillon (nombre total d'échantillons dans les groupes traités et témoins) à un niveau de confiance statistique de 95 % et à un niveau de puissance statistique de 80 %. De gauche à droite, l'ampleur de l'effet minimal détectable se réduit rapidement au fur et à mesure que la taille de l'échantillon approche de 500 (250 unités de traitement et 250 unités de contrôle). Autrement dit, si la taille totale de l'échantillon augmente, la conception de l'EIC est plus précise et permet de détecter des impacts plus petits.

Figure 4 Ampleur des effets minimaux détectables pour des tailles d'échantillons différentes



Remarque: 95 % de pertinence statistique et 80 % de puissance statistique. On considère qu'une approche aléatoire est adoptée dans cette figure.

⁶⁴ Bloom H. S., 1995 propose des orientations pratiques sur la manière de calculer les effets minimaux détectables des approches expérimentales. Pour les approches quasi expérimentales, ce calcul nécessitera un ajustement. En général, les approches quasi expérimentales demandent des échantillons de plus grande taille en comparaison avec les échantillons nécessaires pour une approche expérimentale.

⁶⁵ Le niveau de confiance statistique est une mesure statistique de la fiabilité de la procédure d'estimation, tandis que le niveau de puissance statistique désigne la probabilité qu'un test détecte un effet réel. Ces deux paramètres dépendent également de la taille de l'échantillon et, inversement, ils peuvent être utilisés pour calculer la taille d'un échantillon eu égard aux niveaux adéquats de pertinence. En général, le niveau de confiance statistique est de 95 % et le niveau de puissance statistique est de 80 %.

Lors de la planification d'une EIC, il est utile d'estimer la taille probable des échantillons afin de définir le groupe traité et le groupe témoin; cette estimation se base sur les prévisions concernant le nombre d'unités qui seront traitées, l'approche de la méthode probablement utilisée pour l'EIC et la taille des groupes témoins correspondants à partir des données administratives disponibles ou d'une étude appropriée. Ces informations peuvent ensuite, sous certaines conditions, être utilisées afin de vérifier si les effets minimaux détectables sont suffisamment significatifs et si la mise en œuvre de l'EIC constitue un choix raisonnable. Un des aspects essentiels de cet exercice consiste à déterminer si l'intervention concernée peut produire des effets d'une taille équivalente aux effets minimaux détectables estimés; à cet égard, il peut être utile d'examiner la littérature existante ainsi que les autres évaluations similaires.

Taille probable de l'échantillon

Il convient également de relever que les EIC ont bien souvent pour but d'examiner les effets sur les différentes sous-populations couvertes par l'intervention (homme/femme, jeune/âgé, moins éduqué/plus éduqué). Ces analyses entraînent nécessairement une réduction de la taille du groupe traité et du groupe témoin pour chaque sous-population (homme, femme, jeune, etc.) et donc une baisse de la pertinence statistique. Lorsque l'étude des effets sur les sous-populations constitue un composant important de l'évaluation, il est nécessaire de disposer d'un nombre suffisant de personnes dans le groupe traité et le groupe témoin appartenant à chaque sous-population.

La définition de la taille de l'échantillon nécessite des compétences techniques que l'on ne retrouve pas toujours dans l'AG. Dans ce cas, l'évaluateur doit calculer la taille de l'échantillon et convenir avec l'AG de la faisabilité et de la portée de l'EIC. Il n'est pas possible de définir une taille d'échantillon seuil au-delà de laquelle il est toujours possible de réaliser une EIC car le seuil peut varier en fonction de la méthode adoptée, des analyses attendues des sous-populations, des niveaux acceptés de pertinence statistique et d'autres éléments. Cependant, pour beaucoup d'évaluations, un nombre de 2 000 personnes au total pour le groupe traité et le groupe témoin s'est avéré suffisant pour mener une analyse raisonnablement précise (y compris les principaux sous-groupes, comme homme/femme, etc.). Cette indication ne peut être utilisée comme un «seuil» scientifique mais simplement comme une aide s'agissant d'examiner la mise en œuvre potentielle d'une EIC.

La taille de l'échantillon a une incidence sur la robustesse des résultats de l'EIC. L'encadré ci-dessous donne quelques exemples d'incertitudes concernant l'interprétation des résultats et explique en quoi elles sont liées à la taille de l'échantillon.

Encadré 12 Incertitudes concernant l'interprétation des résultats

Entre autres exemples d'évaluations des interventions financées par le FSE, l'évaluation de plusieurs mesures ciblant les chômeurs de longue durée dans la région des Marches en Italie n'a pas produit de résultats statistiquement significatifs concernant l'une des quatre mesures analysées, les «chèques formation». Cela s'explique probablement par la petite taille de l'échantillon de personnes traitées couvert par l'analyse au regard du nombre important de bénéficiaires réels de cette mesure. Un cas similaire s'est produit lors de l'évaluation d'une mesure visant à promouvoir la formation professionnelle en Pologne, dans la région de Podlachie. Outre les questions liées à la pertinence statistique, cet exemple montre que, lorsque l'on cherche à analyser les effets, le fait de s'appuyer sur de trop petits échantillons de groupes

traités peut entraîner des problèmes de généralisation des constatations.

L'évaluation de la formation linguistique dispensée aux migrants en Allemagne a montré des résultats positifs, contrairement à d'autres études antérieures visant à évaluer les effets des programmes de formation linguistique à destination des migrants. Selon l'auteur de cette évaluation, cette différence pourrait s'expliquer par le fait que le programme allemand analysé se composait également d'un module portant sur l'expérience professionnelle. L'auteur n'a pas pu tester cette hypothèse à l'aide des données disponibles en raison de l'impossibilité de distinguer les composants de formation linguistique de ceux d'expérience professionnelle de l'intervention. Cette hypothèse pourrait faire l'objet de recherches à l'avenir. Cet exemple montre que les exercices contrefactuels doivent parfois être répétés avec des données améliorées et plus riches ou complétés par d'autres approches en matière d'évaluation afin d'expliquer les mécanismes déterminant les résultats d'une intervention.

2.3.6. Liste de contrôle afin de vérifier la préparation et la faisabilité de l'EIC

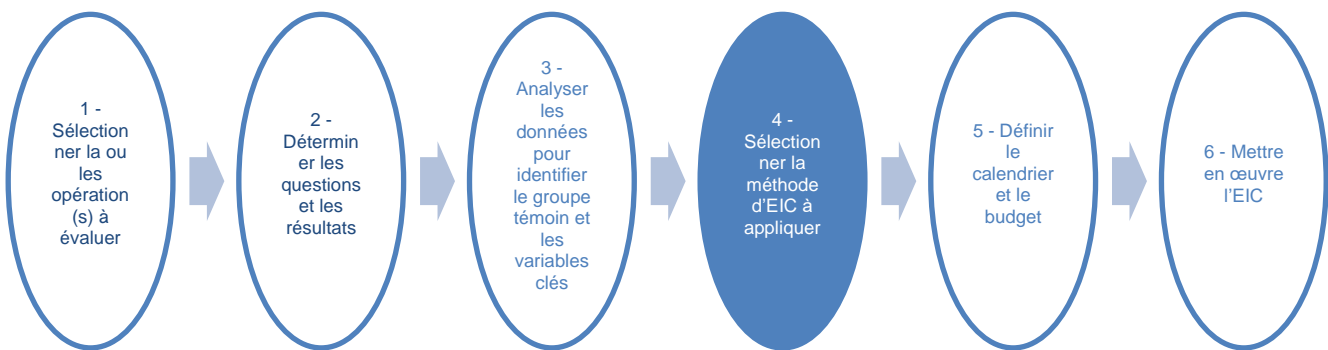
À ce stade, il est utile de résumer à l'aide d'une liste de contrôle de base les facteurs principaux qu'un AG doit prendre en considération et vérifier lorsqu'il prépare une EIC:

Liste des facteurs essentiels à prendre en considération

- L'intervention sélectionnée aux fins de l'évaluation convient-elle à une EIC?
 - o L'intervention est-elle distincte, caractéristique et suffisamment homogène?
 - o La «théorie du changement» de l'intervention laisse-t-elle supposer un mécanisme de causalité convaincant pour les résultats que l'EIC doit examiner?
 - o Est-il facile d'identifier le groupe témoin conformément aux règles de l'intervention?
- Les effets sur les participants sont-ils quantitativement mesurables?
 - o Les résultats de l'intervention à mesurer sont-ils clairs et cohérents au regard de la théorie du changement de l'intervention?
 - o Un délai suffisant s'est-il écoulé depuis la fin du traitement afin de détecter les résultats pour les participants?
- Les questions de l'évaluation nécessitent-elles une mesure de l'impact (effet net) de l'intervention?
- Des données appropriées (registres des groupes traités et témoins, registres des résultats, données de contexte) sont-elles disponibles ou peuvent-elles être mises à disposition?
 - o Les données disponibles sont-elles conformes au RGPD? En cas de réponse négative à cette question, est-il possible d'activer les accords inter-administratifs et les solutions techniques nécessaires afin de mettre ces données en conformité?
 - o Est-il possible de définir un groupe témoin approprié à l'aide des données disponibles?
 - o La taille de l'échantillon est-elle suffisante pour permettre d'atteindre la pertinence statistique requise dans les résultats de l'EIC?

Une réponse positive à toutes ces questions signifie qu'il est possible de concevoir et de mettre en œuvre une EIC. Une réponse négative à certaines de ces questions n'élimine pas la mise en œuvre d'une EIC, mais nécessite des vérifications supplémentaires concernant les conditions de faisabilité et des améliorations de la disponibilité des données ou d'autres facteurs essentiels.

2.4. Méthode d'EIC à appliquer



La sélection de la méthode d'EIC à utiliser est une étape cruciale de la préparation de l'évaluation puisque la méthode retenue a une incidence tant sur la qualité des constatations que sur la qualité générale. En règle générale, aucune méthode n'est supérieure à une autre, mais il se peut qu'une soit plus adaptée à l'exploitation des données alors qu'une autre est mieux indiquée pour enquêter sur un certain type d'intervention; la détermination de la méthode la plus adaptée à une évaluation spécifique constitue une étape essentielle de la conception.

Conditions fondamentales pour la détermination de la méthode la plus appropriée

Les aspects techniques et les caractéristiques intrinsèques de chaque méthode sont examinés en détail au **chapitre 3**. Par commodité, nous allons ici rappeler rapidement ce qu'impliquent en pratique les choix opérés.

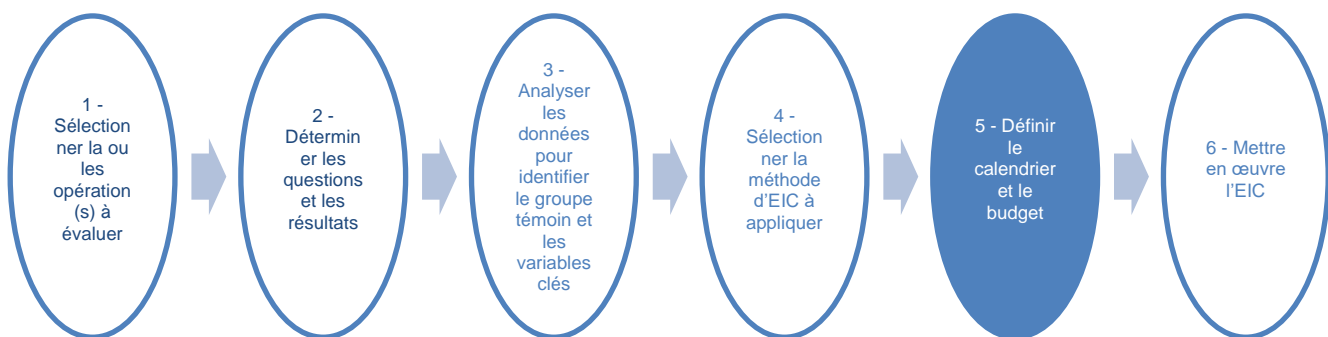
Premièrement, il convient de mettre au point une approche aléatoire (qui répartit aléatoirement les personnes dans le groupe traité ou le groupe témoin) avant le début de l'intervention. Cela nécessite d'avoir achevé la conception de l'évaluation et de disposer d'une organisation capable de mener cette tâche à bien. Cette approche a rarement été adoptée pour les interventions FSE mais, comme nous le verrons, elle est possible sous certaines conditions et s'avère très efficace d'un point de vue méthodologique.

Deuxièmement, le choix de la méthode d'EIC peut dépasser les capacités de l'AG car il nécessite des compétences techniques poussées et une expérience spécifique. Dans un tel cas, c'est l'évaluateur qui choisira la méthode la plus appropriée mais, à l'évidence, seulement après avoir été nommé. Une telle manière de procéder peut présenter l'avantage supplémentaire de constituer un critère d'appréciation de la capacité des évaluateurs lors de la procédure de sélection de l'évaluateur. En tout état de cause, l'AG doit se sentir décidée à réaliser une EIC après avoir répondu par l'affirmative à toutes (ou presque toutes) les questions de la liste de contrôle ci-dessus et après avoir acquis (en interne ou en externe) l'ensemble des compétences nécessaires pour approuver le choix de l'évaluateur.

Troisièmement, les méthodes aléatoires peuvent différer des méthodes quasi expérimentales en ce qui concerne la durée et les coûts et, pour ce qui est des essais contrôlés aléatoires, l'évaluateur devrait participer à l'intervention depuis son commencement. Les méthodes quasi expérimentales ne sont pas significativement différentes les unes des autres; l'évaluateur peut commencer à travailler une fois que l'intervention a commencé (mais suffisamment tôt pour pouvoir organiser et préparer les données nécessaires) et la qualité des données disponibles pourrait être déterminante en ce qui concerne la durée et les coûts.

Enfin, le type d'intervention et la disponibilité des données permettent parfois d'utiliser plusieurs méthodes quasi expérimentales. Dans ce cas, il peut être utile d'appliquer plusieurs méthodes, car le fait de confirmer les mêmes effets grâce à différentes méthodes améliore la robustesse des constatations.

2.5. Calendrier et budget



2.5.1. Quelles sont les ressources disponibles?

La disponibilité des ressources constitue un aspect essentiel dont il faut tenir compte lors de l'élaboration d'un programme d'évaluation d'une EIC. Cela peut comprendre un vaste ensemble de considérations, présentées sous trois titres dans le présent document: a) les ressources spécialisées; b) le temps; et c) les ressources financières.

Quels experts externes et quel personnel interne sont-ils nécessaires pour une EIC?

Dans la plupart des cas, une évaluation d'impact sera confiée à un prestataire externe. Cependant, le contrat devra être géré au sein de l'AG par le personnel qui maîtrise les méthodes d'EIC. Une telle connaissance est nécessaire afin de garantir la qualité et d'assurer une liaison efficace avec des experts externes. D'autres formes d'expertise peuvent également être requises au sein de l'AG, telles que des compétences statistiques et une expertise dans la collecte et la gestion de données. Il est important d'examiner à l'avance si l'AG dispose d'un personnel dûment qualifié et formé, et si ce personnel a les compétences nécessaires pour soutenir l'évaluation.

Personnel interne

Une EIC, pour être efficace, doit être confiée à des contractants qui détiennent les compétences et l'expérience nécessaires pour mener de telles évaluations. De plus, les contractants compétents devront

Personnel externe

comprendre le contexte politique et administratif des États membres, connaître des sources de données potentielles et maîtriser les langues appropriées. Il y a lieu d'examiner les mesures nécessaires afin de développer une base de prestataires d'EIC dans un EM (pour plus de détails sur ce sujet, voir chapitre 4).

Des EIC efficaces nécessitent la coopération des gestionnaires du programme ou de l'intervention évalué(e). Par exemple, il sera nécessaire d'accéder aux registres tenus par les gestionnaires de l'intervention. Ces registres fournissent des informations sur les personnes ou les entreprises qui ont participé à une intervention.

Programmes/interventions de gestion du personnel

Les gestionnaires de programme d'intervention peuvent fournir des conseils et des orientations sur ces types de données. Ils peuvent également être tenus de procéder à une gestion de registres dépassant ce qu'ils devraient faire en l'absence d'une évaluation d'impact.

Afin de surmonter le problème de la collecte de données provenant de diverses sources, les personnes qui planifient une EIC devront assurer la liaison avec le personnel traitant des sources de données officielles (par exemple, le registre du chômage, les données de la sécurité sociale, les services de statistiques, etc.). Cela permettra de prévoir la fourniture des données bien en avance.

Expertise statistique

Quels facteurs sont pertinents pour le calendrier d'une EIC?

Les EIC prennent du temps

La réalisation d'une EIC exige la participation d'un certain nombre de ressources humaines différentes, de plus, ces évaluations sont menées sur des périodes très longues. Un programme d'évaluation doit inclure un calendrier déterminant les principales étapes du projet, qui sera valide pour les personnes participant à l'intervention même, ainsi que pour celles liées à l'évaluation. Le projet de calendrier devra être inclus dans les activités de prestation de l'évaluation et de l'intervention, ainsi que les principales étapes stratégiques.

L'élaboration d'un projet de calendrier significatif et réaliste pour effectuer une EIC est un exercice d'équilibre difficile. D'une part, l'autorité de gestion (ou l'OI) planifiant l'évaluation doit tenir compte des dates clés auxquelles les décisions qui dépendent des résultats de l'évaluation devront être prises. D'autre part, des contraintes inévitables affecteront l'échéancier des rapports. Certains résultats prendront des années pour émerger, et les calendriers de la collecte de données, de l'analyse et de la rédaction de rapports doivent, dans la mesure du possible, en tenir compte (voir la section 2.5.2 ci-après).

Prévoir un décalage dans le temps pour permettre aux impacts d'émerger

L'évaluation doit être menée suffisamment tôt au cours de la période de programmation pour permettre d'apporter des modifications ou de tirer parti des expériences et enseignements tirés pendant le temps restant. Dans certaines circonstances, des interventions identiques ou similaires peuvent bénéficier d'un soutien au cours de périodes de programmation successives. Les résultats d'EIC se concentrant sur les interventions de périodes de programmation précédentes peuvent être extrêmement utiles pour faciliter la mise en œuvre et la conception au cours de périodes de programmation ultérieures.

Se concentrer sur des moments spécifiques

Il est également important d'examiner la manière dont le calendrier d'une évaluation contrefactuelle pourrait être lié au calendrier d'autres

Séquence de plusieurs types d'évaluation

éléments d'évaluation. Il est probable qu'une évaluation fondée sur la théorie doive être réalisée avant une EIC. Pour les interventions innovantes (par ex. les interventions au titre du FSE 2014-2020 apportant un soutien à des enfants en République tchèque ou les services encourageant l'inclusion sociale des personnes recevant le nouveau complément de revenu en Italie, ou les différentes mesures de soutien des travailleurs et des étudiants en télétravail lors du confinement dû à la COVID-19 dans de nombreux EM), il peut s'avérer nécessaire de préparer des rapports sur les éléments clés d'une évaluation de processus soit plus tôt, afin de permettre l'amélioration de la conception d'une EIC, soit plus tard, afin de faciliter la réalisation d'une enquête détaillée sur les causes et les effets. Dans le cadre d'une EIC portant sur une intervention mature en cours, il serait plus approprié de réaliser l'évaluation des processus parallèlement à l'évaluation de l'impact.

Le calendrier sera également influencé par la disponibilité des données. La mise à jour des sources de données peut prendre énormément de temps, ce qui est souvent le cas avec les registres fiscaux. Surmonter les obstacles juridiques et institutionnels pour acquérir les données nécessaires peut également s'avérer long et coûteux. De plus, les étapes consistant à analyser les données provenant de diverses sources, à garantir leur compatibilité, à vérifier leur qualité et à leur donner une forme qui peut être utilisée pour estimer les impacts nécessitent beaucoup de temps et d'efforts.

La collecte des données peut être coûteuse en temps

Comment les coûts peuvent-ils être évalués?

Il est important d'établir un budget indicatif concernant les montants que l'AG souhaite engager et est en mesure de dépenser pour réaliser l'EIC. Le budget comprend deux éléments: coûts internes et coûts externes. Le premier composant fait référence à l'effort demandé aux ressources internes pour suivre l'évaluation; au sein des administrations, ces coûts ne sont généralement pas précisés dans des termes financiers, mais ils doivent être pris en considération afin de garantir le contrôle et le suivi adéquats de l'évaluation. Les coûts liés au fait d'engager des experts externes pour réaliser une EIC doivent être soigneusement estimés afin de garantir une évaluation de haute qualité à un coût raisonnable. L'accent est ici placé sur les coûts externes.

Il convient d'opérer une distinction entre les évaluations des interventions de routine, pour lesquelles les dépenses sont généralement plus basses, et les actions innovantes ou pilotes, pour lesquelles la collecte d'une quantité relativement élevée de données, l'utilisation de nouvelles sources de données et la participation de nouvelles parties prenantes peuvent justifier des dépenses plus importantes. Cependant, il ne s'agit pas là d'une règle et cela doit être apprécié dans le cadre des questions d'évaluation, dont la complexité et le nombre peuvent nécessiter un apport financier plus ou moins important.

Méthode possible pour l'évaluation des coûts

Le choix de la méthode d'évaluation est également décisif à cet égard. Comme nous l'avons indiqué plus haut, une approche aléatoire nécessite la présence de l'évaluateur afin de collecter et contrôler les interventions pour toute la durée de l'intervention. D'autre part, le nombre de sources de données, leur qualité et leur disponibilité

déterminent en grande partie l'effort pour les méthodes quasi expérimentales.

Un document d'orientation délivré par la Commission⁶⁶ donne une indication de la somme nécessaire à l'évaluation d'un programme et précise que les programmes de grande ampleur et de routine ne devraient pas consacrer plus de 1 % de leur budget de programme à l'évaluation. Cette part peut s'élever à 10 % pour les initiatives innovantes ou pilotes. Ce document d'orientation ne traite cependant pas explicitement des besoins en ressources des EIC et nous pouvons supposer que ces pourcentages représentent le plafond des coûts d'évaluation incluant une ou plusieurs EIC.

On peut aisément comprendre qu'il est impossible de fournir une estimation des coûts d'une EIC pour un tel nombre d'interventions différentes et de conditions opérationnelles au titre des programmes FSE+ dans 27 EM. Dans ce cas, il est plus utile de proposer une méthode visant à estimer un budget raisonnable pour les EIC dans différents contextes. Le tableau suivant illustre une approche plus simple: les lignes énumèrent les principales tâches devant être réalisées, tandis que les colonnes indiquent les principaux coûts.

La durée et la complexité de chaque activité déterminent l'effort nécessaire en ce qui concerne les jours de travail de l'équipe d'évaluation et, par conséquent, ses principaux coûts. Les coûts liés à la collecte de données sous la forme d'études quantitatives auprès de participants et de membres du groupe témoins sont considérables. Lorsqu'une EIC s'appuie sur des sources de données administratives existantes et que celles-ci sont relativement simples d'accès, les coûts totaux seront plus bas. Cependant, les données administratives doivent bien souvent être soumises à une activité de «préparation» essentielle (par ex. correction des données manquantes ou erronées, ajustement du format de la base de données, vérification des règles administratives régissant les données, etc.) qui peut entraîner une hausse des coûts.

En l'absence d'autres références, les coûts liés au personnel peuvent être calculés en lien avec les coûts FSE+ pour les services de formation ou d'emploi adoptés dans chaque pays pour les formateurs ou les experts débutants ou expérimentés. Ces frais sont généralement comparables à ceux des chercheurs débutants et expérimentés et, même s'ils ne doivent pas nécessairement être équivalents, ils constituent un paramètre qu'il peut être utile de prendre en considération. Une simple analyse de marché et quelques entretiens menés par des chercheurs peuvent contribuer à préciser les coûts attendus de l'EIC.

⁶⁶ Voir [Commission européenne, 2013](#).

Tableau 3 Structure des principaux coûts d'une EIC

Activités principales	Équipe d'évaluation			Autres coûts (équipement s, matériel, fournitures, déplacements, etc.)	Remarques
	Gestionnaire du programme	Experts confirmés	Experts débutants		
Planification et coordination					Le processus de planification englobe la faisabilité de l'EIC, son organisation et la finalisation de la méthode, en accord avec l'AG. Lors de cette phase, il convient de consacrer suffisamment de temps à l'analyse des lacunes de données potentielles. La coordination couvre toute la durée de l'EIC et comprend l'organisation du travail et les interactions avec l'AG et les autres parties prenantes concernées visant à finaliser les questions d'évaluation.
Analyse bibliographique					L'analyse bibliographique vient appuyer le projet de théorie du changement, la définition des variables à utiliser et la compréhension du contexte. Elle permet également d'affiner les questions d'évaluation, d'exploiter les résultats précédents et de formuler des hypothèses.
Collecte et préparation des données					Les efforts requis pour cette activité varient grandement en fonction de la ou des méthodes de collecte des données adoptées. La collecte de données originales, au moyen d'études, peut s'avérer coûteuse en temps et en argent. D'autres méthodes peuvent nécessiter un investissement technologique (logiciel ou matériel) ou la conclusion d'accords avec les propriétaires des données. Il est possible de réduire les coûts en faisant participer les propriétaires des données à la préparation des données.
Analyse des données					L'analyse des données requiert des compétences et une expérience poussées. La méthode, la quantité d'analyse et la qualité des données ont une influence sur le temps nécessaire à cette activité.
Préparation des rapports					Les efforts requis pour cette activité varient en fonction du nombre et du type de rapports et d'autres outils de communication. Les coûts peuvent couvrir l'impression et la conception graphique en plus du temps de préparation.
Réunions de suivi					Les réunions de suivi et les autres activités de diffusion constituent une étape importante de la diffusion des constatations. Les coûts divers peuvent couvrir la location d'un lieu et la nourriture.
Total					

2.5.2. À quel moment l'intervention doit-elle être évaluée?

Il est essentiel de déterminer à quel stade d'une intervention il est le plus approprié de procéder à une évaluation d'impact, ainsi que de préciser clairement à quel moment les résultats doivent être mesurés et les impacts estimés.

Quand convient-il d'évaluer les nouvelles interventions et les interventions en cours?

La discussion relative au stade d'une intervention auquel il convient de procéder à une évaluation d'impact contrefactuelle sera influencée par la nature de l'intervention, suivant qu'il s'agisse d'une nouvelle intervention ou d'un système mature en cours. S'agissant d'une nouvelle intervention, il faut laisser davantage de temps pour qu'elle devienne mature et atteigne un état stable. La réalisation d'une EIC avant ce stade sera prématurée et fournira potentiellement des informations trompeuses. Dans le cas de nouvelles interventions, une première évaluation des processus, effectuée avant l'EIC, s'avère souvent utile pour relever les problèmes de démarrage et proposer des solutions.

Le calendrier diffère pour les nouvelles interventions et les interventions en cours

Lorsque l'on cherche à déterminer le moment optimal pour réaliser une EIC concernant une nouvelle intervention, il convient de tenir compte d'un ensemble d'autres facteurs, notamment: des mesures visant à garantir que les sources de données appropriées sont disponibles, la création d'une équipe de projet interne composée d'un personnel dûment formé et la nomination d'un contractant externe. En outre, les impératifs du processus décisionnel sur lequel porte, en fin de compte, l'évaluation constitueront une difficulté majeure.

Problèmes de ressources

Pour une intervention en cours, le calendrier d'une évaluation d'impact dépendra principalement d'exigences pratiques et stratégiques. L'intervention devrait déjà être mise en place et avoir atteint un niveau de maturité suffisant pour réaliser une EIC. Un autre aspect dont il faut tenir compte est la présence d'autres réformes entreprises parallèlement à l'intervention évaluée. Les effets de ces réformes peuvent influencer l'impact de l'intervention à l'étude. Les responsables politiques devront déterminer si la présence d'autres réformes dans le paysage politique est pertinente pour les décisions stratégiques qui s'appuieront sur les résultats de l'EIC envisagée.

Les évaluations du FSE sont généralement axées sur une seule période de programmation. Toutefois, notamment dans le cas d'interventions stables qui faisaient déjà partie du programme du FSE pour la période précédente, il peut être utile de combiner une évaluation rétrospective de la période précédente et une évaluation continue de la période actuelle afin de couvrir une plus grande durée de vie de l'intervention visée.

Périodes de programmation multiples

À quel moment faut-il mesurer les résultats et calculer les impacts?

Le deuxième aspect essentiel associé au calendrier d'une évaluation concerne le moment auquel les impacts doivent être mesurés et évalués, ou plus précisément, le moment auquel on prévoit que les effets d'une intervention émergeront.

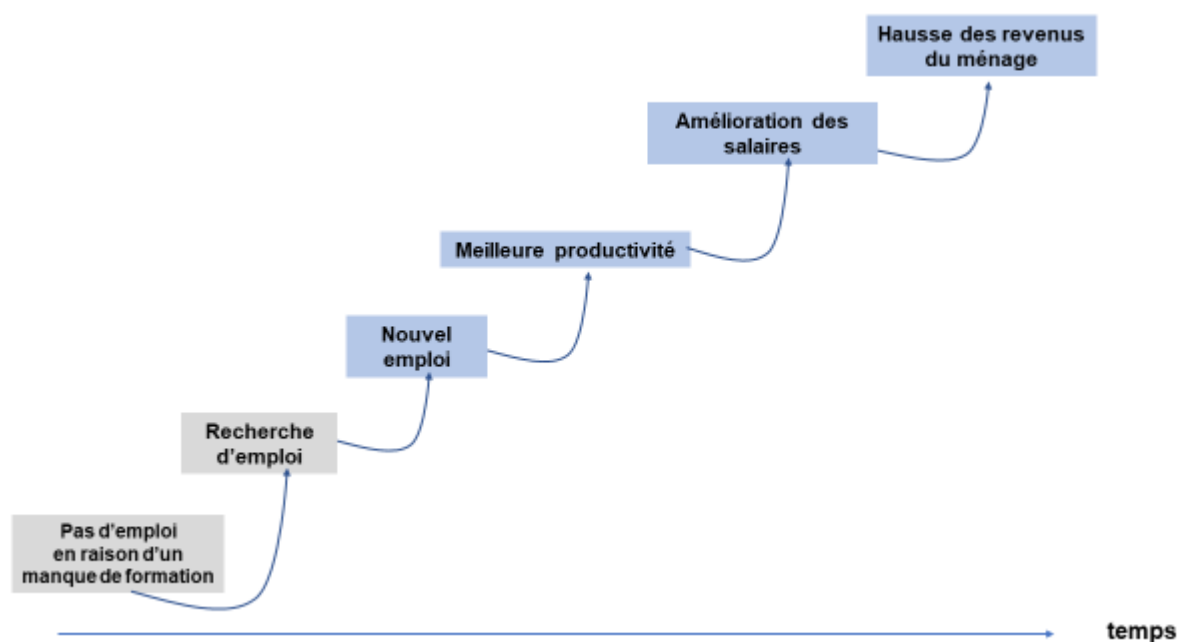
Lors de l'examen d'une intervention de formation visant les chômeurs, la question porte sur la durée à partir de laquelle on peut s'attendre à des taux d'emploi plus élevés. Le fait que les programmes de formation ont tendance, à court terme, à réduire les possibilités d'emploi chez les participants est un phénomène bien connu. Cela est dû à ce que l'on appelle l'effet de «verrouillage». Les interventions sous forme de formation ont tendance à détourner les stagiaires chômeurs de leur recherche d'emploi en raison de leur participation à des cours. À l'inverse, les personnes du groupe témoin sont décidées à trouver un emploi. Si les impacts sont calculés trop tôt, ils pourraient bien être négatifs ou sous-estimés. Lors de la planification d'une

EIC, il est important d'être réaliste quant au moment auquel les répercussions surviennent et auquel elles sont susceptibles d'être mesurables. La figure 5 ci-dessous présente un modèle simplifié d'impacts ultérieurs.

La discussion relative au moment le plus propice pour mesurer les résultats et estimer les impacts devra tenir compte des exigences des responsables politiques concernant la fourniture d'informations à certaines échéances. Dans le cas d'interventions qui visent à améliorer l'employabilité à long terme, il serait logique, d'un point de vue analytique, de suivre les participants pendant deux ans au moins après leur exposition au traitement afin de voir si leurs salaires et le niveau d'emploi sont plus élevés qu'un groupe équivalent de personnes non traitées. En revanche, les gestionnaires de programme ont souvent besoin de résultats rapidement, auquel cas il convient de trouver un compromis entre, d'une part, un intervalle de suivi raisonnable du point de vue de l'intervention et, d'autre part, le besoin des responsables politiques de disposer d'informations concrètes en temps opportun.

*Entre le
raisonnable et le
faisable*

Figure 5 Chronologie simplifiée des résultats d'un programme de formation



Si des mesures de résultats sont obtenues à partir de sources administratives (par exemple, registres d'assurance sociale incluant des détails concernant l'emploi et les revenus), il sera utile d'assurer le suivi des résultats à plusieurs reprises sur une longue période et d'estimer les impacts (même sur une base mensuelle). Ici, le risque réside dans le fait que la nature des résultats peut changer au fil du temps. Si la collecte de données primaires sous la forme d'enquêtes par sondage est nécessaire pour mesurer les résultats, l'estimation des impacts à intervalles réguliers deviendra très coûteuse, à moins que des données rétrospectives sur les résultats puissent être recueillies de manière viable. Toutefois, il ne faut pas sous-estimer les coûts liés à l'extraction des données de plusieurs systèmes administratifs et à la création d'un ensemble unique de données analytiques.

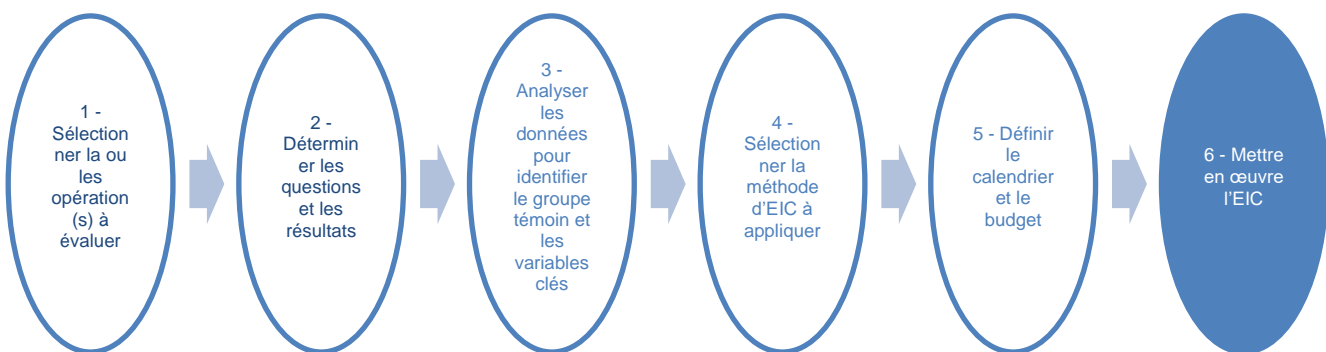
*Se concentrer sur
des moments
spécifiques*

Comme nous en avons discuté à la section 1.5, la formulation d'une théorie du changement (ou d'une logique d'intervention) peut aider à déterminer le moment auquel il convient d'estimer les impacts.

En l'absence d'une théorie du changement (mais qui serait quand même utile pour ceux qui peuvent s'appuyer sur une théorie du changement claire), une alternative pour les personnes planifiant une EIC consiste à procéder à une brève analyse des études antérieures évaluant des interventions semblables à celle envisagée. Un examen minutieux des résultats d'études antérieures peut donner une bonne indication concernant le moment adéquat pour effectuer la mesure des résultats et le calcul des impacts.

...ou sur des analyses d'études antérieures

2.6. Mise en œuvre de l'EIC



Une fois définis le calendrier et le budget, la planification de l'EIC est terminée et il est possible de passer à sa mise en œuvre. La dernière étape du présent guide vise à illustrer les activités essentielles de la réalisation de l'évaluation ainsi que la manière dont l'AG peut surmonter les obstacles qu'il pourrait rencontrer. Plus particulièrement, la présente section applique les enseignements tirés du projet Helpdesk Évaluation, par lequel la Commission et la DG EMPL apportent une assistance aux AG de tous les EM. Le document «EVALSED: La ressource pour l'évaluation du développement socio-économique»⁶⁷ aborde de manière plus générale la mise en œuvre de l'évaluation; le présent guide s'intéresse à quatre activités principales, du point de vue de l'EIC plus précisément:

1. Sélection de l'évaluateur,
2. Supervision de la mise en œuvre de l'EIC,
3. Établissement de rapports,
4. Exploitation des résultats.

2.6.1. Sélection de l'évaluateur

Lors de la sélection de l'évaluateur, la première décision à prendre consiste à déterminer si l'on fait appel à un candidat interne ou externe. Dans ce dernier cas, la sélection de l'évaluateur se fait au moyen d'une procédure de passation de marchés publics.

Indépendance de l'évaluateur

⁶⁷ Disponible à l'adresse suivante: https://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/guide_evalsed.pdf

L'article 44, paragraphe 3, du RPDC reconnaît l'existence d'un problème d'indépendance et déclare que: «*Les évaluations sont confiées à des experts internes ou externes qui sont indépendants sur le plan fonctionnel*». Dès lors, le premier critère à prendre en considération lors de la sélection de l'évaluateur est son indépendance réelle vis-à-vis de l'AG et des processus décisionnels du programme FSE+. Le deuxième critère consiste à déterminer si les compétences et ressources nécessaires à la réalisation d'une EIC sont disponibles en interne. Il s'agit probablement là de l'obstacle le plus fréquent aux évaluations internes, car les compétences techniques nécessaires à la réalisation d'une EIC sont très spécifiques et rares sont les administrations qui les ont à disposition.

Une fois opéré le choix entre un évaluateur interne et un évaluateur externe, l'administration doit énumérer les spécifications techniques de l'évaluation; dans le cas d'un évaluateur externe, les spécifications font partie de la procédure de passation de marchés publics organisée pour le sélectionner. Le cahier des charges d'une EIC doit prêter une attention particulière aux aspects suivants:

*Éléments
principaux du
cahier des charges*

- L'**objectif de l'évaluation** et les questions de recherche doivent être clairs et doivent toujours conduire à une évaluation d'impact et à une EIC lorsque cela est possible;
- Les **données disponibles** doivent être précisées; il s'agit là d'un des principaux problèmes rencontrés lors de la conception et de la mise en œuvre d'une EIC. Cela permet aux évaluateurs de prendre conscience des problèmes existants en matière de disponibilité des données et de rédiger des propositions réalistes et précises;
- **La demande de réalisation d'une EIC** peut être explicite dans le cahier des charges, mais il devrait appartenir à l'évaluateur de désigner les méthodes spécifiques à utiliser. Ainsi, il peut proposer des solutions originales et prouver ses capacités.
- L'**équipe proposée** doit combiner une expertise dans plusieurs domaines, notamment: l'évaluation et l'EIC (savoir comment concevoir et mettre en œuvre une évaluation), l'économétrie (utiliser les bonnes méthodes) et le domaine politique de l'intervention évaluée (reconstruire la théorie du changement et interpréter les constatations dans leur intégralité).
- **Les critères d'attribution** doivent donner la priorité à la qualité plutôt qu'au coût de la proposition. La qualité d'une évaluation dépend largement de la capacité de l'évaluateur à la concevoir et à l'exécuter. Bien souvent, les cadres juridiques applicables à la passation de marchés publics ne conviennent pas parfaitement aux besoins des offres d'évaluation, mais les personnes qui commandent des évaluations devraient adapter, dans la mesure du possible, les règles existantes afin de récompenser la qualité et la capacité technique.
- **Le processus de sélection** devrait inclure des experts en matière d'évaluation et d'EIC dans le comité de sélection afin de pouvoir pleinement apprécier la qualité des propositions et correctement évaluer les différentes solutions méthodologiques et organisationnelles présentées.

Certains EM disposent encore d'une expérience négligeable en matière d'EIC. Le processus de sélection peut donc être confronté à des problèmes en raison du caractère trop restreint du marché pour de tels services. Seul un

nombre très limité d'entreprises ou d'experts pourraient potentiellement être en mesure de remplir les conditions techniques du cahier des charges et de présenter une proposition de bonne qualité. Dans de tels cas, l'organisation par l'AG de séminaires ouverts avec les universitaires et les entreprises intéressés par les évaluations pourrait aider à préparer l'EIC et les appels à propositions qui en découlent. Les séminaires permettraient de collecter les suggestions des experts nationaux et de leur donner le temps d'organiser leur participation à un appel d'offres; afin de respecter la transparence et d'éviter tout conflit d'intérêts, les séminaires devraient être publics et les informations de base devraient également être mises à disposition des personnes n'ayant pas pu y assister.

2.6.2. Supervision de l'EIC

Après avoir choisi l'évaluateur et lancé l'EIC, l'AG doit superviser le processus d'évaluation. Cela consiste principalement à contrôler et valider les éléments livrables de l'évaluateur, à organiser des réunions de suivi avec l'évaluateur et les parties prenantes et, si nécessaire, à conclure des accords avec les propriétaires des données.

L'AG doit être en mesure d'accomplir ces missions. Cela signifie également qu'il doit disposer des ressources nécessaires pour superviser l'intégralité du processus et des compétences requises pour contrôler les solutions techniques de l'EIC. De manière générale, l'AG assigne des membres du personnel spécifiques à la réalisation des évaluations et, lorsqu'ils ne disposent pas d'une expérience suffisante en matière d'EIC, il convient de rechercher l'aide d'experts externes indépendants. Ces experts peuvent être engagés au moyen de contrats à court terme. Lorsqu'ils proviennent d'institutions publiques, leurs services sont gratuits. L'engagement demandé nécessite probablement de 6 à 12 jours de travail, ce qui devrait suffire pour examiner les rapports et participer à quelques réunions avec l'AG, l'évaluateur et les principales parties prenantes.

Compétences internes et externes pour le processus de supervision

La supervision des processus d'évaluation d'un programme prend souvent la forme de la création d'un groupe de pilotage. Un tel groupe se compose souvent de l'AG, de fonctionnaires représentant d'autres services chargés de la mise en œuvre, des principales parties prenantes (partenaires sociaux, ONG, autres administrations) et de certains experts provenant du monde universitaire et/ou des institutions publiques (par ex. l'institut national de statistique). Le groupe de pilotage participe à la définition des questions d'évaluation et discute des constatations de l'évaluation, de leur diffusion et de leur utilisation dans les processus décisionnels. Un sous-groupe plus flexible et plus spécialisé pourrait également aider l'AG à superviser les éléments livrables de l'évaluation.

2.6.3. Établissement de rapports

L'établissement de rapports est l'une des activités essentielles de l'évaluation et constitue le principal moyen de communication des résultats. Les rapports d'évaluation doivent être clairs, concis, intelligibles pour des non-spécialistes et transparents en ce qui concerne les opinions et les recommandations politiques qu'ils formulent. Ils doivent également prouver que l'évaluation est fiable en précisant la méthode et les données utilisées, ainsi qu'en déterminant les limites possibles de l'analyse.

En règle générale, chaque EIC fait l'objet de trois rapports:

- Le **rapport de lancement** est généralement rendu rapidement après la signature du contrat et présente la finalisation de la méthode, conformément aux discussions organisées avec l'AG, ainsi qu'une première sélection des données disponibles.
- Le **rapport intermédiaire** est rendu à une étape intermédiaire de l'évaluation et une EIC peut être réalisée pour montrer les données collectées, la composition du groupe traité et du groupe témoin et l'échantillonnage.
- Le **rapport final** présente les analyses et constatations de l'EIC. Il comprend des détails relatifs à la méthode et aux données et souligne les conséquences politiques émergeant des éléments de preuve. La structure du rapport final d'une EIC doit inclure certaines sections essentielles: les questions d'évaluation; la théorie du changement (ou la logique d'intervention) et la définition des résultats à évaluer; la méthode adoptée; les données utilisées et les caractéristiques du groupe traité et du groupe témoin; les effets estimés; les réponses aux questions d'évaluation et leurs conséquences politiques.

Rapports distincts pour le suivi de l'avancement et l'utilisation des résultats

La personne qui commande l'évaluation examine tous les rapports, avec l'aide d'experts externes si nécessaire, et les observations font l'objet de discussions avec l'évaluateur, qui traitera les clarifications et améliorations demandées avant l'acceptation formelle du rapport.

Les rapports d'EIC devraient éviter de trop entrer dans les détails techniques et formuler les constatations de manière claire et lisible pour les parties prenantes et les non-spécialistes. Des annexes spécifiques peuvent couvrir les étapes plus techniques de l'analyse, ce qui permet de rédiger le rapport principal de manière simple et fluide. Cependant, les détails méthodologiques sont essentiels pour démontrer la fiabilité de l'EIC et ils doivent également être transparents. Le rapport doit couvrir la définition des variables des résultats, la qualité des données, la composition du groupe traité et du groupe témoin et la taille de l'échantillon ainsi que la méthode d'estimation des effets et les tests statistiques utilisés. De plus, conformément aux pratiques universitaires, les données anonymisées utilisées et les méthodes de traitement des données pourraient être mises à disposition des chercheurs qui souhaiteraient répéter et vérifier les analyses.

Le rapport final de toutes les EIC devrait fournir un ensemble normalisé d'informations minimales. Un tel ensemble d'informations minimales a un double objectif: 1) faciliter la comparaison des constatations entre différents PO et différents pays; 2) collecter des informations qui seront disponibles pour des méta-analyses futures.

Ensemble d'informations minimales à fournir dans le rapport final d'une EIC

Le modèle pour la collecte d'un ensemble normalisé d'informations minimales dans chaque évaluation FSE suit les recommandations formulées dans l'étude «Pilot and feasibility study [...]». D'un point de vue opérationnel, ces informations pourraient être incluses dans une annexe du rapport final et pourraient également être envoyées à la Commission. Le

Tableau 4 ci-dessous présente un modèle possible pour la collecte de l'ensemble d'informations minimales nécessaires.

Tableau 4 Informations de base à inclure dans une fiche présentant l'EIC

Catégorie d'information	Informations spécifiques devant être collectées et incluses dans le rapport
1. Informations relatives à l'intervention	<ul style="list-style-type: none"> - Nom de l'intervention évaluée - Caractéristiques de l'intervention (formation, services d'emploi, services d'inclusion sociale, etc.) - Approche du programme FSE+ (axes de priorité, objectif spécifique, action) - Autorité de gestion et/ou autre organisme de mise en œuvre - Dépenses - Durée/intensité du traitement fourni par l'intervention - Groupe de participants par âge, sexe et statut d'admissibilité (bénéficiaire d'une assurance chômage, chômeurs de longue durée, NEET) - Portée territoriale de l'intervention
2. Mesure de l'efficacité de l'intervention	<ul style="list-style-type: none"> - Indicateur précisant si l'EIC estime que le traitement a eu un effet positif et statistiquement significatif sur les résultats pertinents - Ampleur réelle de cet effet - Les deux mesures susmentionnées (+/-) et (ampleur) pour l'ensemble de résultats pertinents et comparables: par ex. taux d'emploi, revenus, etc. - Indicateurs relatifs au rapport coût-efficacité (s'ils sont estimés)
3. Données et méthode	<ul style="list-style-type: none"> - Source des données (étude, données administratives) - Période d'observation - Méthode d'EIC économétrique - Horizon temporel (en règle générale, effets à court terme <= 12 mois à compter de la fin de l'intervention, effets à moyen terme > 12 mois et <= 24 mois, effets à long terme > 24 mois) - Taille de l'échantillon des groupes traités et témoins - Disponibilité / inclusion de données antérieures au programme

2.6.4. Exploitation des résultats

L'utilisation des résultats est importante dans la mesure où, si les résultats ne sont pas diffusés efficacement pour atteindre leur public cible, l'évaluation n'aura que peu d'impact.

La diffusion des constatations de l'EIC est essentielle

La diffusion des conclusions et des résultats de l'évaluation inclut généralement:

- Au moins un rapport d'évaluation écrit, y compris un résumé et une synthèse;
- Au moins une présentation orale des conclusions, appuyée par des diapositives ou autre outil comparable;
- Une section (ou annexe) technique du rapport comprenant un compte-rendu complet de la méthode utilisée, des principales hypothèses formulées et de l'approche adoptée pour l'analyse statistique.

Tous les rapports d'évaluation doivent être rendus publics. Il s'agit d'une disposition du règlement portant dispositions communes pour la période de programmation 2021-2027⁶⁸. De plus les constatations de l'évaluation doivent être présentées et discutées au sein du Comité de suivi du programme; ce passage est obligatoire mais ne suffit pas à lui seul à encourager un débat de fond, car les priorités opérationnelles prévalent généralement lors des réunions des Comités de suivi. Voir l'encadré ci-dessous pour un exemple de

⁶⁸ Article 44, paragraphe 7, du RPDC 2021/1060.

pratique nationale en matière de présentation et de diffusion des constatations d'une évaluation.

Il n'existe pas une seule politique de communication efficace pour la diffusion des leçons retenues et l'ouverture d'un débat à ce sujet; chaque contexte nécessite sa propre stratégie. Il est par conséquent important de prévoir une stratégie efficace et notamment de veiller à ce que les parties prenantes autres que l'AG soient informées des constatations. Il est possible de répartir le «public» de l'EIC en grands groupes et de définir les principaux types de communication; par exemple:

De nombreux publics et de nombreuses approches de communication

- Décideurs politiques et représentants politiques de haut niveau: ils sont habitués à lire des documents courts (notes d'information, notes de synthèse) présentant les principaux résultats et recommandations;
- Fonctionnaires de l'administration participant à la mise en œuvre de l'intervention ou de politiques similaires: des séminaires ou des ateliers peuvent représenter un instrument efficace pour présenter les résultats d'une EIC, en adaptant le niveau technique de la présentation au niveau d'interaction avec les participants;
- Parties prenantes publiques (partenaires sociaux, ONG, bénéficiaires, etc.): un résumé des rapports et les réunions annuelles sur les constatations de l'évaluation peuvent constituer l'instrument combiné permettant de faire participer ces parties au débat. Plus particulièrement, lorsque ces acteurs ont contribué à la définition des questions d'évaluation, la communication des résultats de l'évaluation devrait être liée à ces questions.
- Experts et universitaires: des conférences ou séminaires scientifiques peuvent être les occasions les plus appropriées de présenter les résultats de l'évaluation. Ces acteurs valident les résultats de l'évaluation d'un point de vue scientifique afin d'en renforcer la fiabilité.

Les méta-évaluations⁶⁹ représentent un instrument puissant de diffusion des résultats des EIC. Elles permettent de confirmer et de généraliser les constatations principales qu'une EIC unique ne peut pas produire par elle-même. La réalisation d'une méta-évaluation nécessite de disposer d'un certain nombre d'EIC fiables et de bonne qualité, portant sur des politiques comparables; à l'heure actuelle, les méta-évaluations restent rares et ne se concentrent pas uniquement sur les interventions au titre du FSE. Une utilisation généralisée des méta-évaluations nécessiterait un recours plus large aux EIC et la planification anticipée de celles-ci afin de faciliter leur comparaison ultérieure. Au cours de la période de programmation FSE+ 2021-2027, des méta-évaluations peuvent être mises en œuvre dans des pays accueillant un nombre considérable de programmes ou au moyen de comparaisons entre les pays.

Méta-évaluation afin de renforcer la diffusion et la généralisation

⁶⁹ C'est-à-dire une comparaison systématique, même avec l'appui d'outils statistiques, de nombreuses EIC portant sur des interventions similaires. Une méta-évaluation fournit des indications concernant l'efficacité d'un type d'intervention en examinant les effets produits par des mesures similaires dans de nombreux contextes différents. Cependant, les méta-évaluations ne peuvent pas entrer dans les détails de chaque intervention utilisée pour la comparaison et leurs conséquences politiques portent sur de larges orientations politiques et non pas sur les détails propres à chaque intervention.

Encadré 13 L'expérience polonaise concernant les Conférences d'évaluation

La conférence internationale d'évaluation, organisée par le Ministère des fonds de développement et de la politique régionale et l'Agence polonaise pour le développement des entreprises, est l'un des éléments clés du paysage de l'évaluation de l'intervention publique en Pologne. Depuis son lancement, elle est devenue une plateforme de discussion bien établie pour les parties prenantes chargées de l'élaboration des politiques, de leur mise en œuvre et de l'évaluation de leurs effets.

Lancée en 2005, la conférence s'est tenue tous les ans pendant 10 ans puis tous les deux ans depuis 2015. La 14^e édition de cet événement s'est tenue en 2021. À l'heure actuelle, la conférence est organisée sous la forme d'un événement en deux jours. Le premier jour est consacré aux constatations des recherches et aux effets des interventions; il s'agit d'un jour au cours duquel on discute des politiques. Le deuxième jour se concentre sur les problèmes méthodologiques et sur l'évaluation en tant que processus.

Lors des 14 éditions de la conférence, les méthodes contrefactuelles ont fait l'objet de tables rondes ou de présentations d'experts à plusieurs reprises et ont été abordées sous différents angles pendant les deux jours. Par exemple, en 2017, les effets estimés ont été présentés lors de sessions portant sur l'impact des interventions, tandis que des discussions sur les avantages et les inconvénients des techniques d'estimation ont été organisées séparément.

Cette conférence est également digne d'intérêt en raison de son public varié. En effet, elle s'adresse non seulement aux chercheurs ou aux spécialistes en matière de méthode, mais également aux personnes qui conçoivent et mettent en œuvre les interventions. Les objectifs de cet événement sont doubles: le renforcement des capacités du système d'évaluation en développant les compétences et les connaissances de ses membres, d'une part, et la mise en réseau, en donnant aux parties prenantes concernées un espace de partage des expériences entre les partenaires nationaux, régionaux et étrangers, d'autre part. En effet, chaque édition rassemble entre 300 et 400 représentants de différents secteurs: l'administration publique, le monde universitaire, les entreprises de conseil et les ONG. Ce nombre est considérable si l'on tient compte du fait que les approches fondées sur des données probantes en matière d'élaboration des politiques ne connaissent pas encore un enracinement fort en Pologne.

La Conférence internationale d'évaluation fait partie du système d'évaluation de l'administration publique. Il a été possible de renforcer de manière considérable les capacités grâce aux formations assurées par le Centre commun de recherche ou la Banque Mondiale ou financées au titre de l'assistance technique des programmes opérationnels. La formation à l'évaluation, y compris à l'approche contrefactuelle, a été mise à disposition non seulement des employés des services d'évaluation des autorités de gestion mais également des institutions telles que les agences locales ou régionales pour l'emploi. La conférence constitue une occasion de rassembler, d'aborder et d'échanger les évolutions et les constatations récentes⁷⁰.

⁷⁰ Des informations concernant la conférence sont disponibles à l'adresse suivante: <https://www.ewaluacja.gov.pl/strony/xiv-miedzynarodowa-konferencja-ewaluacyjna/>. Nous souhaiterions remercier Piotr Strzęboszewski et les autres fonctionnaires du Ministère des fonds de développement et de la politique régionale pour les informations qu'ils nous ont fournies concernant le système de conférences.

Chapitre 3. Comment sélectionner la méthode appropriée pour réaliser une EIC

Le présent chapitre présente des méthodes empiriques pour l'évaluation d'impact contrefactuelle. Plus précisément, nous allons nous intéresser à l'approche expérimentale (essai aléatoire contrôlé) ainsi qu'aux méthodes quasi expérimentales les plus courantes: l'appariement (par score de propension), les doubles-différences, l'approche de la régression par discontinuité et les variables instrumentales. Chaque méthode se caractérise par une manière spécifique de générer un groupe témoin afin de répondre à la question suivante: «Que serait-il arrivé au groupe traité, au regard des résultats, s'il n'avait pas participé ou été exposé à l'intervention?». En pratique, la conception retenue peut être adaptée à un contexte donné, en fonction du type d'intervention et du type de données disponibles ou pouvant être collectées⁷¹.

Il n'est pas possible de proposer des orientations détaillées quant au choix de l'approche la plus appropriée pour l'évaluation parmi les scénarios extrêmement variés que rencontrent les AG. Lors du choix de l'approche la plus appropriée pour une EIC dans un ensemble donné de circonstances, les AG doivent examiner ce qui a bien fonctionné lors des évaluations précédentes à la fois au sein même de l'AG, dans l'EM et d'autres EM; les AG peuvent tirer des leçons de ce qui a été accompli précédemment dans les programmes et ailleurs, lorsque des circonstances semblables étaient réunies. Les plates-formes où s'échangent les leçons à retenir pour la conception et la mise en œuvre des évaluations peuvent constituer de précieuses sources d'informations à cet égard. Les recherches, dans la documentation, d'évaluations d'interventions similaires peuvent également constituer une source importante d'informations pour faciliter le processus de conception. Les experts désignés par l'AG se pencheront également sur la meilleure manière d'approcher la conception d'une évaluation. Il importe de retenir que les AG peuvent proposer des compétences techniques et une expérience considérables desquelles s'inspirer.

Sélectionner la bonne approche

3.1. Randomisation – l'approche expérimentale

Caractéristiques essentielles de la randomisation

- Les personnes admissibles à la participation sont réparties aléatoirement entre le groupe traité ou le groupe témoin.
- La randomisation garantit que les deux groupes sont identiques (en moyenne) en ce qui concerne toutes les caractéristiques pertinentes.

⁷¹ Pour d'autres aspects techniques méthodologiques, voir également [Commission européenne, 2019](#).

- Par conséquent, le groupe témoin répond à la question contrefactuelle et la différence de résultats entre le groupe traité et le groupe témoin indique l'effet de causalité de l'intervention.

La conception d'évaluations aléatoires peut prendre de multiples formes. En l'espèce, l'accent est mis sur une approche simple en deux groupes (un groupe traité et un groupe témoin) afin de clarifier les principes essentiels. La figure 1 illustre une conception aléatoire simple.

Le point essentiel dans une répartition aléatoire (ou randomisation) est que les deux groupes soient statistiquement équivalents à tous les égards au moment de la répartition. Après celle-ci, le groupe traité est soumis à l'intervention faisant l'objet de l'évaluation et dont il faut mesurer les impacts ou les effets.

*Groupes
statistiquement
équivalents*

En fonction de la question stratégique au centre des préoccupations, le groupe témoin peut avoir pour consigne de ne bénéficier d'aucune intervention, le groupe bénéficiaire d'une intervention peut être comparé à un groupe bénéficiant d'une autre intervention présentant de l'intérêt (considérée comme représentant l'intervention habituelle) ou il peut exister plusieurs groupes de traitement pour un groupe témoin. Il peut par exemple être intéressant de comparer l'impact d'un programme de formation financé par le FSE avec d'autres formations financées par le budget national, ou avec la fourniture d'autres services aux mêmes groupes.

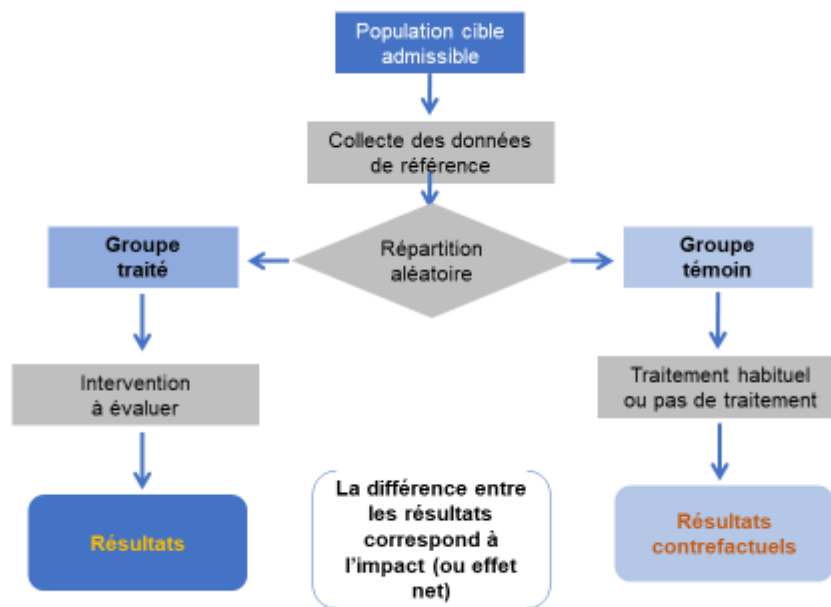
*Pas de traitement
ou autre traitement
pour le groupe
témoin*

Puisque les groupes de traitement et les groupes témoins sont statistiquement équivalents au niveau de la répartition aléatoire et qu'ils sont admis ensuite à des interventions contrôlées, les différences de résultats peuvent être attribuées à l'intervention en cours d'évaluation (sous réserve des incertitudes statistiques habituelles) et toute autre explication des différences observées peut être exclue (voir l'exemple présenté dans l'encadré ci-dessous).

Grâce aux caractéristiques intrinsèques de leur conception, et si elles sont mises en œuvre correctement, les approches aléatoires offrent la possibilité de fournir des preuves solides des effets d'une intervention. C'est pourquoi elles sont nettement privilégiées. Toutefois, elles nécessitent une planification précoce et détaillée et peuvent être difficiles à concevoir et à gérer. Par ailleurs, les gestionnaires de programmes éprouvent d'importantes difficultés à les mettre en œuvre correctement. Par exemple, le processus de répartition aléatoire même pourrait modifier le choix des participants à une intervention: c'est-à-dire que certains participants potentiels pourraient être rebutés par l'idée de la répartition aléatoire et refuser de participer. En outre, les personnes soumises à la répartition aléatoire peuvent ne pas toujours respecter leur statut d'affectation et une multitude d'autres difficultés pourraient se poser. Dans certains cas, les essais aléatoires contrôlés sont coûteux à mettre en œuvre.

*Éléments de
preuve solides...*

Figure 6 Contrôle aléatoire en deux groupes



Pour cette raison, parmi d'autres, il semble improbable que les évaluations d'instruments et interventions financés par le FSE soient réalisées à l'aide d'une approche aléatoire. Toutefois, le présent guide met en garde contre l'envie d'exclure systématiquement la répartition aléatoire au premier abord. L'approche a été largement utilisée et, hormis le cas britannique présenté à l'encadré suivant, les exemples abondent, notamment les expériences GAIN aux États-Unis⁷² (et bien d'autres encore en Amérique du Nord), les expériences menées en Suède⁷³ pour les interventions au titre du FSE 2014-2020 (voir l'encadré suivant) ainsi qu'une étude entreprise en Allemagne pour évaluer les effets des services actifs du marché de l'emploi proposés par des prestataires privés par rapport aux services proposés par le service public de l'emploi⁷⁴.

...mais difficiles à concevoir

Encadré 14 Exemple d'essai aléatoire mené sur un projet du FSE à destination des jeunes

Évaluation du projet «Ung framtid» (Avenir jeune) en Suède

Le projet suédois «Ung framtid» (Avenir jeune) est un projet financé par le FSE mis en œuvre entre 2015 et 2018. Son objectif était d'accompagner de jeunes chercheurs d'emploi âgés de 18 à 24 ans entrant sur le marché du travail dans les régions de Norrland du centre, de Suède du centre-nord et de Suède du sud. Le projet, Arbetsförmedlingen, est venu intensifier et personnaliser les activités de soutien ciblant les jeunes chercheurs d'emploi et était conduit par les Services publics de l'emploi (SPE) suédois; il consistait en une planification individuelle, des informations, des conseils et une assistance concrète concernant les activités de mise en correspondance, de formation et autres. Au total, près de 17 000 jeunes ont été inclus dans ce projet.

L'évaluation a été demandée par le Conseil suédois du FSE à Arbetsförmedlingen au titre du cadre d'évaluation mis au point dans un projet du FSE 2014-2021 cherchant à améliorer la capacité d'évaluation des SPE (Projet du fonds européen 2014-2020 fondé sur des éléments factuels; voir l'encadré dans le chapitre final du guide).

⁷² Voir [Riccio J., Friedlander, D., Freedman S., 1994.](#)

⁷³ Voir [Hagglund, P., 2006.](#)

⁷⁴ Voir [Krug G., Stephan G., 2011.](#)

Lors d'une expérience menée entre juin 2017 et janvier 2018 auprès de huit bureaux locaux pour l'emploi, sur un total de 90 ayant participé au projet, des jeunes admissibles ont été répartis de manière aléatoire entre le groupe traité (personnes recevant une aide intensifiée) et le groupe témoin (personne recevant une aide normale)⁷⁵. 4 689 jeunes ont été répartis de manière aléatoire entre le groupe traité et le groupe témoin de l'expérience et 2 972 d'entre eux ont été intégrés dans le groupe témoin. La répartition aléatoire a permis de former des groupes traités et témoins très semblables au moment de la répartition. Dès lors, toute différence entre les deux groupes en ce qui concerne les variables de résultat (sortie du chômage, proportion de chômeurs et nombre moyen de jours de chômage) mesurées après le début de l'intervention du FSE est attribuable aux activités d'Avenir jeune.

Les constatations de l'étude montrent que le projet Avenir jeune a produit des impacts positifs sur les femmes, mais pas sur les hommes. Pour les femmes, des effets positifs ont été constatés pour toutes les variables de résultat prises en considération: la sortie du chômage des femmes du groupe traité a augmenté de sept points de pourcentage au cours des deux premiers mois suivant l'entrée dans le projet, la proportion de chômeurs a connu une baisse plus importante dans le groupe traité que dans le groupe témoin et, enfin, les femmes du groupe traité ont connu moins de jours de chômage que les femmes du groupe témoin⁷⁶.

L'approche aléatoire se distingue principalement des autres approches par la grande importance qu'elle accorde au contrôle des biais potentiels entre le groupe traité et le groupe témoin à travers les protocoles de recherche. Cet accent mis sur la conception rend l'approche très intuitive mais nécessite une planification avancée. Les approches aléatoires se prêtent souvent mieux à l'évaluation de nouvelles interventions pilotes plutôt qu'aux interventions existantes. Elles demandent en effet un certain degré de contrôle sur la manière dont les participants sont recrutés pour l'intervention à évaluer. Ce «contrôle» est souvent plus difficile à exercer sur les programmes existants que sur les nouvelles interventions, ouvertes aux nouvelles idées.

Répartition aléatoire via les protocoles de recherche

La mise en place d'un protocole contrôlé aléatoire signifie qu'une partie de la population cible admissible n'est pas invitée à participer mais est répartie dans un groupe témoin. Cette répartition se fait de manière totalement aléatoire et n'est pas le fait des candidats ni des administrateurs de l'intervention. Pour cette raison, les responsables politiques peuvent avoir tendance à s'opposer aux RCT pour des raisons éthiques, avant même d'évaluer leur faisabilité d'un point de vue pratique et analytique.

Objections éthiques

Pourtant, des arguments solides plaident en faveur des approches aléatoires. Si la répartition aléatoire est gage de meilleure qualité et de plus grande fiabilité des preuves de l'efficacité des interventions financées publiquement, il est important de l'utiliser davantage pour évaluer les impacts des interventions. D'autre part, si les impacts d'une intervention donnée sont a priori inconnus, ce n'est pas contraire à l'éthique d'exclure certaines personnes, puisque rien ne dit qu'elles auraient bénéficié de l'intervention. Du reste, ces approches sont largement utilisées en médecine et dans d'autres domaines d'étude comme la recherche en matière d'éducation. Enfin, dans certains cas, lorsque les services et le soutien proposés par une intervention sont trop demandés (c'est-à-dire qu'il y a plus de personnes admissibles que l'intervention ne peut en prendre en charge), sélectionner au hasard les personnes qui bénéficieront de l'intervention

...mais aussi des arguments forts en faveur d'une approche aléatoire

⁷⁵ Une phase pilote a été menée entre avril et mai 2017 afin de tester plusieurs activités, comme la gestion de la répartition aléatoire, l'implication des jeunes participants et l'établissement de rapports.

⁷⁶ Axdorph E, Egebark J., Lundström T., Özcan G., 2019.

parmi les candidats admissibles est peut-être la manière la plus éthique de mobiliser des ressources limitées.

3.2. Approches non aléatoires ou quasi expérimentales

Si la répartition aléatoire n'est pas possible, la boîte à outils des EIC comporte un ensemble de méthodes alternatives qui utilisent différentes approches afin de générer un groupe témoin pour répondre à la question contrefactuelle. La présente sous-section définit tout d'abord le défi principal auquel sont confrontées les méthodes quasi expérimentales de manière générale puis présente quatre approches couramment utilisées dans la pratique empirique des EIC.

Lorsque la répartition aléatoire n'est pas possible

3.2.1. Groupe cible et groupe témoin sans répartition aléatoire

Dans les approches quasi expérimentales, les groupes cibles qui bénéficient de l'intervention sont comparés avec un groupe témoin⁷⁷ réunissant des cibles ou des cibles potentielles réparties de manière non aléatoire et qui ne bénéficient pas de l'intervention. À l'instar d'une expérience, l'objectif est d'obtenir une estimation impartiale du changement produit sur les résultats par l'intervention examinée. Puisque les groupes traités et témoins ne sont pas formés au hasard, les approches quasi expérimentales supposent d'accorder une plus grande attention aux méthodes qui prennent en considération les différences potentielles entre les membres du groupe de traitement et les membres potentiels du groupe témoin qui sont susceptibles d'influencer la décision de participer et donc les résultats. La solution consiste à sélectionner un groupe témoin plausible. Ne pas réussir à sélectionner un groupe témoin adéquat ni à expliquer les différences qui subsistent entre les deux groupes dans l'analyse affaiblit la crédibilité des estimations et peut faire échouer les tentatives visant à exclure les autres explications aux effets observés.

Dans le cadre des interventions cofinancées par le FSE, l'approche d'évaluation quasi expérimentale la plus souvent applicable est l'approche par valeur de référence/suivi en deux groupes. Cette approche comprend un groupe témoin et un groupe de traitement, comme pour l'approche aléatoire, sauf que le groupe témoin est constitué (de manière non aléatoire) parmi les groupes existants qui ne participent pas pour qu'il soit le plus semblable possible au groupe de traitement.

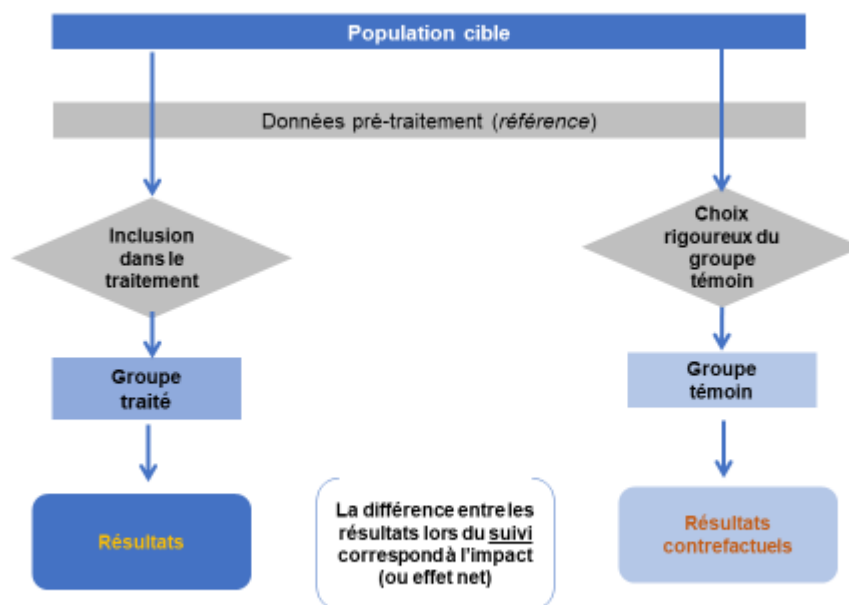
Le groupe traité et le groupe témoin doivent être semblables

Une bonne stratégie pour trouver un groupe témoin pertinent dans une situation quasi expérimentale consiste à sélectionner des candidats témoins qui ont été exclus de l'intervention sur la base de facteurs sans rapport avec leurs caractéristiques et les résultats potentiels. Certaines situations peuvent laisser croire qu'il est possible, même si les groupes témoins n'ont pas été formés explicitement de manière aléatoire, de trouver ultérieurement des personnes ou des entreprises dont la non-participation au traitement s'avère aléatoire par rapport aux résultats potentiels. Si de tels cas surviennent, ils s'approchent d'un idéal dans le cadre d'une approche quasi expérimentale. À titre d'exemple, certains membres d'un groupe cible pour une intervention peuvent être exclus de la participation à l'intervention en raison d'une erreur

⁷⁷ À proprement parler, le terme «groupe témoin» est utilisé pour les approches expérimentales et les approches quasi expérimentales parlent généralement de «groupe de comparaison». En pratique, toutefois, ces termes sont souvent utilisés de manière interchangeable. Le présent guide utilise donc le terme «groupe témoin» pour désigner le groupe utilisé pour estimer le résultat contrefactuel, indépendamment de l'approche de l'EIC.

ou d'une méprise administrative. Il est donc extrêmement important de comprendre le processus de sélection dans le traitement afin de constituer un groupe témoin pertinent - on ne le soulignera jamais assez.

Figure 7 Approche quasi expérimentale stylisée avec groupe traité et groupe témoin



Il existe de nombreuses manières de former un groupe témoin crédible. La première consiste en un appariement statistique: c'est-à-dire que les données sont collectées auprès des personnes traitées, d'une part, et d'un échantillon (généralement très large) de personnes non traitées, d'autre part. Un groupe témoin est alors constitué à partir du groupe de personnes non traitées en retenant celles qui sont les plus semblables aux personnes du groupe traité. Le terme «semblable» fait référence à un ensemble de caractéristiques socio-démographiques comme l'âge, le sexe, le niveau d'éducation, l'emploi et le chômage, etc. mesurées à un moment précédent l'entrée du groupe traité dans le programme. On procède alors à l'«appariement» des personnes non traitées et des personnes traitées. En pratique, la liste potentiellement longue des caractéristiques socio-démographiques peut être résumée à l'aide de la valeur du «score de propension», qui en facilite l'application.

Appariement des personnes traitées et non traitées

3.2.2. Appariement par score de propension (propensity score matching)

Principales caractéristiques de l'appariement par score de propension

- L'appariement imite la répartition aléatoire en constituant ex post un groupe témoin aussi semblable que possible au groupe traité en ce qui concerne ses principales caractéristiques.
- Contrairement à la répartition aléatoire, seules les caractéristiques observables (âge, sexe, niveau d'éducation, etc.) peuvent être appariées,

alors qu'il n'est pas possible de prendre en considération les caractéristiques non observables (par ex. la motivation).

- La validité de cette approche dépend fortement de la disponibilité des données.

L'appariement par score de propension (ASP) consiste à estimer un modèle statistique pour l'échantillon entier (personnes traitées et candidats potentiels) qui fournit une estimation de la propension à participer de chaque personne ou entreprise - indépendamment du fait qu'elle participe ou non⁷⁸. Les personnes ou entreprises traitées sont ensuite appariées - soit avec une personne ou entreprise non traitée, soit avec plusieurs personnes ou entreprises non traitées - en fonction du score de propension⁷⁹. Cette procédure permet de définir un groupe témoin qui peut servir ultérieurement à déduire une estimation du résultat contrefactuel. Un tel appariement garantit une prise en considération dans les estimations d'impact des différences observables entre le participant à un programme et les témoins et permet, à condition que toutes les différences antérieures à l'appariement soient observables, d'obtenir une estimation impartiale des effets de l'intervention. Toutefois, si la sélection des bénéficiaires se fonde sur des facteurs non issus de l'observation, un point d'interrogation subsistera quant à l'adéquation de l'appariement avec sa capacité à éliminer les biais. L'hypothèse décisive qui sous-tend l'approche d'appariement veut que la procédure de sélection soit caractérisée par des données observables.

La propension à participer comme manière de définir le groupe traité et le groupe témoin

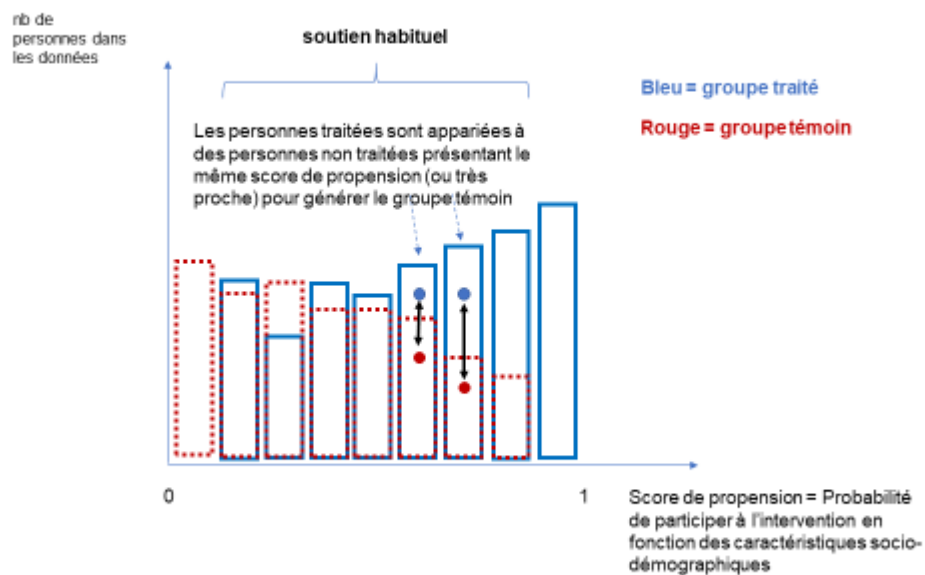
La figure ci-dessous représente une illustration intuitive et simplifiée de l'approche d'appariement par score de propension. L'ordonnée représente le nombre de personnes dans le groupe traité et le groupe non traité, classé en fonction du score de propension en abscisse. Généralement, les personnes traitées ont tendance à avoir des scores de propension relativement plus élevés, tandis que les personnes non traitées ont tendance à avoir des scores de propension plus bas. La zone où les scores de propension des deux groupes se rejoignent est désignée comme la partie de soutien commun.⁸⁰C'est là que les cas traités et non traités sont appariés entre eux. Deux exemples sont donnés dans le diagramme, mais le processus est en fait répété jusqu'à ce que tous les cas traités et non traités aient été appariés dans la partie de soutien commun. Dans la figure, l'appariement se fait entre «voisins les plus proches». Par «voisin le plus proche» d'un membre du groupe traité, on entend le membre du groupe témoin qui affiche le score de propension le plus proche. Une fois que les deux groupes sont formés, leurs résultats moyens peuvent être comparés pour obtenir une estimation de l'impact. Dans la pratique, l'appariement par score de propension peut devenir un processus très complexe incluant une variété de points à prendre en considération. La plupart de ces points sont ignorés dans le cas d'espèce pour veiller à la clarté des principes essentiels. L'encadré suivant donne un exemple pratique d'une évaluation du FSE par l'approche d'appariement.

⁷⁸ Pour simplifier cette analyse, nous considérons que les responsables politiques souhaitent connaître l'effet de l'intervention sur ceux qui ont effectivement bénéficié des services du programme (dans de nombreux cas, il s'agit d'un sous-ensemble du groupe cible ayant reçu cette possibilité). On parle d'une «analyse TOT» (treatment on the treated, intervention sur le participant).

⁷⁹ Il existe une grande variété d'approches possibles pour un appariement par score de propension. Pour un aperçu accessible, voir [Caliendo M., Kopeinig S., 2008](#).

⁸⁰ L'étendue de la partie de soutien commun a des conséquences sur la taille de l'échantillon et l'utilité des résultats pour le projet, d'autant plus lorsqu'un grand nombre de cas traités se situent en dehors de cette partie.

Figure 8 Illustration de l'approche par score de propension



Une sélection sur la base de données observables. La plausibilité de l'approche par score de propension repose notamment sur l'hypothèse selon laquelle la sélection parmi les bénéficiaires peut être entièrement caractérisée par des données observables. Autrement dit, il n'y a aucune différence non observée entre le groupe de traitement et le groupe témoin liée aux résultats et/ou à la décision de participer à l'intervention. La plausibilité de cette hypothèse est confirmée par l'intégration d'un riche éventail de variables dans l'estimation des scores de propension, variables sélectionnées sur la base des connaissances et théories préalables. Plus précisément, dans le cadre des interventions sur le marché du travail, il est fortement recommandé d'inclure les antécédents individuels sur le marché du travail lorsque l'on cherche d'éventuelles différences non observées⁸¹.

Sélection basée sur les données observables

Encadré 15 Exemple d'une évaluation par l'approche d'appariement⁸²

Impacts des interventions financées par le FSE en 2014-2020 concernant le chômage de longue durée dans la région des Marches⁸³

Une approche d'appariement a été utilisée pour évaluer l'impact d'un certain nombre d'interventions du FSE (stages, bourses, expérience professionnelle dans des communes et chèques formation) pour les chômeurs de longue durée financées dans la région des Marches en Italie. Les interventions ne ciblaient pas spécifiquement les chômeurs de longue durée, mais ces derniers représentaient la majorité des participants à toutes les interventions au titre du FSE analysées.

L'objectif de l'évaluation consistait à déterminer l'impact des mesures du FSE sur la probabilité d'être employé à 6, 9, 12, 15 et 18 mois après le début des interventions. Plusieurs variables de résultat ont été utilisées: la probabilité d'être employé à un moment donné après les interventions, la probabilité d'être employé en vertu d'un contrat à durée indéterminée et le nombre de jours travaillés au cours d'une période après les interventions.

⁸¹ Voir [Caliendo, Mahlstedt and Mitnik, 2017](#) et [Kluve, Lehmann, and Schmidt, 2008](#).

⁸² D'autres exemples d'application pratique des méthodes (ASP, approche de discontinuité de la régression et double différence) sont disponibles à l'adresse suivante [Commission européenne, 2020](#).

⁸³ Voir [Pompili M., Giorgetti I., 2020a](#)

Une approche de score de propension a été adoptée (appariement entre «voisins les plus proches») afin d'identifier les différences entre les groupes traités et non traités et les quatre méthodes d'interventions au titre du FSE susmentionnées ont fait l'objet d'analyses distinctes. Pour chaque intervention, le groupe traité se composait de participants qui avaient commencé l'intervention avant la fin du mois d'août 2019 (526 pour les stages, 1 058 pour les bourses, 236 pour les expériences professionnelles dans les communes et 241 pour les chèques formation); le groupe témoin se composait de chômeurs inscrits aux SPE au cours de la période 2016-2018, ayant connu une durée de chômage d'au moins 12 mois (77 255 inscriptions).

Plusieurs variables ont été prises en considération pour calculer l'ASP tiré des caractéristiques socio-démographiques des personnes traitées et non traitées, comme: le sexe, l'âge, la nationalité, le niveau d'éducation, le lieu de résidence, la date d'entrée dans l'intervention et les antécédents professionnels remontant à 36 mois avant l'intervention.

Pour mesurer les variables de résultat pour le groupe traité et le groupe témoin (avant et après les interventions), on a eu recours à des données provenant des COB (les archives des notifications obligatoires de contrat de travail effectuées par les sociétés auprès des SPE).

Les résultats de l'étude sont mitigés, avec des impacts positifs déclarés pour les stages et les bourses, tandis que les expériences professionnelles dans les communes ont produit des effets négatifs et que les chèques formation n'ont entraîné aucun effet significatif⁸⁴.

3.2.3. Double différence

Caractéristiques essentielles de la double différence

- L'approche de la double différence est une approche intuitive qui compare les différences avant/après en matière de résultats entre le groupe traité et le groupe témoin.
- Puisque le changement connu par le groupe témoin avec le temps mesure ce qui serait arrivé au groupe traité en l'absence d'intervention (le scénario contrefactuel), toute différence dans les résultats avant/après du groupe traité indique l'effet de causalité de l'intervention.
- Cette méthode est simple et peut être appliquée dans de nombreux cas.

Que ce soit séparément ou en parallèle avec l'appariement, les mesures de référence (ou de prétraitement) des variables de résultat peuvent servir à produire des estimations de la double différence. Dans ce cas, la différence d'un résultat avant et après le traitement dans un groupe témoin est déduite de cette même différence observée dans le groupe traité afin d'obtenir une estimation de l'impact d'une intervention. À nouveau, il est essentiel de sélectionner un groupe témoin plausible. Les impacts calculés sur la base de la double différence sont généralement observés au sein d'un cadre de régression qui représente également d'autres différences observées entre les groupes traités et témoins. Par ailleurs, cette approche vérifie les différences non observées entre les deux groupes, stables au fil du temps, mais également les différences qui évoluent avec le temps et qui touchent de la même manière les groupes témoins et traités (par exemple, des facteurs affectant l'économie dans son ensemble). En raison de cette capacité à vérifier certains aspects des différences non observées entre les groupes traités et témoins, l'approche de la double différence présente dans la plupart des cas une amélioration par rapport à une stratégie d'appariement par

Avant et après le traitement

⁸⁴ En ce qui concerne les chèques formation, comme l'a souligné le rapport d'évaluation, les constatations peuvent être considérées comme des constatations préliminaires puisque l'échantillon examiné est trop petit au regard du nombre total de participants.

échantillon. La figure 4 propose une représentation visuelle de l'approche de la double différence.

L'abscisse représente le passage du temps et l'ordonnée représente une échelle sur laquelle les résultats sont notés. Admettons, dans le cas d'espèce, que les résultats représentent les salaires. YT1 indique les salaires moyens du groupe de traitement dans la période de prétraitement, YC1 ceux du groupe témoin. Dans la période de post-traitement, YT2 et YC2 indiquent respectivement les salaires des groupes traités et témoins. Par conséquent, la ligne continue supérieure représente l'évolution des salaires au sein du groupe de traitement et la ligne continue inférieure l'évolution au sein du groupe témoin.

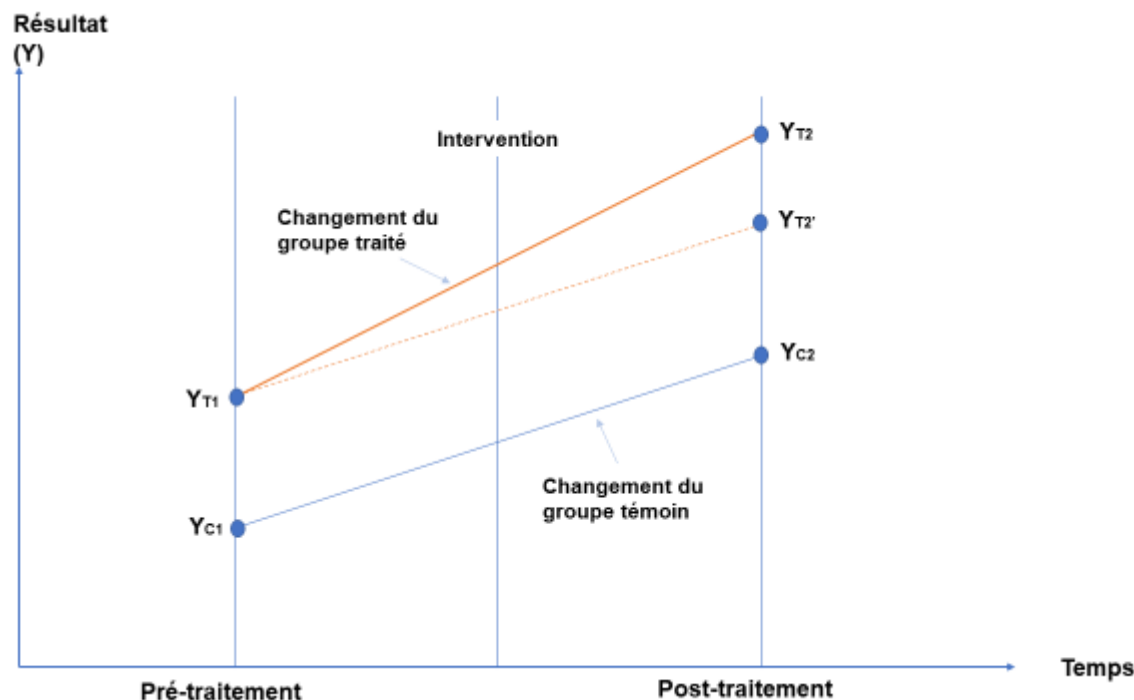
Une estimation brute de l'impact de l'intervention serait possible en comparant les salaires des groupes traités et témoins dans la période de post-traitement, autrement dit $YT2 - YC2$. Ce calcul serait toutefois incorrect puisqu'il ignore les différences des salaires de prétraitement. Une manière d'envisager l'estimation de la double différence consiste à la calculer en soustrayant une estimation du biais de prétraitement de la différence de résultats en post-traitement. En conséquence, la différence de salaires de post-traitement ($YT2 - YC2$) est ajustée en lui soustrayant la différence des salaires de prétraitement ($YT1 - YC1$) et, par conséquent, l'estimation de l'impact par la double différence peut être exprimée très simplement par la formule suivante:

$$(YT2 - YC2) - (YT1 - YC1).$$

La logique de la double différence

Si les différences de salaires de post-traitement ne sont pas ajustées en fonction des différences préexistantes entre les groupes traités et témoins, il peut en découler des estimations biaisées. Autrement, il a déjà été dit que l'approche de la double différence peut être envisagée en soustrayant le changement de résultats au sein du groupe témoin au changement observé dans le groupe de traitement. Le changement observé dans le groupe témoin est perçu comme celui qui serait survenu dans le groupe de traitement en l'absence de l'intervention.

Figure 9 Illustration de l'approche des doubles différences



Dans le scénario le plus simple, la principale hypothèse sur laquelle repose l'approche de la double différence est celle de la tendance commune: à savoir que l'évolution dans le temps des résultats dans les groupes traités et témoins est la même sans traitement. Il n'est pas possible de tester directement cette hypothèse, bien que la plausibilité de l'approche soit étayée par les cas où de nombreuses mesures de prétraitement sur les variables de résultat sont disponibles pour le groupe traité et le groupe témoin et où ces mesures montrent des tendances parallèles. Voir l'encadré suivant pour un exemple d'approche de la double différence.

Hypothèse principale: des tendances communes sont équivalentes en l'absence de traitement

Encadré 16 Exemple d'une évaluation par l'approche de la double différence

Évaluation des mesures destinées aux travailleurs âgés dans la région de Lublin financées par le FSE au cours de la période de programmation 2007-2013⁸⁵

L'objectif principal de l'étude consistait à évaluer les interventions au titre du FSE visant à étendre l'activité économique des personnes âgées dans le contexte d'une situation démographique défavorable dans la région de Lublin en Pologne. Les interventions étaient financées au titre du programme opérationnel «Capital humain» 2007-2013.

Le groupe traité, un échantillon de 145 personnes, se composait de personnes qui étaient au chômage au moment où elles ont intégré les interventions. Le groupe témoin a été extrait des données anonymisées des registres fournis par les bureaux pour l'emploi, contenant des données socio-démographiques ainsi que les antécédents concernant les événements inscrits au registre portant sur le statut professionnel de personnes spécifiques (les données ont été fournies par huit bureaux pour l'emploi et contenaient des informations relatives à 67 102 personnes). La variable de résultats correspondait au «statut d'inscrit au registre du chômage»⁸⁶ et pouvait être observée avant et après l'intervention, permettant la mise en œuvre d'une approche de la double différence. Le statut professionnel a été mesuré à 5 moments: 6 mois avant la participation au projet, pendant la participation au projet et 6, 12 et 18 mois après la fin du projet.

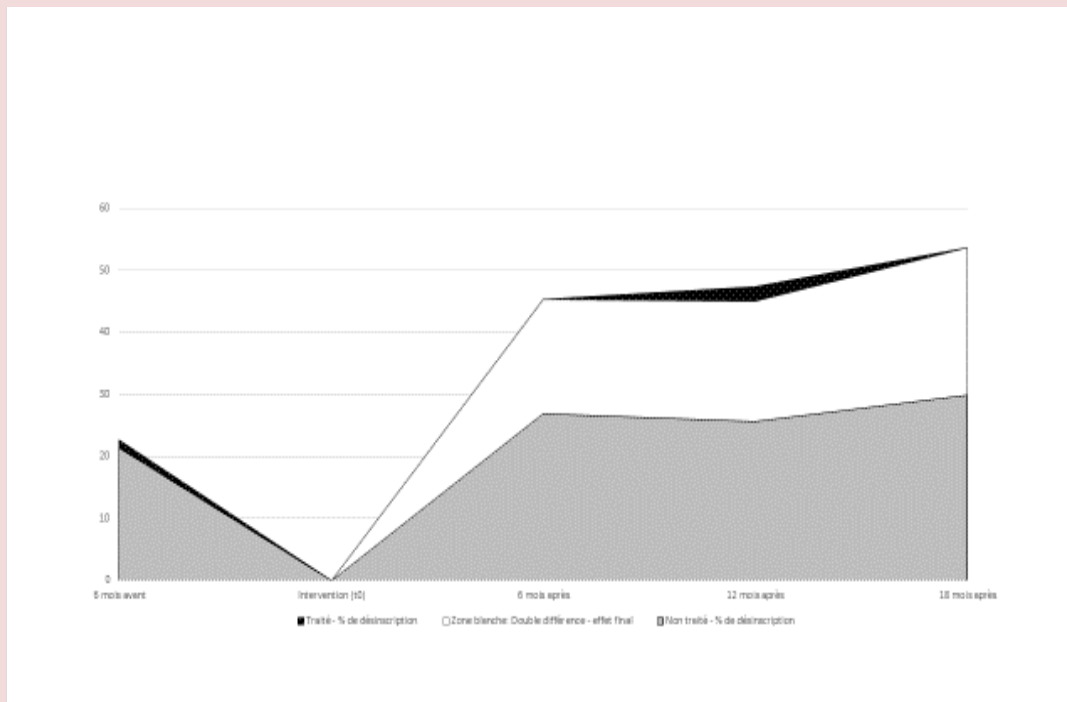
La sélection du groupe témoin dans cette approche de la double différence a été améliorée à l'aide de

⁸⁵ Voir [Re-source Pracownia Badań i Doradztwa, 2015](#).

⁸⁶ Bien que le fait d'être enregistré comme «chômeur» ne soit pas équivalent au fait d'être «chômeur», il était nécessaire de formuler cette hypothèse pour pouvoir mener l'étude.

méthode d'appariement permettant de trouver une personne non traitée pour chaque personne traitée, en s'appuyant sur un certain nombre de covariables observables, comme le sexe, l'âge, le niveau d'éducation, le taux de chômage et le lieu de résidence.

L'analyse a montré que le soutien apporté aux chômeurs âgés de plus de 50 ans a produit un effet net positif sur l'emploi: les chômeurs âgés de plus de 50 ans traités avaient 2,5 fois plus de chances de se désinscrire en tant que chômeurs qu'une personne non traitée. La figure ci-dessous montre l'évolution de la variable de résultats pour le groupe traité et le groupe témoin, ainsi que l'effet net estimé une fois prise en considération l'hypothèse parallèle de l'approche de la double différence.



3.2.4. Approche de la régression par discontinuité

Caractéristiques essentielles de la discontinuité de la régression

- L'approche de la régression par discontinuité se fonde sur l'idée qu'une valeur spécifique («point limite») d'un score ou d'une note détermine si une personne participe ou non à l'intervention.
- Les personnes proches de cette valeur sont ensuite considérées comme comparables, à la seule différence que celles se trouvant d'un côté du point limite participent (le groupe traité) et celles se trouvant de l'autre côté ne participent pas (le groupe témoin).
- La comparaison de ces groupes donne généralement une mesure précise et intuitive des effets de l'intervention; mais la méthode n'est applicable que si un processus de sélection fondé sur un score ou une note est en place.

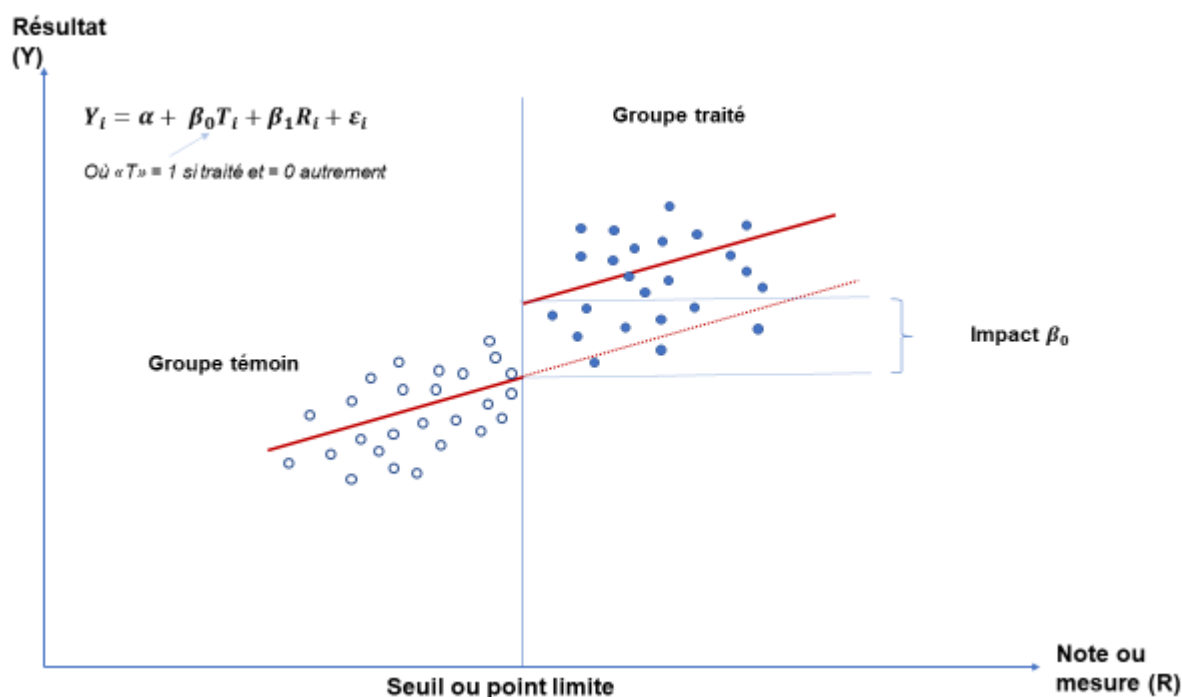
Une approche de la **régression par discontinuité** peut être préférée lorsque l'accès à une intervention est fixé selon un point limite parallèlement à une évaluation, une échelle ou une mesure continue. Par exemple, l'accès à une formation peut dépendre des résultats à un test d'aptitude, où seuls ceux qui

Un point limite sépare les personnes appartenant au groupe traité et

obtiennent un résultat supérieur au seuil fixé (ou au point limite) accèdent à la formation. Pour que l'approche soit valable, le point limite devrait être déterminé sans connaître les scores des personnes formées potentielles; les candidats qui se situent à proximité immédiate du point limite seront très semblables entre eux, étant entendu que les candidats juste au-dessus participeront à l'intervention, à l'inverse de ceux juste en dessous. La comparaison des résultats des candidats au-dessus et en dessous du point limite fournit une estimation de l'impact de l'intervention au point limite.

celles appartenant au groupe témoin

Figure 10 Illustration de l'approche de discontinuité de la régression



Une approche de la régression par discontinuité peut être adoptée lorsque le point limite définit complètement le groupe de traitement (conformité parfaite), auquel cas il s'agit d'une discontinuité *stricte*, ou lorsque, dans certaines conditions, tous les candidats d'un côté donné du point limite ne respectent pas à la lettre leur consigne dans le traitement (discontinuité *floue*).

Discontinuité stricte ou floue

La figure 10 ci-dessus montre un exemple stylisé d'une approche de la régression par discontinuité. Il s'agit là de l'illustration la plus simple d'une discontinuité stricte. L'intervention produit des effets constants à chaque valeur de la note et les impacts sont estimés selon un modèle de régression linéaire (la forme fonctionnelle de la régression de l'impact n'intervient pas). En pratique, l'analyse devra invariablement être beaucoup plus élaborée que celle présentée à la figure 10.

Les points de la figure 10 représentent les différentes unités, par exemple les personnes en formation. L'abscisse indique la note ou la mesure utilisée pour répartir ces personnes entre les groupes du programme de formation. Les personnes dont la note se trouve à droite de la ligne verticale continue indiquant le seuil pour cette note ou mesure (un test d'aptitude par exemple) intègrent la formation et forment le «groupe traité». Les stagiaires potentiels ayant un résultat en deçà du seuil pour la note ou la mesure ne participent pas à la formation et forment le groupe témoin.

L'élément essentiel est que la note utilisée pour répartir le groupe cible entre groupe de traitement et groupe témoin soit une variable quantitative continue mesurée avant le traitement et qu'une personne est admise dans la formation selon que ses résultats sont supérieurs ou inférieurs à un point limite ou un seuil prédéfini.

Le résultat est illustré sur l'axe de l'ordonnée. En substance, l'impact du traitement est déterminé en appliquant sur les données un modèle estimatif de régression linéaire (vu l'hypothèse susmentionnée); il s'agit de faire régresser la variable de résultat par rapport à la mesure de la note, parallèlement à une variable factice (un indicateur de traitement) qui définit si un score se situe au-dessus ou en dessous du point limite (si la personne est placée dans le groupe traité ou dans le groupe témoin).

Cette équation de régression de l'impact est représentée à la figure 10. L'effet ou l'impact de la formation dans notre exemple s'obtient à partir du coefficient sur l'indicateur de traitement, soit β_0 ⁸⁷. Il s'agit effectivement de vérifier s'il y a une coupure ou une discontinuité à hauteur du point limite, représenté dans la figure 10 par un déplacement vers le haut de la ligne de régression à partir du seuil ou du point limite. Dans notre exemple, on observe un impact positif de la formation sur le résultat.

Une autre manière de comprendre l'estimation de l'impact est d'observer l'extension en pointillé de la ligne du groupe témoin à la figure 10. On peut considérer qu'il s'agit d'une estimation contrefactuelle relative au groupe de traitement - le rapport entre la mesure de la note et la mesure du résultat qui aurait prévalu en cas d'absence de l'intervention - la différence entre la ligne en pointillé et la ligne d'évolution du groupe de traitement représente l'effet ou l'impact du traitement. Notons qu'en l'absence de traitement, il n'y a aucune discontinuité dans la ligne et l'on considère alors que le résultat varie continuellement en fonction de la note ou de la mesure.

Cette approche fonctionne parce que les observations réalisées sur le groupe de traitement et le groupe témoin à proximité du point limite sont similaires, hormis le fait que, dans l'exemple en question, les unités au-dessus du point limite reçoivent une formation, à l'inverse des autres. La situation ne diffère donc pas de l'approche aléatoire pour les observations proches du point limite. Il y a toutefois une réserve non négligeable. Dans la plupart des cas, l'estimation d'un impact à l'aide d'une approche de discontinuité de la régression informe seulement le responsable politique des effets pour les personnes proches du point limite ou seuil. Des généralisations peuvent être faites pour les personnes plus éloignées du seuil, mais seulement de manière restreinte.

L'approche de discontinuité de la régression peut s'avérer utile lorsque les candidats sont désignés pour une intervention sur la base du besoin mesuré par une note continue ou un résultat. Toutefois, l'analyse peut se compliquer lorsque le point limite est flou et en l'absence de conformité et lorsque des problèmes de forme fonctionnelle se posent dans le modèle de régression de l'impact. De fait, un éventail d'hypothèses doivent être envisagées et la validité des hypothèses ne peut pas toujours être confirmée.

Bonne validité interne mais la généralisation peut être limitée

⁸⁷Dans un scénario simple, il s'agirait de l'effet de l'intention d'apporter un traitement à partir du point limite (voir [Bloom H. S., 2009](#)).

L'encadré suivant montre un exemple pratique d'utilisation d'une approche de la régression par discontinuité pour l'évaluation d'un système de garantie pour la jeunesse.

Encadré 17 Exemple d'une évaluation par l'approche de discontinuité de la régression

Formation professionnelle pour les jeunes chômeurs en Lettonie: éléments de preuve issus d'une approche de la régression par discontinuité⁸⁸

Les chercheurs du Centre commun de recherche de la Commission européenne ont utilisé une approche de la régression par discontinuité afin d'évaluer l'impact produit par un programme de formation professionnelle sur les résultats sur le marché du travail de jeunes chômeurs en Lettonie, financé au titre d'un système de garantie pour la jeunesse pour la période 2014-2020 et ciblant les jeunes âgés de 15 à 29 ans ne travaillant pas, ne suivant pas d'études ou de formation (NEET).

Les données utilisées provenaient de l'Agence nationale lettone pour l'emploi, elles ont fourni des informations sur les participants et les non-participants inscrits en tant que chômeurs à des dates spécifiques puis elles ont été mises en correspondance avec des données provenant du service national des impôts, qui contiennent des informations sur les revenus d'une personne à des dates spécifiques avant et après le programme (janvier 2012 et juin 2017). Après le nettoyage des ensembles de données, l'échantillon final du groupe traité comportait 898 personnes et celui du groupe témoin comportait 10 717 personnes.

Les évaluateurs se sont appuyés sur une règle du gouvernement lettone qui donne la priorité aux chômeurs âgés de moins de 25 ans pour la participation au programme de formation professionnelle. Par conséquent, l'âge représente la variable décisive qui détermine la probabilité de la participation au programme de formation professionnelle, l'âge de 25 ans représentant le seuil en-deçà duquel la participation au programme augmente considérablement en raison des règles en matière de priorité. Puisque les personnes ne peuvent pas contrôler leur âge, l'admission au programme de formation professionnelle peut être considérée comme aléatoire à hauteur du seuil. Les règles en matière de priorité et le fait que la participation était volontaire ont contraint les chercheurs à appliquer une approche de la régression par discontinuité floue. La stratégie d'identification s'appuyait sur le fait que les personnes proches du point limite (âge de 25 ans) étaient semblables, en dehors de leur participation au programme de formation professionnelle.

Les constatations ont montré que l'impact sur l'emploi futur et le revenu mensuel probables jusqu'à 3,5 ans après l'intégration au programme est positif mais pas statistiquement significatif. Un effet positif de la règle en matière de priorité pour la participation au programme a cependant été observé. Puisque la validité d'une approche de discontinuité de la régression floue dépend du fait que les participants potentiels ne peuvent pas contrôler la variable décisive (l'âge), plusieurs essais ont été mis en œuvre pour confirmer cet effet.

⁸⁸ Voir [Bratti M. et al., 2018.](#)

3.2.5. Variables instrumentales

Caractéristiques essentielles des variables instrumentales

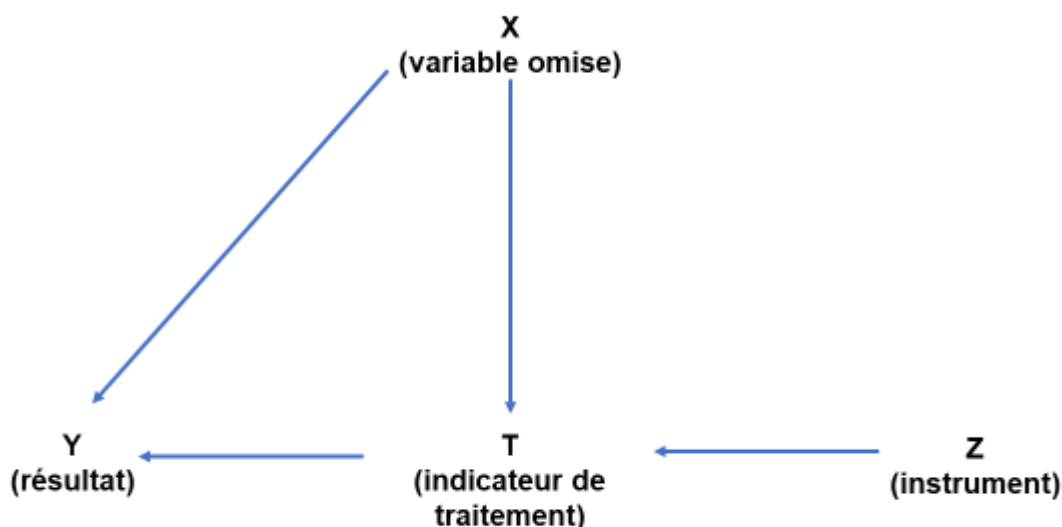
- Le principe des variables instrumentales est que certains processus (prédéterminés/exogènes) autres que la procédure de sélection réelle déterminent la participation.
- Le processus prédéterminé, et ses variables essentielles, peuvent être utilisés pour tenir compte de toute auto-sélection par les personnes lorsqu'elles décident ou non de participer.

Dans l'approche par **variables instrumentales**, la sélection pour le traitement est au moins en partie déterminée par un facteur exogène qui n'est en rien lié aux résultats si ce n'est par le traitement. Ainsi, le facteur exogène a une influence sur la participation, mais pas les résultats directement. Généralement, ces facteurs exogènes sont des erreurs ou des omissions administratives ou d'autres variations aléatoires du traitement reçu.

Comment les facteurs exogènes affectent la participation

La figure 11 illustre l'approche par variables instrumentales. Quatre variables sont représentées dans un système de causalité très simplifié. Les variables représentent des données recueillies auprès d'une population visée hypothétiquement par une intervention de formation (ceux qui reçoivent une formation et ceux qui agissent comme groupe témoin).

Figure 11 Illustration d'une approche par variables instrumentales



«Y» représente le résultat examiné. Dans le cas d'une intervention de formation, il pourrait s'agir des revenus. «T» est un indicateur qui montre si un individu a suivi la formation⁸⁹.

«X» est une variable omise qui n'est pas observée, mais qui est liée à la fois au résultat «Y» et à l'indicateur de traitement «T»; si l'on reprend l'idée d'un

⁸⁹ En d'autres termes, il y a une pleine conformité, et tous les membres du groupe de traitement participent à la formation.

programme de formation, il peut s'agir par exemple d'une mesure de référence de la capacité. Dans ce cas, la capacité est liée à la participation à la formation et aux revenus. Par exemple, davantage de membres compétents du groupe cible peuvent choisir de suivre la formation tout en profitant de salaires plus élevés.

L'existence de «X» motive la recherche d'un instrument et signifie que l'impact de la formation sur les revenus est faussé par son existence. En d'autres termes, l'estimation est biaisée en raison de l'existence de «X» et du fait que cette variable n'est pas observée et ne peut pas être directement prise en considération dans l'analyse.

Enfin, la variable «Z» est un instrument qui, selon Morgan et Winship⁹⁰, peut être considéré comme un élément influençant «T» (choc) et qui est indépendant de «X». C'est pour cette raison qu'aucune ligne de la figure 11 ne relie Z à X. De plus, «Z» n'affecte «Y» qu'en passant par «T», il n'existe aucune autre voie par laquelle «Z» affecte «Y». Cela signifie que «Z» peut être utilisé pour générer une variation dans «T» (le traitement) qui n'est pas corrélé avec la variable parasite «X». Ainsi, une mesure objective de l'effet de «T» sur «Y» peut être obtenue en exploitant cette variation⁹¹.

L'identification d'un instrument valide est cruciale

Les circonstances les plus simples dans lesquelles une approche par variables instrumentales pourrait être adoptée sont décrites ci-dessous, en excluant nombre des complexités impliquées. En pratique, il est souvent difficile de trouver un instrument convaincant. La plausibilité des différents instruments possibles dépend fortement du contexte et les hypothèses d'identification sous-jacentes ne peuvent en général pas être testées de manière statistique (en réalité, il est possible de tester de manière statistique la corrélation nécessaire entre «Z» et «T», mais pas l'«indépendance» de «Z» et «X»). Par exemple, une stratégie pourrait consister à utiliser la distance qui sépare les centres de formation (l'emplacement physique de la formation) du domicile d'un stagiaire comme un instrument pour estimer l'effet de la formation sur la rémunération nette des stagiaires. On peut observer que les stagiaires qui vivent plus près des centres de formation sont plus susceptibles de participer à un programme de formation et que la distance entre le domicile d'un stagiaire et un centre de formation n'est pas liée à d'autres déterminants des revenus et de la participation à la formation (par exemple, des mesures du capital humain). Dès lors, cette mesure de distance ne peut influencer les revenus que par son effet sur la formation⁹².

Les variables instrumentales peuvent être utilisées dans une large gamme de contextes. Les estimations peuvent être obtenues grâce à différentes méthodes d'estimation en fonction de la variable étudiée. Jusqu'à présent, cette approche n'a pas été souvent utilisée dans les évaluations du FSE. L'encadré 18 ci-dessous illustre un exemple d'analyse des effets de causalité d'une politique appliquant une approche par variables instrumentales.

⁹⁰ Morgan S.L., Winship C., 2014.

⁹¹ L'effet de causalité de «T» sur «Y» est calculé en présence d'un instrument en estimant la relation entre «Z» et «Y», puis en le divisant par la relation estimée entre «Z» et «T».

⁹² L'interprétation des résultats de cette analyse peut être compliquée par le fait que l'instrument est en corrélation avec la variation de l'effet du traitement (voir [Bryson, et al., 2002](#)).

Encadré 18 Exemple d'une étude réalisée selon l'approche par variables instrumentales**Effets sur l'emploi d'une formation linguistique dispensée aux migrants sans emploi⁹³**

L'acquisition de compétences dans la langue du pays hôte est un facteur important pour les migrants s'agissant d'atteindre un niveau élevé d'intégration sur le marché du travail.

Dans une étude récente (Lang J., 2021), l'auteur utilise une approche par variables instrumentales afin d'estimer l'effet de causalité produit par un programme de formation à la langue allemande à des fins professionnelles sur les résultats en matière d'emploi de migrants participant au programme deux ans après leur inscription. Le programme a été mis en œuvre par le bureau fédéral allemand pour la migration et l'asile (BAMF) et financé par le FSE; le programme FSE-BAMF a fonctionné jusqu'à la fin de l'année 2017. Plusieurs composants formaient ces cours: formation à la langue allemande, renforcement des compétences professionnelles et placements professionnels. Les personnes issues de l'immigration et actives sur le marché du travail allemand étaient admissibles, quelles que soient leur nationalité ou leur date d'immigration.

Pour traiter l'hétérogénéité non observée en ce qui concerne les compétences linguistiques, l'auteur utilise la variation exogène dans l'intensité de la formation à la langue locale au niveau de l'agence pour l'emploi. Cet instrument est considéré comme valide puisque les agences pour l'emploi disposent d'un large pouvoir d'appréciation en ce qui concerne la mise en œuvre de différentes combinaisons de programmes en fonction de leur «style politique» et que cette variable est exogène aux résultats en matière d'emploi des chômeurs. Les cours mis en œuvre en 2014 ont été analysés.

L'évaluation s'appuie sur un large ensemble de données administratives, appelé «biographies intégrées sur l'emploi» (BIE), une base de données fusionnée d'informations administratives de l'agence fédérale allemande pour l'emploi. Les BIE comprennent des informations relatives à la participation à la formation linguistique, des informations détaillées concernant les antécédents professionnels (sauf le travail indépendant), la recherche d'emploi et les paiements de transfert effectués au cours de la période de chômage et les données à caractère personnel des personnes concernées. D'autres données (WGH, Werdegangshistorie) ont également été utilisées afin de disposer d'informations détaillées concernant la participation aux cours de langue dispensés par les BAMF et de couvrir, si possible, les données manquantes des BIE par des informations relatives aux périodes de travail indépendant ou de congé parental.

Les constatations montrent qu'après une période marquée par des effets de verrouillage, durant un peu plus d'un an, deux ans après l'intervention, la probabilité que les migrants participant au programme soient employés était supérieure de 9 points de pourcentage à celle du groupe témoin et cette probabilité ne se limite pas aux emplois peu qualifiés. Il a également été constaté que la fourniture rapide (peu de temps après l'arrivée en Allemagne) d'une formation linguistique favorise l'intégration des migrants sur le marché du travail.

⁹³ Voir [Lang, J., 2021](#).

Tableau 5 Comparaison des caractéristiques essentielles des principales approches d’EIC

Approche	Caractéristiques essentielles	Avantages	Données requises	Limitations
Randomisation – approche expérimentale	<ul style="list-style-type: none"> - Les personnes admissibles à la participation sont réparties aléatoirement entre le groupe traité et le groupe témoin. - La randomisation garantit que les deux groupes sont identiques (en général) en ce qui concerne toutes les caractéristiques pertinentes. - Par conséquent, le groupe témoin répond à la question contrefactuelle et la différence de résultats entre le groupe traité et le groupe témoin donne l'effet de causalité de l'intervention. 	<ul style="list-style-type: none"> - Si la randomisation est effectuée correctement, les estimations de l'impact sont «impartiales» - Les résultats sont transparents et facilement compréhensibles - Les conclusions sont moins sujettes à la qualification et au doute - Variété de variantes de conception disponibles pour faire face à une variété de contextes politiques et de circonstances d'intervention 	<ul style="list-style-type: none"> - Exigence de base visant à contrôler la sélection dans l'intervention au moyen d'une randomisation - Enregistrement des personnes et de leur répartition dans les groupes - La collecte de données de référence est essentielle - Les résultats mesurés doivent être enregistrés pour le groupe traité comme pour le groupe témoin 	<ul style="list-style-type: none"> - Bien souvent, le groupe traité et le groupe témoin ne respectent pas les règles d'affectation au traitement - Le consentement des participants est souvent nécessaire - La randomisation peut influencer sur la sélection des groupes qui participent à/ sollicitent une intervention - Si les participants sont au courant de leur statut d'affectation, cela peut modifier leur comportement et influencer les résultats - Préoccupations d'ordre éthique - Exigences considérables en matière de planification et de conception - Peut être onéreuse (mais pas nécessairement)
Appariement par score de propension	<ul style="list-style-type: none"> - L'appariement imite la répartition aléatoire en constituant ex post un groupe témoin aussi semblable que possible (en général) au groupe traité en ce qui concerne ses principales caractéristiques. - Contrairement à la répartition aléatoire, seules les caractéristiques observables (âge, sexe, niveau d'éducation, etc.) peuvent être appariées, alors qu'il n'est pas possible de prendre en considération les caractéristiques non observables (par ex. la motivation). - La validité de cette approche dépend donc fortement de la disponibilité des données. 	<ul style="list-style-type: none"> - Nécessite une bonne connaissance des processus de sélection, mais ne nécessite pas de contrôle direct sur la sélection dans l'intervention - Peut être appliquée rétrospectivement, si les bonnes données sont disponibles et dans une variété de contextes - Techniquement, une méthode semi-paramétrique d'évaluation; nécessite moins d'hypothèses paramétriques (par exemple, pas besoin d'hypothèse de régression standard) - Peut être utilisée pour évaluer des effets de traitement multiples 	<ul style="list-style-type: none"> - Identification précise des participants à l'intervention - Sources de données à partir desquelles prélever un échantillon - Concept clair de participation et bonne compréhension de la sélection au traitement - Données riches, idéalement recueillies au début, à partir desquelles effectuer l'appariement - Mesures de résultats de l'intervention pour les participants et le groupe témoin 	<ul style="list-style-type: none"> - Nécessite des quantités considérables de données qui permettent une caractérisation complète du processus de sélection - La validité dépend de la qualité des contrôles et de leur sélection rigoureuse, ainsi que du degré de soutien commun - Repose sur l'hypothèse que la sélection dans l'intervention peut être caractérisée de manière adéquate par des données observables - La gamme de différentes approches d'appariement disponibles nécessite une analyse de sensibilité - Il peut être difficile d'expliquer et d'interpréter les résultats et ceux-ci

Approche	Caractéristiques essentielles	Avantages	Données requises	Limitations
				peuvent être potentiellement ambigus
Double différence	<ul style="list-style-type: none"> - L'approche de la double différence est une approche intuitive qui compare les différences avant/après en matière de résultats du groupe traité aux différences avant/après du groupe témoin - Puisque le changement connu par le groupe témoin avec le temps mesure ce qui serait arrivé au groupe traité en l'absence d'intervention (le scénario contrefactuel), toute différence dans les résultats avant/après du groupe traité indique l'effet de causalité de l'intervention - Cette méthode est simple et peut être appliquée dans de nombreux cas en pratique. 	<ul style="list-style-type: none"> - Contrôles de certains aspects des différences non observées entre les participants et les groupes témoins - Peut être utilisée conjointement avec l'appariement - Fonctionne avec les données avant et après l'intervention, comme les données de panel (données individuelles au fil du temps) ou les données transsectorielles répétées (données sur les personnes collectées à différents moments) 	<ul style="list-style-type: none"> - Les données requises sont similaires à d'autres approches, mais avec l'exigence supplémentaire de mesurer les résultats avant l'intervention - Afin de tester les principales hypothèses, plusieurs observations de prétraitement sur les résultats sont nécessaires pour les groupes traités et témoins 	<ul style="list-style-type: none"> - Nécessite l'invocation de l'hypothèse de la tendance commune dans les résultats entre les participants et les groupes témoins - L'analyse peut devenir très complexes et être sujette à une mauvaise interprétation - Des données de prétraitement approfondies sur les résultats sont nécessaires pour tester l'hypothèse de la tendance commune - Ne peut pas être utilisée pour évaluer des effets de traitement multiples⁹⁴
Approche de la régression par discontinuité	<ul style="list-style-type: none"> - L'approche de discontinuité de la régression se fonde sur l'idée qu'une valeur spécifique («point limite») d'un score ou d'une note détermine si une personne participe ou non à l'intervention. - Les personnes proches de cette valeur sont ensuite considérées comme comparables, à la seule différence que celles se trouvant d'un côté du point limite participent (le groupe traité) et celles se trouvant de l'autre côté ne participent pas (le groupe témoin). - La comparaison de ces groupes donne généralement une mesure précise et intuitive des effets de l'intervention; mais la méthode n'est applicable que si un processus de sélection fondé sur un score ou une note est en place. 	<ul style="list-style-type: none"> - Des approches de discontinuité de la régression strictes ou floues sont disponibles. - Peut fournir les impacts impartiaux des effets du traitement sous certaines conditions. 	<ul style="list-style-type: none"> - Le choix du seuil doit être indépendant des valeurs de la notation attribuée à chaque membre d'un groupe cible. - Des données concernant les personnes appartenant au groupe traité et au groupe témoin sont nécessaires concernant la note ou la mesure, le seuil ou le point limite et les résultats. 	<ul style="list-style-type: none"> - Cette approche n'est pas valide sans une mesure ou une notation continue qui détermine le traitement. - Les analyses peuvent devenir complexes et incertaines lorsque les problèmes liés à la forme fonctionnelle de la régression de l'impact deviennent importants, lorsqu'il y a une non-conformité ou lorsque la taille de l'échantillon autour du seuil est limitée. - Des dangers peuvent survenir dans l'interprétation des conclusions et dans l'extrapolation des généralisations.
Variables instrumentales	<ul style="list-style-type: none"> - Le principe des variables instrumentales est qu'il existe des processus (prédéterminés/exogènes) autres que le 	<ul style="list-style-type: none"> - Peut fournir des estimations de haute qualité sur des effets de causalité ou des preuves de leur 	<ul style="list-style-type: none"> - Nécessite des données de référence, des données sur les résultats de l'intervention et la participation à celle- 	<ul style="list-style-type: none"> - Il peut être difficile de trouver un instrument plausible.

⁹⁴ Voir [Frolich, M., 2004.](#)

Approche	Caractéristiques essentielles	Avantages	Données requises	Limitations
	<p>processus de sélection réellement utilisé qui déterminent la participation.</p> <ul style="list-style-type: none"> - Ce processus prédéterminé peut alors être utilisé pour tenir compte de toute auto-sélection par les personnes lorsqu'elles décident ou non de participer. 	<p>existence.</p> <ul style="list-style-type: none"> - Résout le problème de la distorsion de variable omise (ou du biais de sélection). - Peut être appliquée rétrospectivement. 	<p>ci, en plus de pouvoir identifier un instrument.</p> <ul style="list-style-type: none"> - Un instrument doit être lié à la participation à l'intervention et n'affecter le résultat que sur cette base. L'instrument ne doit pas être corrélé avec d'autres déterminants de résultats. 	<ul style="list-style-type: none"> - Peut être difficile à expliquer aux non-initiés. - L'interprétation des résultats n'est pas simple; testabilité limitée de l'identification des hypothèses.

Chapitre 4. Faire évoluer le programme d'EIC

Le présent guide vise à encourager et à aider les AG à réaliser des EIC plus larges et de bonne qualité. Il apporte des conseils aux responsables de la planification et de la commande des évaluations d'impact des interventions cofinancées par le FSE+. Les sections précédentes ont abordé la planification et la mise en œuvre d'une EIC et ont examiné plusieurs questions clés qui demandent de l'attention. Cependant, la période de programmation 2014-2020 a mis en lumière un certain nombre d'autres «problèmes plus larges», ainsi que des défis auxquels il convient de faire face afin de renforcer les évaluations du FSE+ et l'utilisation des EIC.

Cette section du guide formule quelques suggestions pour aborder ces questions plus larges. Il est plus particulièrement question d'aborder les points suivants:

Problèmes plus larges à résoudre

- Le manque de connaissance des approches des EIC au sein des AG et parmi la communauté plus large des responsables politiques au sein des EM;
- L'absence de contractants externes, dûment qualifiés et expérimentés, capables de réaliser des EIC, dans les EM;
- L'abolition des obstacles juridiques auxquels il faut faire face de manière générale dans les EIC;
- La progression vers une meilleure planification des EIC;
- L'élargissement du champ d'application des EIC.

4.1. Améliorer les niveaux de compréhension des parties prenantes

Pour la période de programmation 2021-2027, le RPDC prévoit dans son article 44, paragraphe 1, que *«l'autorité de gestion réalise des évaluations des programmes au regard d'un ou de plusieurs des critères suivants: l'efficacité, l'efficience, la pertinence, la cohérence et la valeur ajoutée de l'Union, dans le but d'améliorer la qualité de conception et de mise en œuvre des programmes. Les évaluations peuvent également porter sur d'autres critères pertinents, tels que le caractère inclusif, la non-discrimination et la visibilité, et couvrir plusieurs programmes»*. Comme nous l'avons dit dans les sections précédentes, cela signifie que l'AG demandant l'EIC et les évaluateurs cherchant à obtenir le contrat doivent disposer de la capacité d'évaluation appropriée.

Dans certains cas, les services du Helpdesk Évaluation ont constaté que certaines administrations avaient des lacunes en matière de capacité de réalisation d'EIC, c'est-à-dire en matière d'appel d'offres et d'accompagnement de la mise en œuvre. En conséquence, les évaluateurs ont rencontré des difficultés pour réaliser des EIC car il n'avait pas été procédé à l'avance à l'identification de questions d'évaluation claires, de la disponibilité des données de base et de la planification éclairée de la mise en œuvre.

Parfois, notamment dans les pays de petite taille, le marché du conseil connaît des lacunes en matière de capacité à réaliser une EIC ou de fourniture d'expertise technique. En conséquence, il est grandement nécessaire de stimuler la demande comme l'offre en matière d'EIC. La question de l'offre peut s'améliorer si les AG et les EM commencent à demander des EIC ou font connaître leurs demandes pour de telles études. La rapidité des réponses à une demande accrue d'EIC dépendra des compétences préalables, de l'expérience et de l'existence d'institutions au sein des EM capables d'exécuter ces approches. Cependant, la demande peut être en partie stimulée par l'amélioration des connaissances et de la compréhension des méthodes d'EIC au sein des AG.

Stimuler l'offre et la demande

Pour répondre à ce problème, les AG peuvent notamment proposer des formations aux méthodes d'EIC à leur personnel. La formation doit souligner les avantages des méthodes d'EIC pour les AG. Par ailleurs, il convient d'insister sur les questions de responsabilisation et d'apprentissage des méthodes qui fonctionnent. Un fil conducteur pour les formations est proposé à l'annexe 2.

Mise au point de formations aux méthodes d'EIC

4.2. Développement des capacités

Une autre question soulevée lors la période 2014-2020 et mentionnée dans les sections précédentes a trait au besoin de développer la capacité de réaliser des EIC au sein des communautés de chercheurs/universitaires/consultants dans les EM. Dans certains cas, il était évident que les compétences nécessaires pour ce faire étaient disponibles dans les EM mais que des obstacles sont survenus lorsque l'on a tenté d'appliquer ces compétences dans un cadre d'évaluation (par ex. accès limité à des données utilisables ou problèmes liés à la définition d'un groupe témoin raisonnable).

Il existe plusieurs moyens d'accroître l'offre de services d'évaluation. La plupart des questions soulevées s'appliquent tant aux EIC qu'aux évaluations en général. Trois méthodes sont généralement employées pour améliorer l'offre d'évaluation:

Renforcer les institutions et les communautés de praticiens

- Nouer des liens avec les établissements d'enseignement, en particulier les universités;
- Organiser et consolider une communauté indépendante de consultants;
- Encourager l'organisation d'une communauté de professionnels de l'évaluation.

Les universités

Il est important de tisser des liens avec les universités, et ce pour deux raisons. Premièrement, le corps enseignant des universités possède peut-être les compétences et les connaissances requises pour réaliser des EIC. De nombreux microéconomistes, économétriciens et quantité de sociologues ou psychologues présentent les types de compétences nécessaires, par exemple. Dans de nombreux EM, il est possible que les compétences requises soient présentes mais que ceux qui en sont dotés n'aient pas pensé auparavant à les appliquer à l'évaluation des interventions. Peut-être ne reçoivent-ils aucune incitation dans ce sens, ce qui doit être rectifié.

Renforcement des compétences universitaires

Certains EM ont une tradition de chercheurs universitaires activement engagés dans la recherche appliquée. Dans un tel contexte, les universitaires sont habitués à travailler avec le gouvernement et les AG. Dans d'autres EM, où les universités et les universitaires ne participent pas autant aux travaux appliqués, un changement de culture est peut-être nécessaire. Une méthode efficace pour élaborer une base de prestataires dans le milieu universitaire consiste à ce que les autorités des EM et les AG financent en grande partie les coûts des centres de recherche spécialisés en méthodes d'EIC.

Les universités et les universitaires peuvent également jouer un rôle dans la formation de la prochaine génération d'évaluateurs. En travaillant étroitement avec les universités, il est possible d'encourager celles-ci à inclure des méthodes d'évaluation de programmes dans leurs cursus et, dans le cadre de ce projet, de veiller à ce que les méthodes d'EIC soient incluses dans les programmes d'enseignement. Dans certains EM, les universités peuvent également jouer un rôle dans l'organisation de cours suivis de développement professionnel sur l'évaluation d'impact et les méthodes d'EIC. Ces cours peuvent s'adresser aux responsables politiques, aux spécialistes techniques des AG mais également à d'autres prestataires potentiels comme les consultants indépendants. Les EM peuvent envisager de financer de telles formations.

*Former la
prochaine
génération*

Les consultants indépendants

Pour certaines évaluations à grande échelle, il existe un marché international. C'est certainement le cas des grandes EIC. Pourtant, de nombreux EM souhaitent développer une capacité nationale à mener des EIC. Une stratégie pour y arriver consiste à mettre en place des alliances stratégiques entre les prestataires nationaux potentiels et les consultants internationaux.

*Faire évoluer le
marché*

Plusieurs pistes pour élaborer une base nationale de prestataires chargés de réaliser des EIC sont présentées ci-après et peuvent être appliquées par les AG (ou d'autres organes) qui commandent des EIC:

- Insister sur les offres de consortiums ou de partenariats qui englobent toujours l'un ou l'autre consultant local;
- Rédiger des contrats d'évaluation qui permettent aux nouveaux arrivants nationaux sur le marché de l'évaluation de réaliser des petites évaluations à faible risque;
- Veiller à ce que les critères techniques et financiers associés aux offres d'évaluation ne soient pas trop restrictifs et permettent aux nouveaux arrivants de participer;
- Privilégier pour la sélection l'aspect technique et de savoir-faire plutôt que les procédures administratives complexes avec lesquelles les consultants moins expérimentés ne se sentent pas à l'aise;
- Organiser des réunions de briefing avec les consultants potentiels pour répondre aux questions et encourager les offres dans un environnement compétitif;
- Soutenir la prise de contacts entre les consultants en évaluation relativement isolés afin d'encourager le développement de l'esprit d'équipe, la formation de consortiums et autres réseaux et associations professionnels, y compris au niveau international; et

- faire prendre conscience à ceux qui commandent des évaluations qu'ils doivent peut-être adopter une approche de gestion plus interactive des nouveaux contractants pour accélérer l'acquisition de connaissances et d'expérience.

Encadré 19 Exemple de projet visant à renforcer la culture et les capacités en matière d'EIC

Au cours de la période de programmation 2014-2020, un projet du FSE a été mené en Suède dans le but d'améliorer la culture et la capacité d'évaluation des Services publics de l'emploi (SPE). Le projet, intitulé «Projet du fonds européen fondé sur des éléments factuels», s'est déroulé de 2016 à 2018 et a fait participer les gestionnaires et autres membres du personnel des SPE à des ateliers et autres activités de formation dans le domaine de l'évaluation.

Les activités ne se concentraient pas sur des problèmes techniques mais plutôt sur l'importance d'une évaluation, les étapes principales d'une évaluation d'impact, l'importance des données de bonne qualité et l'assurance qualité d'une évaluation. Un site internet a été mis en place, sur lequel du matériel pédagogique a été téléchargé.

Le projet a eu pour principaux résultats l'amélioration des connaissances des gestionnaires des SPE en matière d'évaluation, de leur propension à la réalisation d'évaluations et de leur attitude coopérative. Plus concrètement, selon les personnes interrogées, «les effets des projets sur le nombre d'évaluations d'impact réalisées» ne sont pas encore élevés, mais il est évident que des progrès ont eu lieu compte tenu du fait qu'aucune EIC n'avait été effectuée en Suède au cours de la précédente période de programmation 2007-2013.

L'évaluation du projet «Ung framtid» (Avenir jeune), qui a adopté une approche aléatoire, constitue un exemple concret, en pratique, de participation des gestionnaires des SPE au processus d'évaluation⁹⁵.

Créer des communautés professionnelles

Il est important de créer des communautés professionnelles d'évaluation dans les EM, qui disposent d'un espace leur permettant de discuter des méthodes d'EIC et de partager leur expérience. La création de communautés professionnelles est importante pour le soutien et l'apprentissage mutuels mais également pour le maintien des normes de qualité. Une stratégie intéressante pourrait consister à nouer des liens avec les sociétés d'évaluation nationales compétentes et à les encourager à promouvoir les EIC des interventions du FSE+ à travers des formations, des conférences ou des séminaires spécialisés ou des sessions de sensibilisation.

Création de communautés professionnelles

Partager l'expérience

La CE est favorable à des évaluations d'impact du FSE+ plus rigoureuses et a largement recommandé les EIC au cours de la période 2014-2020⁹⁶. À l'heure actuelle, nous pouvons affirmer que le recours aux EIC a augmenté, mais qu'elles restent limitées dans un certain nombre d'AG et d'EM. Le partage d'expériences en matière de méthodes d'EIC est l'un des premiers moyens pour renforcer les capacités et soutenir la diffusion des EIC dans l'UE. Les plates-formes d'apprentissage mutuel existantes, comme les évaluations par les pairs des politiques de l'emploi et de l'insertion sociale et les communautés de pratique au sein du FSE+, devraient être utilisées à cette

Utiliser les plateformes existantes

⁹⁵ Les informations s'appuient sur un entretien avec un fonctionnaire du SPE suédois Arbetsförmedlingen.

⁹⁶ L'annexe XI du RPDC pour la période 2014-2020 (règlement n° 1303/2013) préconise «l'existence d'un système d'indicateurs de résultats efficace requis pour suivre l'avancement vers la production des résultats et pour entreprendre l'évaluation des incidences». Cette exigence faisait partie des conditions ex ante pour la période de programmation 2014-2020.

fin. Les initiatives des AG ou EM individuels, comme les conférences internationales ou les séminaires, pourraient également favoriser le partage d'expériences.

4.3. S'attaquer aux obstacles juridiques

L'un des problèmes majeurs les plus significatifs rencontrés par les chercheurs qui réalisent des EIC dans les EM concerne l'accès aux données. Plus particulièrement, les chercheurs font souvent face à des obstacles juridiques destinés à protéger la confidentialité des personnes représentées dans les ensembles de données. Comme indiqué plus haut en ce qui concerne les caractéristiques du RGPD, la réponse à ces problèmes réside dans la réalisation de réformes plus larges et dans la conclusion d'accords permettant de mettre les données à disposition des évaluateurs de manière contrôlée et continue.

Lever les obstacles juridiques à l'accès aux données

Les versions analytiques des ensembles de données administratives pourraient être élaborées de manière régulière à partir des données administratives détenues par les autorités des EM, enregistrées et ensuite classées dans des archives à l'accès contrôlé. Les contractants agréés peuvent extraire des données de ces collections moyennant une autorisation. Si un consentement spécifique s'avérait impossible à obtenir, les données seraient rendues tout à fait anonymes grâce à des identifiants personnels cryptés afin de respecter le RGPD. Des collections de données de ce genre ont été créées dans plusieurs pays. Cependant, lorsque l'accès reste un obstacle en raison de divergences d'interprétation des règles en matière de vie privée, une initiative nationale au niveau gouvernemental devrait promouvoir les accords et systèmes capables de fournir des données à des fins de recherche dans un temps relativement court.

Création d'ensembles de données analytiques

Si des doutes subsistent quant à la confidentialité des données à caractère personnel, l'établissement de laboratoires de données devrait être envisagé. De cette manière, les évaluateurs travaillant sur des ensembles de données administratives auraient accès aux registres exclusivement dans des lieux sécurisés, où l'accès aux données est strictement surveillé et contrôlé. Les données seraient traitées et analysées sur place et seuls les résultats des analyses pourraient quitter les locaux.

Création de laboratoires de données

4.4. Vers des approches plus orientées vers l'avenir

Une caractéristique commune aux quelques EIC réalisées à ce jour sur des interventions financées par le FSE est qu'elles étaient toutes de nature rétrospective plutôt que prospective. Autrement dit, les experts évaluateurs ont été chargés d'évaluer des interventions élaborées sans penser à l'évaluation et, dans certains cas, pour lesquelles la planification de l'évaluation d'impact était faible, voire inexistante. Ainsi, les évaluateurs ont dû se constituer des sources de données à l'aide de moyens qui coûtent du temps, de l'argent et sont loin d'être optimaux, en fonction des données qui étaient disponibles, plutôt que des sources de données élaborées avec l'évaluation d'impact à l'esprit.

À l'inverse, une approche prospective supposerait de faire participer les évaluateurs à la planification d'une EIC à la première occasion et permettrait d'influencer les interventions (nouvelles ou existantes), souvent de manière subtile, pour les rendre plus souples aux EIC. Planifier à l'avance une EIC

Les approches orientées vers l'avenir incluent les

peut faire la différence entre pouvoir réaliser une évaluation avec rigueur et ne pas pouvoir réaliser d'évaluation du tout. Faire participer du personnel interne dûment formé ou engager des contractants externes dès le début d'une intervention ou au moment de prendre les décisions en matière de financement suppose que:

évaluateurs dès le départ

- Une tenue de registres adéquate est intégrée à la réalisation des programmes et interventions;
- Des sources de données nécessaires peuvent être trouvées dès le départ et que les questions d'accès et de protection des données sont traitées en temps voulu;
- La collecte des données de référence peut être ciblée et des enquêtes proposées le cas échéant;
- Les questions pratiques relatives à la manière dont les participants sont sélectionnés pour l'intervention sont abordées d'une manière qui montre que les procédures de sélection répondent aux critères stricts de l'évaluation.

La participation d'évaluateurs formés aux méthodes d'EIC (qu'il s'agisse d'évaluateurs internes de l'AG ou d'experts externes désignés) à la procédure d'élaboration des nouvelles interventions du FSE+, ou aux décisions relatives aux interventions existantes permettrait de faire coïncider le début de l'évaluation d'impact et le début de la période du programme et produirait des bénéfices importants en matière de prise de décisions fondées sur des données probantes.

4.5. Élargissement du champ d'application des EIC

Un dernier point très important concerne la couverture politique des EIC et implique plusieurs dimensions: les politiques couvertes, les résultats analysés et l'exhaustivité de l'interprétation des constatations.

Au cours de la période 2014-2020, la diffusion des EIC se concentrait presque exclusivement sur les politiques actives du marché du travail et sur les effets liés au statut professionnel. Cela peut probablement s'expliquer par le fait que les données administratives sur l'emploi sont généralement plus faciles d'accès et que ces données fournissent la principale variable de résultat pour l'analyse de nombreuses politiques actives du marché du travail.

Mettre davantage l'accent sur les EIC concernant les politiques d'inclusion sociale et d'éducation

Les politiques en matière d'éducation ou d'inclusion sociale n'ont pas, ou très peu, fait l'objet d'EIC, malgré leur importance en ce qui concerne la stratégie du FSE et l'attribution de fonds; l'encadré 20 ci-dessous présente quelques-uns des rares exemples disponibles en Espagne, en Pologne et au Portugal.

Afin de neutraliser ce manque d'EIC et d'améliorer les connaissances sur ces politiques importantes, un ensemble d'initiatives combinées est nécessaire:

Faire participer des acteurs supplémentaires et créer des sources de données appropriées

- La planification et la préparation anticipées d'EIC portant sur des mesures d'inclusion sociale ou d'éducation, afin de pouvoir évaluer en temps opportun la disponibilité des données; activer les collaborations nécessaires entre les différents acteurs pour accroître cette disponibilité; et définir le groupe traité et le groupe témoin à un stade précoce.
- La promotion de la participation des propriétaires des données à la conception de l'EIC; une participation plus large aux EIC des

administrations chargées des politiques éducatives et sociales, lorsqu'elles sont différentes de l'AG, pourrait produire un engagement plus fort et plus généralisé en faveur de l'évaluation et de la mise à disposition des données.

- La stimulation des accords institutionnels et des outils logiciels afin de rendre les données administratives sur les conditions sociales et l'éducation utilisables aux fins des EIC. À cet effet, la participation des instituts nationaux de statistique, en tant qu'institutions de liaison et facilitateurs, peut s'avérer importante car ces instituts disposent des compétences nécessaires et préparent déjà de nombreux ensembles de données administratives afin de produire des statistiques nationales;

L'appui aux collectes de données pertinentes auprès des participants appartenant au groupe traité et au groupe témoin au début de l'intervention, de manière à disposer d'un ensemble de données «avant/après» cohérent. Lorsque les données administratives ne sont pas disponibles ou pas pertinentes, une étude spécifique peut fournir les données nécessaires, mais elle doit commencer au début de l'intervention et également inclure le groupe témoin afin de pouvoir procéder aux comparaisons nécessaires.

Encadré 20 Exemples d'évaluations dans le domaine de l'éducation

Interventions du FSE en matière de décrochage scolaire mises en œuvre dans les Asturies⁹⁷

Dans le cadre d'une évaluation générale du PO Principado de Asturias du FSE pour la période 2014-2020, une analyse d'impact spécifique du projet intitulé «Diversificación Curricular y de Mejora del Aprendizaje y del Rendimiento» (PMAR) a été réalisée. Le PMAR est une mesure du FSE mise en œuvre au titre du PI 10.I, dont l'objectif est de lutter contre le décrochage scolaire. Dans le cadre du PMAR, les élèves sont répartis dans des groupes spécifiques (comportant entre 8 et 15 élèves, mais de tailles différentes dans des circonstances particulières) afin d'étudier la linguistique, les sciences sociales, les sciences et les mathématiques ainsi que les langues étrangères (les autres matières étant enseignées dans le groupe principal). La mesure a été mise en œuvre lors de l'année scolaire 2016-2017, en deuxième et troisième année d'enseignement secondaire (ESO). Les élèves ayant redoublé au moins une année à quelque moment que ce soit et les élèves qui, après avoir terminé la première année d'ESO, ne sont pas en mesure de passer à l'année suivante, sont tous admissibles pour participer. La sélection finale est effectuée à la suite de l'évaluation individuelle des élèves (sur des critères académiques et psycho-pédagogiques) réalisée par l'équipe enseignante.

Au total, 1 053 élèves ont participé, dont 512 en deuxième année d'ESO et 541 en la troisième année. Ces élèves ont été comparés à d'autres élèves présentant des difficultés d'apprentissage⁹⁸ afin de sélectionner un groupe témoin comparable, au regard des critères d'admissibilité, aux élèves traités (3 852 personnes au total).

Une approche par appariement a été adoptée, à l'aide des données administratives provenant de la base de données SAUCE (Sistema Informático para la Administración Unificada de Centros Educativos), fournie sous forme anonymisée par le Ministère de l'éducation. Ces données ont été utilisées pour calculer le score de propension⁹⁹ et pour mesurer la variable de résultat, qui correspond au passage à l'année supérieure. L'évaluation a constaté que les interventions avaient produit un effet positif et statistiquement significatif: le taux de réussite des élèves de deuxième année du groupe traité était supérieur de 18 points de pourcentage à celui du groupe témoin; pour les élèves de troisième année du groupe traité, la différence

⁹⁷ Voir [Díaz J.M. et al., 2019](#)

⁹⁸ Ces élèves ont été définis comme des élèves qui, lors de l'année scolaire 2015-2016, étaient en première et deuxième année d'enseignement secondaire, qui ont terminé l'année en ayant échoué à une ou plusieurs matières et qui ont redoublé au moins une année dans leur parcours scolaire.

⁹⁹ Les variables utilisées pour appairer les deux groupes étaient les suivantes: résultats scolaires au cours de l'année scolaire précédente, type d'école (privée ou publique), inscription à d'autres cours supplémentaires lors de l'année scolaire précédente (oui ou non), sexe, revenu moyen de la commune de résidence, lieu de résidence, pays de naissance (Espagne ou reste du monde).

correspond à environ 16 points de pourcentage. Dans les deux groupes, l'impact était plus positif pour les femmes que pour les hommes.

Interventions du FSE pour améliorer l'enseignement professionnel en Podlachie¹⁰⁰

L'étude a examiné les effets du projet intitulé «Bon métier - Bonne vie», qui avait pour objectif de promouvoir l'enseignement et la formation professionnels dans la région de Podlachie en Pologne. Le projet ciblait le personnel des établissements de premier cycle de l'enseignement secondaire, les élèves et les parents et un ensemble d'activités ont été mises en œuvre à destination des institutions et des personnes (stratégies de marketing et de communication adaptées pour les écoles, formations pour les professeurs, coopération avec les entreprises et les dirigeants de l'industrie locale, mentorat pour les élèves et les parents, conseils en matière de formation et d'éducation pour les élèves). Le projet a été mené entre 2017 et 2019.

La population traitée se composait d'élèves qui participaient au projet évalué. Sur 9 403 élèves, 1 500 ont été sélectionnés au moyen d'un échantillonnage stratifié aléatoire et 200 élèves ont participé à l'enquête de l'étude. Les informations relatives aux élèves qui ont participé au projet provenaient des données de surveillance. Le groupe témoin a été défini comme les élèves présentant des caractéristiques semblables à celles des élèves du groupe traité mais n'ayant pas participé au projet. Un groupe témoin se composant d'environ deux fois plus de membres que le groupe traité était prévu et a été sélectionné aléatoirement à partir du registre national de la population (PESEL, un registre attribuant un numéro individuel à chaque citoyen polonais). Cependant, en raison de difficultés causées par la COVID-19, il n'a pas été possible d'employer la méthode prévue. Finalement, la technique du recrutement en chaîne («boule de neige») a été utilisée comme principale méthode d'échantillonnage et a conduit à la réalisation de 401 entretiens, dont 384 ont été utilisés dans l'étude.

Les données relatives aux variables de résultat¹⁰¹ pour le groupe traité et le groupe témoin ont été collectées grâce à des enquêtes par questionnaire réalisées lors d'entretiens en personne, par téléphone ou par internet. L'étude a utilisé une pondération par le score de propension comme méthode contrefactuelle pour évaluer l'impact (seulement au niveau général, pas pour les sous-groupes spécifiques). Les analyses contrefactuelles menées lors de l'étude n'ont pas donné de constatations concluantes. La pertinence statistique des résultats n'était pas satisfaisante.

Évaluation du système de bourse de l'enseignement supérieur pour les étudiants issus de milieux défavorisés au Portugal¹⁰²

Une deuxième évaluation, pas encore publiée officiellement, examine un programme de bourses pour les étudiants issus de ménages à faibles revenus ayant pour objectif de favoriser leur accès à l'enseignement supérieur et d'augmenter leur niveau d'assiduité. Ce programme bénéficiait de l'appui du FSE (depuis la période 2007-2013) dans les régions Nord, Centre et Alentejo et de l'État dans les autres régions. Les étudiants admissibles sont ceux dont les ressources sont inférieures à un seuil spécifique (entre 6 800 et 7 900 EUR par personne) et qui ont obtenu un nombre minimal de crédits au cours de l'année d'étude précédente pour les étudiants qui ne sont pas en première année d'étude. Les mêmes règles d'admissibilité s'appliquent pour les universités publiques et privées. Le montant de la bourse est proportionné aux revenus du ménage.

Depuis 2011, le programme a couvert près de 70 000 étudiants par an, mais l'évaluation ne couvre que la période commençant en 2012 en raison de la disponibilité des données cette année-là. De plus, l'évaluation se concentre sur les étudiants demandant la bourse pour la première fois et inscrit en première année d'un cursus diplômant (licence ou master); l'admissibilité est déterminée par le seul critère des revenus¹⁰³.

Deux ensembles de données administratives ont été utilisés: des données sur les étudiants ayant demandé la bourse entre 2012 et 2018 [fournies par la direction générale de l'enseignement supérieur au Portugal (DGES)], qui ont été fusionnées au moyen d'un identifiant unique avec un autre ensemble de données

¹⁰⁰ Voir [Zub M. et al., 2020](#).

¹⁰¹ Les variables de résultat (sur la base des réponses à l'enquête) étaient les suivantes: fréquentation d'une école technique au moment de l'étude, fréquentation d'une école professionnelle du 1^{er} degré au moment de l'étude, fréquentation d'une école technique ou professionnelle du 1^{er} degré au moment de l'étude, candidature à une école secondaire, candidature à une école technique, candidature à une école professionnelle du 1^{er} degré, candidature à une école dispensant une formation professionnelle (école technique ou professionnelle du 1^{er} degré), le fait que l'école réponde dans une très large mesure aux attentes de l'élève, une profession tout à fait cohérente avec les intérêts de l'élève.

¹⁰² Voir Guthmuller S., Meroni E.C., pas encore publiée.

¹⁰³ L'analyse de l'échantillon complet des étudiants (pas uniquement de ceux en première année) est envisagée mais n'a pas encore été réalisée.

contenant des informations sur le parcours scolaire et la progression [fourni par la direction générale pour les statistiques et la science (DGEEC)]. L'échantillon final faisant l'objet de l'analyse se composait de 156 002 étudiants, dont 130 602 ont été traités et 25 400 ne l'ont pas été. Le seuil de revenu (la variable décisive) permettait à l'évaluateur d'appliquer une approche de la régression par discontinuité et plusieurs variables de résultat ont été mesurées: sur le court terme, le pourcentage d'étudiants encore inscrits à la fin de la première année et les crédits obtenus à la fin de la première année; sur le long terme, la probabilité d'obtenir un diplôme, la note finale obtenue et le nombre d'années nécessaires pour obtenir un diplôme.

Lors de la période initiale, les bourses ont produit des effets positifs concernant les taux d'inscription, mais pas les crédits obtenus; sur le long terme, les étudiants traités ont plus souvent obtenu un diplôme, en moins de temps et avec de meilleures notes que le groupe témoin. Les effets étaient plus élevés pour les hommes, les étudiants résidant dans les territoires couverts par le FSE et les étudiants inscrits en master et dans les universités d'État.

Nous avons déjà souligné l'utilité et la capacité explicative des méta-évaluations auparavant. Une hausse des EIC, et donc une hausse potentielle des méta-évaluations, permettrait de renforcer le débat sur les mesures spécifiques en faisant participer plusieurs AG et EM à ce débat en même temps. Comme nous l'avons indiqué plus haut, l'amélioration de la planification des EIC au niveau national et international devrait permettre de surmonter la complexité méthodologique de ces évaluations.

Augmenter le nombre de méta-évaluations

Dans les interventions du FSE, les EIC mesurent en général les effets à l'aide de variables quantitatives ou binaires (oui/non), comme les revenus ou le statut professionnel. Cependant, les «résultats difficilement quantifiables», portant sur la perception de soi et les capacités des participants, représentent des mesures importantes de la réussite des interventions. Par exemple, l'«employabilité», l'estime de soi ou les compétences acquises sont des conditions préalables au fait d'obtenir un emploi ou d'être actif sur le marché du travail. Dans de nombreux contextes, les analyses des résultats difficilement quantifiables seraient encore plus utiles que les chiffres de sortie du chômage et permettraient de faire la lumière sur la correspondance entre les interventions et les besoins des personnes.

Prendre en considération les «résultats difficilement quantifiables»

Les EIC peuvent être appliquées à ces résultats difficilement quantifiables, à condition de collecter les informations avant et après l'intervention auprès du groupe traité et du groupe témoin, conformément à la théorie du changement de l'intervention. Ces études doivent être planifiées à l'avance et faire participer tous les acteurs concernés, notamment les bénéficiaires qui se trouvent en contact direct et constant avec les participants. L'encadré 21 ci-dessous présente un exemple d'EIC évaluant les impacts produits sur les «résultats difficilement quantifiables».

Encadré 21 Exemple d'évaluation des effets sur les «résultats difficilement quantifiables» en Allemagne

Les programmes de création d'emploi améliorent-ils l'intégration sociale et le bien-être des chômeurs de longue durée?¹⁰⁴

L'étude examine les effets d'un programme de création d'emploi ciblant un groupe de personnes vulnérable, les chômeurs de longue durée qui ont bénéficié de l'aide sociale pendant au moins quatre ans et qui ont des problèmes de santé et/ou des enfants. Le programme a été opérationnel de 2015 à 2018 et a couvert environ 20 000 personnes. La mesure subventionnait des contrats de travail normaux, avec un temps de travail hebdomadaire de 30 heures, pour une durée maximale de 36 mois (principalement auprès

¹⁰⁴ Voir [Ivanov B. et al., 2020](#).

d'employeurs publics ou d'organisations caritatives).

Puisque la mesure pour l'emploi subventionnée, d'une durée maximale de 36 mois, avait pour objectif explicite d'encourager l'intégration sociale des groupes cibles, les auteurs ont essayé d'en évaluer l'impact à cet égard.

L'analyse a utilisé un large ensemble de données, intégrant des données administratives (les «biographies intégrées sur l'emploi» allemandes) à une enquête sur un panel de participants au programme et de personnes formant le groupe témoin. L'enquête sur panel a permis de collecter des informations sur des mesures subjectives relatives à la qualité de vie: la satisfaction à l'égard de la vie, la santé mentale, l'intégration sociale et le statut social. Des échelles de Likert ont été utilisées pour collecter des informations sur ces mesures, d'une manière cohérente avec d'autres enquêtes (comme l'enquête nationale PASS) afin de produire des constatations comparables. Les principales étapes de l'enquête étaient les suivantes: les personnes traitées et non traitées ont été définies et appariées dans les données administratives; le groupe traité et le groupe témoin qui en ont résulté ont passé des entretiens (trois vagues); l'ensemble de données final a été constitué, par nettoyage des données et élimination des cas avec des données manquantes. L'échantillon se composait de 2 531 paires appariées lors de la phase 1; de 1 191 paires lors de la phase 2 et de 450 paires lors de la phase 3 (pour un échantillon total d'environ 62 000 personnes, dont 12 400 appartenaient au groupe traité et 49 600 au groupe témoin).

En s'appuyant sur l'appariement par score de propension et en mesurant les effets à 7, 18 et 29 mois après l'entrée dans le programme, les auteurs ont conclu que les interventions produisaient des effets positifs sur les mesures relatives au bien-être, mais à des degrés différents: par exemple, la satisfaction à l'égard de la vie a connu une hausse considérable, mais le statut social n'a que modérément progressé. Cependant, les effets ont tendance à baisser tout au long du programme, ce qui pourrait s'expliquer par une hausse du nombre de participants quittant le programme ainsi que du nombre de personnes appartenant au groupe témoin qui trouvent un emploi. Pour les personnes les plus vulnérables, les effets positifs étaient plus élevés.

Comme nous l'avons souligné au début du présent guide, l'EIC représente un instrument analytique puissant pour évaluer «quel» effet a produit l'intervention du FSE+. Pour comprendre «comment» ou «pourquoi» les effets mesurés se sont produits, il est nécessaire d'utiliser d'autres instruments.

EIC intégrées aux EFT

Dans certains cas, lorsque les interventions sont bien connues et ont déjà été évaluées en profondeur, une analyse relativement limitée de la mise en œuvre ainsi que quelques entretiens avec les bénéficiaires suffisent à déterminer les principaux éléments à l'origine des effets mesurés. Pour des interventions plus complexes, moins connues ou innovantes, il est souvent nécessaire d'intégrer l'EIC à d'autres évaluations.

Une évaluation de mise en œuvre ou de processus, réalisée avant l'EIC, peut permettre de souligner la manière dont l'intervention s'est déroulée, les problèmes rencontrés lors de l'exécution et la manière dont les différents acteurs ont collaboré afin d'en assurer le succès. Les constatations de cette évaluation peuvent alimenter l'EIC en indiquant les effets sur lesquels il faudrait enquêter, la manière dont le groupe témoin devrait être composé eu égard aux caractéristiques de participation et le moment auquel il convient de réaliser l'EIC.

Il est possible de réaliser une évaluation fondée sur la théorie (EFT) en parallèle de l'EIC. Les deux évaluations peuvent s'alimenter l'une l'autre: l'EFT fournit des explications concernant les mécanismes contextuels, sociaux et individuels qui ont une influence sur les effets et les rendent possibles, tandis que l'EIC calcule l'effet net et précise la mesure dans laquelle l'intervention a produit des résultats significatifs. Dans une évaluation d'impact, il est recommandé d'intégrer des d'EIC et des EFT afin de fournir un aperçu complet des effets et de leurs causes.

Glossaires

Acronymes

PAMT	Politique active du marché du travail
CBA	Analyse de rentabilité (Cost-benefit analysis)
EIC	Évaluation d'impact contrefactuelle
RPDC	Règlement portant dispositions communes
DG EMPL	Direction générale de l'emploi, des affaires sociales et de l'inclusion
DG REGIO EMPL	Direction générale de la politique régionale
DD	Double différence
CE	Commission européenne
PE	Plan d'évaluation
FEDER	Fonds européen de développement régional
Fonds structurels	Fonds social européen
UE	Union européenne
RGPD	Règlement (UE) 2016/679 (règlement général sur la protection des données)
OI	Organisme(s) intermédiaire(s)
BIE	Biographies intégrées sur l'emploi
VI	Variable instrumentale
JLD	Ensemble de données longitudinales sur les demandeurs d'emploi
EFT	Enquête sur les forces de travail
AG	Autorité(s) de gestion
EM	État(s) membre(s)
ONG	Organisations non gouvernementales
PO	Programme(s) opérationnel(s)
PSE	Service(s) public(s) de l'emploi UE
PSM	Appariement par score de propension (propensity score matching)
RCT	Essai aléatoire contrôlé (randomised control trial)
RDD	Approche de la régression par discontinuité (regression discontinuity design)
PME	Petites et moyennes entreprises
IEJ	Initiative pour l'emploi des jeunes

Définitions

Signification	Définition
Données de référence	Données sur les variables mesurées avant qu'une unité (particulier ou entreprise) soit soumise à l'intervention. Dans de nombreux cas, les mesures des résultats de l'intervention sont relevées avant le traitement dans les groupes traités et témoins.
Bénéficiaire	Selon l'article 2, point a), du règlement 2021/1060 du Conseil, un bénéficiaire est «un organisme de droit public ou privé, ou une entité avec ou sans personnalité juridique ou une personne physique, responsable du lancement ou à la fois du lancement et de la mise en œuvre des opérations». Dans le contexte des instruments financiers (parfois utilisé pour les micro-crédits aux indépendants dans le cadre du FSE+, par exemple), un bénéficiaire est l'organisme qui met en œuvre le fonds à participation ou, lorsqu'il n'existe pas de structure de fonds à participation, l'organisme qui met en œuvre le fonds spécifique ou, lorsque l'autorité de gestion gère l'instrument financier, l'autorité de gestion.
Groupe témoin	Un groupe de personnes, d'entreprises ou d'autres unités le plus semblable possible au groupe traité ou groupe de traitement, mais qui ne bénéficie pas du traitement, et qui sert à estimer les résultats contrefactuels. À proprement parler, le terme «groupe témoin» fait référence aux approches expérimentales (voir «Randomisation» ci-dessous) et le terme «groupe de comparaison» fait référence aux approches quasi expérimentales, mais ces deux termes sont utilisés de manière interchangeable en pratique.
Analyse contrefactuelle	Une comparaison entre ce qui s'est réellement passé et ce qui se serait passé en l'absence de l'intervention, en ce qui concerne les résultats. Puisque la différence entre les résultats réels et contrefactuels définit l'effet de causalité de l'intervention, l'analyse contrefactuelle englobe toutes les approches ayant pour objectif d'évaluer la part des changements observés qui peut être attribuée à l'intervention évaluée.
Double différence	Dans sa forme la plus simple, la différence d'un résultat avant et après le traitement dans un groupe témoin est déduite de cette même différence observée au sein du groupe traité afin d'obtenir une estimation de l'impact d'une intervention. Les impacts calculés grâce à la double différence sont généralement déduits à l'aide d'un cadre de régression.
Efficacité	Elle fait référence à la «réalisation des objectifs» et est évaluée en comparant ce qui a été accompli avec ce qui était prévu (ou avec une situation de référence) ou en comparant les observations après une action avec ce qui se serait passé sans l'action (situation contrefactuelle).
Efficience	Terme qui se définit par la réalisation d'un résultat donné à un coût minimal ou, de la même manière, par l'obtention d'un maximum de résultats avec un niveau déterminé de ressources. L'efficience est déterminée à travers une analyse de rentabilité ou d'efficacité.
Plan d'évaluation	Conformément à l'article 44 du règlement 2021/1060 du Conseil (RPDC), l'État membre ou l'autorité de gestion élabore un plan d'évaluation, qui peut porter sur plusieurs programmes et qui doit être transmis au comité de suivi au plus tard un an après la décision portant approbation du programme.
Évaluation externe	Une évaluation réalisée par un évaluateur externe indépendant à la suite d'une procédure d'appel d'offres.
Impact	Dans le cadre des EIC, les impacts font référence à l'effet net, définis comme la différence entre les résultats moyens du traitement et les résultats contrefactuels. Au sens du présent guide, les termes «impact» et «effet net» sont interchangeables.
Évaluation d'impact contrefactuelle	Un type d'évaluation d'impact qui cherche à isoler les effets causals des interventions à travers une estimation des résultats contrefactuels moyens et en soustrayant ces résultats des résultats moyens observés parmi les unités traitées. Les estimations des résultats contrefactuels sont généralement déduites à partir de groupes témoins soigneusement sélectionnés pour être le plus ressemblant possible au groupe traité.
Approche par variables instrumentales	Lors de l'application de cette méthode, la sélection dans le traitement doit être au moins en partie déterminée par un facteur (ou un instrument) exogène sans autre rapport avec les résultats que le traitement. Ainsi, le facteur exogène influence la participation, mais

Signification	Définition
	pas les résultats directement.
Évaluation interne	Une évaluation réalisée en interne, c'est-à-dire directement commandée par une institution publique ou une unité indépendante (de l'AG ou de l'OI) sans procédure d'appel d'offres, ou sous la forme d'une procédure étendue de suivi et d'analyse.
Interventions	Terme qui fait généralement référence aux actions des programmes opérationnels du FSE ou aux projets cofinancés par le FSE.
Appariement	Il s'agit d'une méthode dans laquelle l'intervention et les échantillons de contrôle sont assortis en fonction de leurs caractéristiques observées.
Approche non aléatoire ou quasi expérimentale	Une approche de l'évaluation d'impact contrefactuelle selon laquelle les groupes témoins sont formés à l'aide d'autres méthodes que la répartition aléatoire (ou randomisation).
Résultat	Les effets à court et moyen termes probables ou obtenus des résultats d'une intervention ¹⁰⁵ . Le terme «résultat» est semblable au terme «produits» mais il est plus souvent employé dans les évaluations d'impact.
Réalisation	Une réalisation correspond à tout ce qui est directement lié à la mise en œuvre d'une opération financée par le FSE. Les réalisations se mesurent en unités physiques ou monétaires. Les réalisations sont principalement mesurées en matière de nombre de personnes aidées, d'entités aidées, de biens et de services fournis et de projets mis en œuvre.
Participants	Un participant est une personne physique bénéficiant directement d'une opération, sans être responsable du lancement ou de la mise en œuvre de cette opération. Dans le contexte du FSE, ce terme fait référence aux personnes à qui les interventions du FSE apportent un appui.
Évaluation des processus	L'évaluation des processus concerne la mise en œuvre des programmes, notamment mais pas exclusivement la manière dont les services sont proposés, les différences entre la population visée et la population touchée par le programme, l'accès au programme et les pratiques de gestion.
Appariement par score de propension	L'appariement par score de propension consiste à estimer un modèle statistique pour l'échantillon entier (personnes traitées et candidats potentiels) qui fournit une estimation de la propension à participer de chaque personne ou entreprise - indépendamment du fait qu'elle participe ou non. Les personnes ou entreprises traitées sont ensuite appariées, soit avec une personne ou entreprise non traitée, soit avec plusieurs personnes ou entreprises non traitées, en fonction du score de propension.
Randomisation (ou répartition aléatoire)	Il s'agit d'une méthode selon laquelle les membres d'un groupe cible sont désignés de manière aléatoire pour une série de traitements ou de situations de contrôle. La randomisation garantit une équivalence statistique des groupes à tous les égards au moment de la répartition.
Approche de la régression par discontinuité	Cette méthode peut être appliquée lorsque l'accès à une intervention est fixé par un point limite le long d'une note ou d'une échelle ou d'une mesure continue. Cette approche repose sur le fait que ceux qui se situent à proximité immédiate du point limite seront très semblables, hormis le fait que les candidats du dessus bénéficient de l'intervention, à l'inverse de ceux d'en dessous. La comparaison des résultats des candidats au-dessus et en dessous du point de limite fournit une estimation de l'impact de l'intervention.
Pertinence	Terme qui fait référence à l'adéquation des objectifs explicites d'une intervention avec les problèmes socio-économiques que l'intervention est supposée résoudre ¹⁰⁶ .
Résultat	Les effets sur les participants ou les entités produits par une opération, par exemple en ce qui concerne leur situation professionnelle, leurs revenus, leur score à des tests scolaires standardisés, etc. Les effets peuvent être mesurés à court terme ou à long terme. Dans le contexte du FSE+, lorsque l'on mesure les indicateurs de résultats, on entend par court terme immédiatement après la participation (4 semaines) et par long

¹⁰⁵ Voir [OCDE, 2010](#).

¹⁰⁶ Voir [Commission européenne, 2013](#).

Signification	Définition
Groupe de traitement	terme au moins six mois après ¹⁰⁷ . Cependant, dans le cadre d'une analyse d'impact, on entend souvent par long terme une période plus longue, d'au moins 24 mois après les interventions. Un groupe de personnes, d'entreprises ou d'autres unités qui bénéficient d'une intervention ou y sont soumises (que ce soit l'offre de traitement ou sa réception effective).

¹⁰⁷ Voir [Commission européenne, 2021a](#) – et [Commission européenne, 2018](#).

Bibliographie

- Ashenfelter, O. (1978). «Estimating the effect of training programmes on earnings», *Review of Economics and Statistics*, 6, pages 47-57. <https://www.jstor.org/stable/1924332?origin=crossref>
- Axdorph E, Egebark J., Lundström T., Özcan G. (2019) *Effekter av förstärkta förmedlingsinsatser för unga arbetssökande – Resultat från utvärderingen av Ung framtid*. Arbetsförmedlingen
- Baran J. et al. (2016). *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój. I Raport Wskaźnikowy*. <http://files.evaluationhelpdesk.eu/Evaluations/PLE49.pdf>
- Baran J. et al. (2017). *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój II Raport wskaźnikowy*. <http://files.evaluationhelpdesk.eu/Evaluations/PLE196.pdf>
- Baran J. et al. (2018). *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój III Raport wskaźnikowy*. <http://files.evaluationhelpdesk.eu/Evaluations/PLE197.pdf>
- Baran J. et al. (2018a). *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój. II Raport Tematyczny*. <http://files.evaluationhelpdesk.eu/Evaluations/PLE221.pdf>
- Bazzoli M., De Poli S., Rettore E., Schizzerotto A.(2018). «Are Vocational Training Programmes Worth Their Cost? Evidence from a Cost-Benefit Analysis». In *Politica Economica/Journal of Economic Policy* Vol. XXXIV(3), pages. 215-240. <https://www.rivisteweb.it/doi/10.1429/92119>
- Bloom, H. S. (1995). «Minimum detectable effects: A simple way to report the statistical power of experimental designs», *Evaluation Review*, 8(2), 225-246. <https://cpb-us-e2.wpmucdn.com/sites.uci.edu/dist/1/1159/files/2021/03/Bloom-MDES-Eval-Rev-1995-Bloom.pdf>
- Bloom, H. S. (2009). «Modern regression discontinuity analysis», *MDRC Working Papers on Research Methodology*, New York: MDRC. <https://www.mdrc.org/publication/modern-regression-discontinuity-analysis>
- Boockmann B. et al. (2017). *Evaluation des ESF-Bundesprogramms zur Eingliederung langzeitarbeitsloser Leistungsberechtigter nach dem SGB II auf dem allgemeinen Arbeitsmarkt. Im Auftrag des Bundesministeriums für Arbeit und Soziales*. <http://files.evaluationhelpdesk.eu/Evaluations/DEE189.pdf>
- Boockmann B. et al. (2018). *Evaluation des ESF-Bundesprogramms zur Eingliederung langzeitarbeitsloser Leistungsberechtigter nach dem SGB II auf dem allgemeinen Arbeitsmarkt. Im Auftrag des Bundesministeriums für Arbeit und Soziales*. <http://files.evaluationhelpdesk.eu/Evaluations/DEE52.pdf>
- Boockmann B. et al. (2019). *Evaluation des ESF-Bundesprogramms zur Eingliederung langzeitarbeitsloser Leistungsberechtigter nach dem SGB II auf dem allgemeinen Arbeitsmarkt. Im Auftrag des Bundesministeriums für Arbeit und Soziales*. <http://files.evaluationhelpdesk.eu/Evaluations/DEE94.pdf>
- Boockmann B. et al. (2021). *Evaluation des ESF-Bundesprogramms zur Eingliederung langzeitarbeitsloser Leistungsberechtigter nach dem SGB II auf dem allgemeinen Arbeitsmarkt. Im Auftrag des Bundesministeriums für Arbeit und Soziales*. <http://files.evaluationhelpdesk.eu/Evaluations/DEE202.pdf>
- Borik V. et al. (2015). *The net effects of graduate work experience and the promotion of self-employment. Technical Report*. <http://files.evaluationhelpdesk.eu/Evaluations/SKE8.pdf>

- Bratti M. et al. (2018). *Vocational Training for Unemployed Youth in Latvia: Evidence from a Regression Discontinuity Design*. IZA DP N° 11870. <http://ftp.iza.org/dp11870.pdf>
- Bratu C. et al. (2014). *Knowledge gaps in evaluating labour market and social inclusion policies identified 39 ESF counterfactual impact evaluations in the 2007-2013 programming period*. doi:10.2788/083390. <https://op.europa.eu/en/publication-detail/-/publication/d0e73217-2a40-47af-a5c3-a700dcd47da3/language-en>
- Bredgaard, T. (2015). «Evaluating what Works for whom in active Labour market policies». *European Journal of Social Security*, 17(4), 436-452. <https://doi.org/10.1177/138826271501700403>
- Bryson, A., Dorsett, R. and Purdon, S. (2002). *The use of propensity score matching in the evaluation of active labour market policies*, Department for Work and Pensions, document de travail n° 4. [http://eprints.lse.ac.uk/4993/1/The use of propensity score matching in the evaluation of active labour market policies.pdf](http://eprints.lse.ac.uk/4993/1/The%20use%20of%20propensity%20score%20matching%20in%20the%20evaluation%20of%20active%20labour%20market%20policies.pdf)
- Caliendo M., Kopeinig S., (2008), «Some practical Guidance for the implementation of propensity score matching». *Journal of Economic Surveys*; Volume 22, Numéro 1 - Pages 31-72. <https://doi.org/10.1111/j.1467-6419.2007.00527.x>
- Caliendo, M., Mahlstedt, R., & Mitnik, O. A. (2017). «Unobservable, but unimportant? The relevance of usually unobserved variables for the evaluation of labour market policies». *Labour Economics*, 46, 14-25. <https://doi.org/10.1016/j.labeco.2017.02.001>
- Card, D., Ibarra, P. and Villa, J. M. (2011). *Building in an evaluation component for active labour market programs: a practitioner's guide*, document de réflexion n° 6085, Bonn, Allemagne: IZA. <http://ftp.iza.org/dp6085.pdf>
- Card D., Kluve J., Weber A. (2017). «What works? A meta-analysis of recent active labor market program evaluations». *Journal of the European Economic Association*, Volume 16, Numéro 3, juin 2018, Pages 894–931, <https://doi.org/10.1093/jeea/ivx028>
- Centre de prévention et de contrôle des maladies CDC (2013). *Good Evaluation Questions: Checklist to Help Focus Your Evaluation*. Département de la Santé et des services aux personnes - États-Unis. http://www.cdc.gov/asthma/program_eval/AssessingEvaluationQuestionChecklist.pdf
- Danmarks Statistik (2017). *Effektmåling og monitorering 2016 - strukturfondsindsatsen i Syddanmark*. <http://files.evaluationhelpdesk.eu/Evaluations/DKE7.pdf>
- Danmarks Statistik (2018). *Effektmåling af den Virksomhedsrettede Strukturfondsindsats 2007-2013*. <http://files.evaluationhelpdesk.eu/Evaluations/DKE10.pdf>
- Diaz J. M. et al. (2019). *Evaluación del PO-FSE 2014/2020 del Principado de Asturias para el informe anual a presentar en 2019*. <http://files.evaluationhelpdesk.eu/Evaluations/ESE101.pdf>
- Ecorys, Ismeri Europa (2020). *Study supporting the evaluation of ESF support to education and training (Thematic Objective 10)*. Étude réalisée pour le compte de la Direction générale de l'emploi, des affaires sociales et de l'inclusion. <https://op.europa.eu/en/publication-detail/-/publication/d0c1a558-077d-11eb-a511-01aa75ed71a1/language-en/format-PDF/source-173162502>
- Commission européenne (2007): Orientations indicatives sur les méthodes d'évaluation Évaluation pendant la période de programmation. Document de travail n° 5. DG Politique régionale. https://ec.europa.eu/regional_policy/sources/docoffic/2007/working/wd5_ongoing_fr.pdf
- Commission européenne (2013): EVALSED: La ressource pour l'évaluation du développement socio-économique, version mise à jour.

https://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/guide_evalsed.pdf

Commission européenne (2017): Boîte à outils pour une meilleure réglementation, Outil #47 sur les *Critères et questions d'évaluation* https://ec.europa.eu/info/sites/default/files/file_import/better-regulation-toolbox-47_en_0.pdf

Commission européenne (2018): Suivi et évaluation du Fonds social européen de la politique européenne de cohésion. Document d'orientation
<https://ec.europa.eu/sfc/en/system/files/ged/ESF%20monitoring%20and%20evaluation%20guidance.pdf>

Commission européenne (2019): *Advanced counterfactual evaluation methods*. Document d'orientation. doi:10.2767/464242.
<https://ec.europa.eu/social/main.jsp?catId=738&langId=en&pubId=8254&furtherPubs=yes>

Commission européenne (2020): *Counterfactual impact evaluation of European Social Fund interventions in practice*. Document d'orientation pour les autorités de gestion. doi:10.2767/55495. <https://op.europa.eu/en/publication-detail/-/publication/f82c5fb8-bb40-11ea-811c-01aa75ed71a1/language-en>

Commission européenne (2020): *How to use administrative data for European Social Funds counterfactual impact evaluations. A step-by-step guide for managing authorities*.
<https://op.europa.eu/en/publication-detail/-/publication/d96feed3-f30c-11ea-991b-01aa75ed71a1>

Commission européenne (2021): Règlement (UE) 2021/1060 du Parlement européen et du Conseil du 24 juin 2021 portant dispositions communes relatives au Fonds européen de développement régional, au Fonds social européen plus, au Fonds de cohésion, au Fonds pour une transition juste et au Fonds européen pour les affaires maritimes, la pêche et l'aquaculture, et établissant les règles financières applicables à ces Fonds et au Fonds «Asile, migration et intégration», au Fonds pour la sécurité intérieure et à l'instrument de soutien financier à la gestion des frontières et à la politique des visas.

Commission européenne (2021a): *Common Indicators Toolbox*. Document de travail. Juin 2021.
https://ec.europa.eu/sfc/en/system/files/2021/ged/toolbox_june_2021_final.pdf

Fondazione G. Brodolini, Metis GmbH, Applica, Ockham IPS (2020). *Study for the Evaluation of ESF Support to Employment and Labour Mobility (Thematic Objective 8)*, étude réalisée pour le compte de la Commission européenne, Direction générale de l'emploi, des affaires sociales et de l'inclusion. <https://ec.europa.eu/social/BlobServlet?docId=22899&langId=en>

Frolich, M. (2004). «Programme evaluation with multiple treatments», *Journal of Economic Surveys*, 18(2), pages 181-224.
<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.620.9209&rep=rep1&type=pdf>

Guthmuller S., Meroni E.C., *Evaluation of the higher education grant system for less privileged students in Portugal*. Rapport technique du JRC. Non encore publié.

Hagglund, P. (2006). *A description of three randomised experiments in Swedish labour market policy*, Institute for Labour Market Policy Evaluation, rapport 2006:4.
<https://www.ifau.se/globalassets/pdf/se/2006/r06-04.pdf>

Heckman, J. J., Ichimura, H., Smith, J. et Todd, P. (1998). «Characterizing selection bias using experimental data». In *Econometrica*, Vol. 66, N° 5. P. 1017-1098.
http://jenni.uchicago.edu/papers/Heckman_Ichimura_etal_1998_Econometrica_v66_n5_r.pdf

Hunger K. and Sattler K. (2017). *Evaluationsbericht zum spezifischen Ziel A 1.1. im Rahmen der Evaluierung des Operationellen Programms des Europäischen Sozialfonds in Baden-Württemberg 2014–2020*. <http://files.evaluationhelpdesk.eu/Evaluations/DEE26.pdf>

Holland, P. W. (1986). «Statistics and Causal Inference», *Journal of the American Statistical Association*, 81 (396), 945-960 DOI: 10.1080/01621459.1986.10478354.

<http://people.umass.edu/~stanek/pdf/causal-holland.pdf>

IAW Institut für Angewandte Wirtschaftsforschung _ ISG Institut für Sozialforschung und Gesellschaftspolitik, GmbH (2015). *Evaluation der Modellprojekte "Bürgerarbeit". Endbericht.* <http://files.evaluationhelpdesk.eu/Evaluations/DEE7.pdf>

ICF, Cambridge Econometrics and Eurocentre (2020). *Study supporting the 2020 evaluation of promoting social inclusion, combating poverty and any discrimination by the European Social Fund (Thematic Objective 9).* Étude réalisée pour le compte de la Commission européenne, Direction générale de l'emploi, des affaires sociales et de l'inclusion. <https://op.europa.eu/en/publication-detail/-/publication/8788ec85-2308-11eb-b57e-01aa75ed71a1>

Indecon (2016). *Indecon Evaluation of JobBridge Activation Programme.* <http://files.evaluationhelpdesk.eu/Evaluations/IEE2.pdf>

Instytut Badań Strukturalnych, Imapp, IQS. (2015). *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój. I Raport Tematyczny.* <http://files.evaluationhelpdesk.eu/Evaluations/PLE34.pdf>

Ires Piemonte (2019). *Misure di sostegno all'imprenditorialità.* <http://files.evaluationhelpdesk.eu/Evaluations/ITE61.pdf>

Isfol (2016). *Primo rapporto di valutazione del piano italiano Garanzia giovani.* <http://files.evaluationhelpdesk.eu/Evaluations/ITE13.pdf>

Ismeri Europa (2018). *I Rapporto tematico di valutazione – I risultati di PIPOL.* <http://files.evaluationhelpdesk.eu/Evaluations/ITE36.pdf>

Ismeri Europa – Ecorys – Institute for Employment Studies. (2019). *Pilot and feasibility study on the sustainability and effectiveness of results for European Social Fund participants using counterfactual impact evaluations.* doi:10.2767/39339. <https://op.europa.eu/fr/publication-detail/-/publication/84cc9eb9-b33d-11e9-9d01-01aa75ed71a1>

Ivanov B. et al., 2020, «Do Job creation schemes improve the social integration and well-being of the long-term unemployed?», *Labour Economics*. <https://doi.org/10.1016/j.labeco.2020.101836>.

Kalinowski H. (2020). *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój V Raport wskaźnikowy.* <http://files.evaluationhelpdesk.eu/Evaluations/PLE296.pdf>

Kalinowski H. et al. (2020). *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój Raport końcowy.* <http://files.evaluationhelpdesk.eu/Evaluations/PLE339.pdf>

Kluve, J., Lehmann, H., & Schmidt, C. M. (2008). «Disentangling treatment effects of Active Labor Market Policies: The role of labor force status sequences». *Labour Economics*, 15(6), 1270-1295. <https://doi.org/10.1016/j.labeco.2007.12.002>

Krug, G., et Stephan, G. (2011). *Is contracting-out intensified placement services more effective than in-house production? Evidence from a randomized field experiment*, Documents de réflexion LASER - Document N° 5, <http://doku.iab.de/externe/2011/k110912303.pdf>

Lammers M. et Kok L.(2021). «Are active labor market policies (cost-)effective in the long run? Evidence from the Netherlands». In *Empirical Economics*, 60:1719–1746. <https://doi.org/10.1007/s00181-019-01812-3>

Lang, J. (2021). «Employment effects of language training for unemployed immigrants». *J Popul Econ* (2021). <https://doi.org/10.1007/s00148-021-00832-7>

Martini, A. (2009). *Counterfactual impact evaluation: what it can (and cannot) do for cohesion policy*, document préparé à l'occasion de la 6e conférence sur l'évaluation de la politique de

cohésion, Varsovie, 30 novembre 2009.

https://ec.europa.eu/regional_policy/archive/conferences/evaluation2009/abstracts/martini.doc

Morgan, S. L. and Winship, C. (2014). *Counterfactual and causal inference: Methods and principles for social research*. 2^e édition. Cambridge and New York: Cambridge University Press.

OCDE (2010), Glossaire des principaux termes relatifs à l'évaluation et la gestion axée sur les résultats, Paris, <https://www.oecd.org/dac/evaluation/2754804.pdf>

Openfield (2019). *Analiza skuteczności i efektywności dotacji na założenie działalności gospodarczej udzielonych w ramach 8 osi priorytetowej RPO WM - komponent 3*. <http://files.evaluationhelpdesk.eu/Evaluations/PLE312.pdf>

Palczyńska M. et al. (2019). *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój IV Raport wskaźnikowy*. <http://files.evaluationhelpdesk.eu/Evaluations/PLE274.pdf>

Pomatto G. (2017). *L'attuazione del Buono per Servizi al Lavoro nella Regione Piemonte*. Ires Piemonte. <http://files.evaluationhelpdesk.eu/evaluations/ITE55.pdf>

Pomatto G. (2019). *Buoni per servizi al lavoro nella Regione Piemonte: qualità percepita dai destinatari e meccanismi dell'attuazione*. Ires Piemonte. <http://files.evaluationhelpdesk.eu/evaluations/ITE77.pdf>

Pompili M., Giorgetti I. (2020). *Rapporto di Placement. Servizio di attività di valutazione del POR FSE 2014/2020 – Regione Marche*. <http://files.evaluationhelpdesk.eu/Evaluations/ITE115.pdf>

Pompili M., Giorgetti I. (2020a). *Rapporto tematico "Disoccupazione di lunga durata". Servizio di attività di valutazione del POR FSE 2014/2020 – Regione Marche*. <http://files.evaluationhelpdesk.eu/Evaluations/ITE214.pdf>

Poy S. (2019). *Gli effetti occupazionali del buono per servizi al lavoro nella Regione Piemonte: prime evidenze. Misura per disoccupati da almeno 6 mesi, anno 2017*. Ires Piemonte. <http://files.evaluationhelpdesk.eu/evaluations/ITE123.pdf>

Poy S (2020). *Nuove evidenze sull'effetto occupazionale del buono per servizi al lavoro. Target persone disoccupate da almeno 6 mesi*. Ires Piemonte. <http://files.evaluationhelpdesk.eu/evaluations/ITE245.pdf>

Re-source Pracownia Badań i Doradztwa (2015). *Ocena działań na rzecz wydłużenia aktywności zawodowej osób starszych w kontekście niekorzystnej sytuacji demograficznej w województwie lubelskim*. <http://files.evaluationhelpdesk.eu/evaluations/PLE126.pdf>

Riccio J., Friedlander, D., Freedman S. (1994). *GAIN: Benefits, Costs, and Three-Year Impacts of a Welfare-to-Work Program*, MDRC, NYC <https://www.mdrc.org/publication/gain-benefits-costs-and-three-year-impacts-welfare-work-program>

Scheller F. and Seidel K. (2020). *Zweiter Evaluationsbericht zum spezifischen Ziel A 1.1: Teilnehmer*innenperspektive und Wirkungsanalyse. im Rahmen der Evaluierung des Operationellen Programms des Europäischen Sozialfonds in Baden-Württemberg 2014–2020*. <http://files.evaluationhelpdesk.eu/Evaluations/DEE171.pdf>

Shadish, W. R., Cook, T. D. and Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalised causal inference*, Boston, US: Houghton Mifflin Company.

Stern E. et al. (2012). *Broadening the range of designs and methods for impact evaluations*, Department for International Development of the UK, Working Paper 38. <https://assets.publishing.service.gov.uk/media/57a08a6740f0b6497400059e/DFIDWorkingPaper38.pdf>

WK Kellogg Foundation (2004). *Logic Model Development Guide*. <https://www.wkcf.org/resource-directory/resources/2004/01/logic-model-development-guide>

Zub M. et al. (2020). *Ocena wpływu wsparcia RPOWP na popularyzację szkolnictwa zawodowego w województwie podlaskim. Raport końcowy.*
<http://files.evaluationhelpdesk.eu/Evaluations/PLE359.pdf>

Liste des annexes

Annexe 1. Lectures complémentaires

La liste suivante comprend des lectures conseillées pour le personnel des autorités de gestion qui souhaite approfondir les questions évoquées dans le présent guide. La documentation sur l'évaluation est très fournie. Cette liste signale les analyses majeures dignes de foi qui fournissent des informations directement utiles à la planification des EIC. Pour la plupart des sources, une courte description suit directement la référence.

L'évaluation en général

- Rossi, Peter H., Mark W. Lipsey, et Gary T. Henry. (2018). *Evaluation: A Systematic Approach*. 8^e édition. Thousand Oaks, CA: Publications SAGE.

Un ouvrage classique sur les pratiques et méthodes d'évaluation. Inclut des méthodes et des exemples.

- Gertler, Paul J.; Martinez, Sebastian; Premand, Patrick; Rawlings, Laura B.; Vermeersch, Christel M. J. (2016). *L'évaluation d'impact en pratique*, Second Edition. Washington, DC: Banque interaméricaine de développement et Banque Mondiale. © Banque mondiale. <https://openknowledge.worldbank.org/handle/10986/25030> (disponible en anglais, en portugais et en espagnol).

À l'instar du présent guide, ce manuel aborde d'abord l'évaluation classique (essai aléatoire contrôlé) et ensuite les alternatives. Bien que destiné aux gestionnaires de programmes dans les pays à faible revenu, le manuel présente une pertinence directement applicable dans les États membres de l'UE.

- White, H., & Raitzer, D. A. (2017). *Impact evaluation of development interventions: A practical guide*. Banque asiatique de développement. <https://www.adb.org/sites/default/files/publication/392376/impact-evaluation-development-interventions-guide.pdf>

À l'instar de Gertler et al. (2016) mentionné ci-dessus, ce guide, qui traite de l'évaluation d'impact des interventions dans les pays à revenu faible et intermédiaire, présente une discussion complète concernant les méthodes générales d'EIC. Il est plus technique que Gertler et al. (2016), mais contient cependant de nombreuses réflexions pratiques qui pourraient intéresser les gestionnaires du programme FSE.

- Commission européenne (2019) *Advanced counterfactual evaluation methods*. Document d'orientation. doi:10.2767/464242. <https://op.europa.eu/en/publication-detail/-/publication/11968bbb-fac9-11e9-8c1f-01aa75ed71a1/language-en>

Ce document présente des méthodes d'évaluation d'impact contrefactuelle récentes et plus avancées, comme l'analyse séquentielle, l'appariement dynamique et les contrôles de synthèse.

- Csillag Marton, Kreko Judi et Scharle Agota. (2020). *Counterfactual evaluation of youth employment policies*. Préparé dans le cadre du projet «Youth Employment PartnerSHIP». <http://yepartnership.ibs.org.pl/content/uploads/2021/02/Methodological-guide.pdf> [Disponible en anglais, en espagnol, en hongrois (traduction en cours), en italien et en polonais]

Il s'agit d'une introduction «étape par étape» à l'évaluation contrefactuelle des politiques du marché du travail ciblant les jeunes, qui se concentre sur l'utilisation des données administratives. Les problèmes sont présentés en s'appuyant sur les problèmes pratiques rencontrés dans le cadre de quatre EIC portant sur des subventions à l'emploi de jeunes en

Espagne, en Hongrie, en Italie et en Pologne.

- Trésor britannique (Royaume-Uni) (2020). *The Magenta Book: Guidance for evaluation*. Londres: The Agency. <https://www.gov.uk/government/publications/the-magenta-book>

Le Magenta book apporte des détails sur la méthode d'évaluation. Ces documents sont des exemples intéressants de la perspective d'évaluation intra-gouvernementale.

Essais aléatoires contrôlés

- White H. (2013). «An introduction to the use of randomised control trials to evaluate development interventions», *Journal of Development Effectiveness*, 5:1, 30-49, DOI: 10.1080/19439342.2013.764652 <https://www.tandfonline.com/doi/pdf/10.1080/19439342.2013.764652>
- White, H., Sabarwal S. & T. de Hoop, (2014). Essais contrôlés randomisés (RCT), Notes méthodologiques: Évaluation d'impact 7, Centre de recherche de l'UNICEF, Florence. https://www.unicef-irc.org/publications/pdf/brief_7_randomized_controlled_trials_eng.pdf

Deux introductions non techniques à la logique des essais aléatoires contrôlés, avec une discussion portant sur différentes approches et une critique des essais aléatoires contrôlés.

- Glennerster, R., & Takavarasha, K., (2013). *Running Randomized Evaluations: A Practical Guide*, Princeton University Press, Princeton, NJ.

Ce livre fournit une orientation étape par étape sur la manière de concevoir et mettre en œuvre des essais aléatoires contrôlés dans le domaine des programmes sociaux. Il s'appuie sur les exemples concrets d'essais aléatoires contrôlés réalisés par Abdul Latif Jameel, Poverty Action Lab.

Double différence

- Card, David, Pablo Ibararán, et Juan Miguel Villa. (2011). *Building in an Evaluation Component for Active Labor Market Programs: A Practitioner's Guide*. Document de réflexion de l'IZA n° 6085. Bonn: Forschungsinstitut zur Zukunft der Arbeit. <http://ftp.iza.org/dp6085.pdf>

Compare l'approche par double différence et les essais aléatoires contrôlés.

- Lechner, M. (2011). «The Estimation of Causal Effects by Difference-in-Difference Methods». *Foundations and Trends in Econometrics*. Vol. 4, N° 3 (2010) 165–224 DOI: 10.1561/0800000014. https://michael-lechner.eu/ml_pdf/journals/2011_Lechner_DiD_2011_ECO%20403%20Lechner_darf%20au%20Netz.pdf

Cet article traite de l'approche par double différence en profondeur ainsi que de certains des principaux problèmes rencontrés lors de son application: Les extensions de la double différence, comme les applications non linéaires et l'appariement par score de propension avec la double différence, sont également présentées.

- Fredriksson A. and de Oliveira G. M. (2019). «Impact evaluation using Difference-in-Differences». *RAUSP Management Journal*, Vol. 54 N° 4, p. 519-532. DOI 10.1108/RAUSP-05-2019-0112 <https://www.emerald.com/insight/content/doi/10.1108/RAUSP-05-2019-0112/full/pdf?title=impact-evaluation-using-difference-in-differences>

Un aperçu des méthodes par double différence, avec des recommandations pratiques.

- Card, David et Alan B. Krueger. (1994). «Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania». *American Economic Review*, 84 (4), 774–775. <https://davidcard.berkeley.edu/papers/njmin-aer.pdf>

Un exemple classique d'application d'une technique par double différence.

Variables instrumentales

- Morgan, Stephen L., et Christopher Winship. (2014). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. 2^e édition. Cambridge and New York: Cambridge University Press.

Il s'agit d'une analyse un peu technique des méthodes d'EIC par la terminologie issue de la sociologie. Le chapitre 9, «Instrumental Variable Estimators of Causal Effects» (p. 291-324), propose une synthèse de la logique et des procédures d'estimation par les variables instrumentales.

- Kuhn, Andreas, Jean-Philippe Wuellrich, and Josef Zweimüller. 2010. *Fatal Attraction? Access to Early Retirement and Mortality*. Document de réflexion de l'IZA n° 5160. Bonn: Forschungsinstitut zur Zukunft der Arbeit. <http://ftp.iza.org/dp5160.pdf>

Utilisation des variations régionales dans le changement de l'âge de la retraite en Autriche comme variable instrumentale dans l'étude de l'effet de la retraite anticipée sur la santé des travailleurs.

- Galiani, S., Rossi, M. A., & Schargrotsky, E. (2011). «Conscription and crime: Evidence from the Argentine draft lottery». *American Economic Journal: Applied Economics*. <https://www.aeaweb.org/articles?id=10.1257/app.3.2.119>

Utilisation innovante et parlante d'une conscription par tirage au sort (pour le service militaire) en tant que variable instrumentale. Étude très lisible, constitue un prototype pour une ligne de recherche utilisant les procédures de tirage au sort comme instruments.

Appariement

- Heinrich, Carolyn, Alessandro Maffioli et Gonzalo Vázquez. 2010. *A Primer for Applying Propensity Score Matching. Impact-Evaluation Guidelines Technical Notes N° IDB-TN-161*. Washington: Banque interaméricaine de développement. <https://publications.iadb.org/publications/english/document/A-Primer-for-Applying-Propensity-Score-Matching.pdf>

Comme le guide sur la régression par discontinuité ci-dessous, cette référence est destinée aux gestionnaires d'évaluation avertis.

- Caliendo M., Kopeinig S., (2008), «Some practical Guidance for the implementation of propensity score matching». *Journal of Economic Surveys*; Volume 22, Numéro 1 - Pages 31-72. <https://doi.org/10.1111/j.1467-6419.2007.00527.x>

Un texte classique sur les choix à faire dans le cadre de la mise en œuvre d'une approche par appariement par score de propension (en ce qui concerne les estimations, les algorithmes d'appariement, l'évaluation de la qualité de l'appariement, la sensibilité des effets estimés du traitement).

- Harris, H. et Horst, S. J. (2016). «A Brief Guide to Decisions at Each Step of the Propensity Score Matching Process», *Practical Assessment, Research, and Evaluation*: Vol. 21, Article 4. DOI: <https://doi.org/10.7275/yq7r-4820>. Disponible à l'adresse suivante: <https://scholarworks.umass.edu/pare/vol21/iss1/4>.

Semblable à l'article précédent.

Approche de la régression par discontinuité (regression discontinuity design)

- Jacob, Robin, Pei Zhu, Marie-Andrée Somers, et Howard Bloom. (2012). *A Practical Guide to Regression Discontinuity*. New York: MDRC. https://www.mdrc.org/sites/default/files/RDD%20Guide_Full%20rev%202016_0.pdf

Un examen exceptionnellement accessible et complet de la méthode de discontinuité de la régression qui inclut une bibliographie soigneusement sélectionnée.

- Lee, D. S. et Lemieux T. (2009), *Regression discontinuity designs in economics*, NBER

Document de travail N° 14723, National Bureau of Economic Research, Cambridge, MA, 2009. http://www.nber.org/papers/w14723.pdf?new_window=1

Cet article constitue une sorte de «guide d'utilisation» pour l'approche de la régression par discontinuité, qui fournit une description de la logique de la méthode et illustre différentes manières d'estimer les approches par discontinuité de la régression.

Annexe 2. Suggestions de cours relatifs aux EIC en ligne

Un cours d'introduction aux EIC à destination des AG peut couvrir les thèmes suivants:

- Introduction aux approches de l'évaluation: évaluations de processus (pourquoi sont-elles importantes et en quoi sont-elles liées aux EIC) et évaluation d'impact (approches contrefactuelles et fondées sur la théorie)
- Que sont les EIC? À quelles questions d'évaluation peuvent répondre les EIC efficaces? Pour quels critères d'évaluation les EIC fournissent-elles des éléments de preuve importants? Pourquoi les EIC sont-elles importantes?
- Comment fonctionnent-elles? (Groupe traité) et groupe témoin, comparaisons «avant/après», principes d'inférence
- Aperçu des méthodes d'EIC (caractéristiques, points forts et points faibles de chaque méthode)
 - o Essai aléatoire contrôlé (randomised control trial)
 - o Appariement
 - o Double différence
 - o Approche de la régression par discontinuité (regression discontinuity design)
 - o Variable instrumentale
 - o Indication d'autres méthodes possibles (par ex. techniques d'analyses spatiales)
- Données requises
 - o Exigences pour les groupes traités et témoins
 - o Sources possibles (suivi du FSE, ensembles de données administratives, enquêtes)
 - o Règles du RGPD
- Synthèse des étapes de mise en œuvre:
 - o Planification des EIC (plan d'évaluation, faisabilité d'une EIC)
 - o Commandes d'EIC (cahier des charges, critères de sélection et méthodes)
 - o Gestion des EIC (vérification des éléments livrables et interactions avec l'évaluateur)
 - o Diffusion des constatations des EIC (types de public et voies de communication)

Un cours structuré de la manière exposée plus haut devrait inclure des exemples à étudier lors d'ateliers et durerait 2 ou 3 jours. Une approche pour donner un tel cours consisterait à adopter une méthode d'apprentissage fondée sur un problème et à utiliser une mesure politique du programme comme cas concret à étudier.

Annexe 3. Évaluation d'impact contrefactuelle - Exemples mentionnés dans le Guide

Tableau 6 Caractéristiques des EIC mentionnées dans le guide comme exemples

Pays	Titre - année	Politique évaluée et période de traitement observée	Méthode	Groupe traité et groupe témoin - définition et taille	Variables de résultat	Période d'observation	Données
Suède	Effets d'un soutien intensifié apporté aux jeunes chercheurs d'emploi — Résultats de l'évaluation réalisée par Young Future (2019)	Soutien intensifié apporté par le service public de l'emploi aux chercheurs d'emploi Juin 2017 - janvier 2018	Essai aléatoire contrôlé (randomised control trial)	Groupe traité: jeunes âgés de 18 à 24 ans, sélectionnés pour recevoir le soutien intensifié Groupe témoin: jeunes âgés de 18 à 24 ans, sélectionnés pour recevoir le soutien habituel Groupe traité: 2 792 - Groupe témoin: 1 897	- part de chômeurs - % de sortie vers un emploi- nombre moyen de jours de chômage	12 mois après l'entrée	- Registres publics de l'emploi pour les variables de résultat- Enquête auprès des fonctionnaires et intermédiaires pour l'emploi afin de mesurer l'intensité du soutien apporté
Italie (Marches)*	Impacts des interventions financées par le FSE en 2014-2020 concernant le chômage de longue durée dans la région des Marches (2020)	Stages, expériences professionnelles, placements professionnels, chèques formation pour les chômeurs (dont les chômeurs de longue durée) 2017 — juillet 2019	Appariement par score de propension	Groupe traité: chômeurs de longue durée traités Groupe témoin: chômeurs de longue durée sur la période 2016-2018 inscrits dans les SPE Groupe traité: 526 pour les stages, 1 058 pour les placements professionnels, 236 pour les expériences professionnelles dans les communes et 241 pour les chèques formation. 77255	- probabilité d'être employé à un point dans le temps donné après les interventions- probabilité d'être employé en vertu d'un contrat à durée indéterminée- nombre de jours travaillés au cours d'une période suivant les interventions	6, 9, 12, 15 et 18 mois après l'entrée dans les interventions	- Données de suivi du FSE pour le groupe traité- données administratives provenant des SPE concernant les personnes enregistrées - Données administratives sur les contrats de travail activés dans la région (Comunicazioni Obbligatorie - COB)
Pologne (région de Lublin)*	Évaluation des mesures destinées aux travailleurs âgés dans la région de Lublin financées par le FSE au cours de la période de programmation 2007-2013 (2015)	Plusieurs types de politiques actives du marché du travail 2007-2013	Appariement par score de propension + Double différence	Groupe traité: échantillon de chômeurs âgés de plus de 50 ans ayant reçu un soutien Groupe témoin: chômeurs âgés de plus de 50 ans qui n'ont participé à aucune mesure et qui étaient inscrits dans un registre national de chômage au moment où les membres du groupe traité ont participé au programme évalué Groupe traité: 145 - Groupe témoin: 67102	- probabilité de sortie du statut de chômeur inscrit	6, 12 et 18 mois après l'entrée dans les interventions	- données administratives provenant des SPE

Pays	Titre - année	Politique évaluée et période de traitement observée	Méthode	Groupe traité et groupe témoin - définition et taille	Variables de résultat	Période d'observation	Données
Lettonie	«Vocational Training for Unemployed Youth in Latvia: Evidence from a Regression Discontinuity Design» (2018)	Formation professionnelle ciblant les NEET âgés de 15 à 29 ans 2014-2015	Approche de la régression par discontinuité (RDD)	Groupe traité: chômeurs ayant participé à la formation professionnelle dans l'année suivant la date d'inscription Groupe témoin: chômeurs inscrits entre juin 2013 et décembre 2015 n'ayant participé à aucune mesure Groupe traité: 898 - Groupe témoin: 10717	- probabilité d'être employé à un point dans le temps donné après les interventions- revenus	12-36 mois après l'inscription	- données administratives provenant de l'agence nationale lettone pour l'emploi - informations sur les participants et les non-participants enregistrés en tant que chômeurs à des dates spécifiques- données administratives provenant du service national des impôts - informations sur les conditions d'emploi à différentes dates et sur les revenus des personnes
Allemagne	Effets sur l'emploi d'une formation linguistique dispensée aux migrants sans emploi (2021)	Cours de langue (+ expériences professionnelles pour certaines des personnes traitées) pour les chercheurs d'emploi migrants 2014	Variation instrumentales	Groupe traité: participants migrants qui ont commencé une formation linguistique à des fins professionnelles en 2014 Groupe témoin: échantillon aléatoire de non-participants ayant connu au moins une période au cours de laquelle ils ne disposaient pas de la nationalité allemande et ne participant à aucune mesure de politique active du marché du travail ni aucun cours d'intégration Groupe traité: 8968 - Groupe témoin: 26463	- probabilité d'être employé (emploi régulier, emploi régulier > 6 mois, emploi rémunéré au-dessus du seuil de risque de pauvreté, emploi qualifié)- revenus quotidiens et cumulés provenant d'un emploi régulier-jours cumulés en emploi régulier	24 mois après l'entrée dans les interventions	- Biographies intégrées sur l'emploi pour la détermination des deux groupes et des variables de résultat- Données collectées rétrospectivement au cours d'entretiens avec des chercheurs d'emploi et de personnes chargées de la gestion des dossiers (Werdegangshistorie, WGH) afin de collecter plus d'informations sur les périodes de travail indépendant ou de congé parental ainsi que sur l'éducation et l'expérience professionnelle à l'étranger.
Italie (Trente)	«Are Vocational Training Programmes Worth Their Cost? Evidence from a Cost-Benefit Analysis» (2018)	Cours de formation professionnelle à destination des chômeurs 2010-2011	PSM	Groupe traité: chômeurs participant à des formations professionnelles, financées par les ressources provinciales et par le FSE Groupe témoin: personnes inscrites comme chômeurs à la date de début des formations individuelles dans les registres du SPE local Groupe traité: 818 (formations de la province), 114 (formations du FSE) - groupe témoin: 17 236 (liés aux formations de la province), 1 152 (liés aux formations du FSE)	- probabilité d'être employé -revenus	3, 6, 12, 18, 24 et 36 mois après l'entrée dans les interventions	- données de suivi du FSE pour le groupe traité- données administratives provenant des SPE concernant les personnes enregistrées - données administratives sur les contrats de travail activés dans la région (Comunicazioni Obbligatorie - COB)- données administratives sur les déclarations fiscales provenant de l'INPS - données administratives provenant du service de l'éducation de la province

Pays	Titre - année	Politique évaluée et période de traitement observée	Méthode	Groupe traité et groupe témoin - définition et taille	Variables de résultat	Période d'observation	Données
Allemagne	«Do Job creation schemes improve the social integration and well-being of the long-term unemployed?» (2020)	Programme de création d'emploi (emplois subventionnés pour une durée maximale de 36 mois) pour les personnes vulnérables (chômeurs de longue durée) 2015 – juin 2017	PSM	Groupe traité: échantillon des participants ayant répondu à l'ensemble du questionnaire Groupe témoin: non-participants inscrits auprès des SPE et présentant les mêmes critères d'admissibilité que les personnes du groupe témoin ayant répondu à l'ensemble du questionnaire Groupe traité: 2 531 pour la vague 1, 1 191 pour la vague 2 et 450 pour la vague 3 - groupe témoin: comme ci-dessus	- probabilité d'être employé - mesures subjectives de la satisfaction à l'égard de la vie, de la santé mentale, de l'intégration sociale, du statut social	7, 18, 29 mois à partir de l'entrée dans le programme	- Biographies intégrées sur l'emploi pour la détermination des deux groupes et des variables de résultat- étude sur les variables de résultat subjectives (résultats difficilement quantifiables)
Espagne (Asturies)*	Évaluation du PO du FSE pour les Asturies (2019)	Méthodes d'enseignement pour des groupes d'élèves à risque de décrochage scolaire (deuxième et troisième année d'enseignement secondaire) 2016-2017	PSM	Groupe traité: élèves participant aux interventions Groupe témoin: élèves présentant des critères d'admissibilité comparables Groupe traité: 1053 - Groupe témoin: 3852	- probabilité de passer à l'année supérieure	12 mois après l'entrée	- Données administratives sur l'éducation: base de données SAUCE (Ministère de l'éducation)
Portugal	Évaluation du système de bourse de l'enseignement supérieur pour les étudiants issus de milieux défavorisés au Portugal (2020) - pas encore publiée	Bourses pour les étudiants de l'enseignement supérieur issus de familles à faibles revenus 2021-2018	RDD	Groupe traité: étudiants demandant à entrer en première année de master ou de licence Groupe témoin: étudiants non admissibles car leurs revenus sont supérieurs au seuil défini Groupe traité: 130602 - Groupe témoin: 25400	- probabilité d'être inscrit à la fin de la première année- nombre de crédits obtenus à la fin de la première année - probabilité d'obtenir un diplôme - nombre d'années nécessaires à l'obtention d'un diplôme- note finale pour l'obtention du diplôme	12-48 mois à partir de la bourse	- Données administratives sur les étudiants demandant la bourse (DGES)- données administratives avec des informations sur le parcours et la progression scolaires (DGEEC)
Pologne (Podlachie)*	Évaluation de l'impact sur la popularisation de l'enseignement professionnel produit par le soutien apporté par le PO de Podlachie sur la période 2014-2020 (2020)	Projet du FSE «Bon métier - Bonne vie», visant à promouvoir l'enseignement et la formation professionnels et ciblant les écoles secondaires du premier cycle 2017-2019	PSM	Groupe traité: échantillon d'élèves ayant participé Groupe témoin: élèves présentant des caractéristiques semblables à ceux du groupe traité mais ne participant pas au projet Groupe traité: 200 - Groupe témoin: 384	- fréquentation d'une école technique au moment de l'étude- fréquentation d'une école professionnelle du 1 ^{er} degré au moment de l'étude- fréquentation d'une école technique ou d'une école professionnelle du 1 ^{er} degré au moment de l'étude- candidature à une école secondaire, une école technique, une école professionnelle du 1 ^{er} degré, une école dispensant une formation professionnelle	12-36 mois après les interventions	- Données de surveillance - Registre national de la population (PESEL) pour la sélection du groupe témoin- Enquête pour mesurer les variables de résultat

COMMISSION EUROPÉENNE

Pays	Titre - année	Politique évaluée et période de traitement observée	Méthode	Groupe traité et groupe témoin - définition et taille	Variables de résultat	Période d'observation	Données
Italie (Piémont)*	Effets sur l'emploi du programme de chèques pour des services à l'emploi (Buono servizi lavoro - BSL financé par le PO du FSE dans le Piémont au cours de la période 2014-2020) (2019)	Chèques pour des services à l'emploi (orientation + formation + stage) pour les chômeurs vulnérables connaissant une période de chômage d'au moins 6 mois 2018	PSM	Groupe traité: chômeurs traités Groupe témoin: chômeurs inscrits auprès des SPE présentant des caractéristiques semblables (âgés de 30 ans et chômeurs depuis 6 mois) Groupe traité: 8125 - Groupe témoin: 130000	- probabilité d'être employé- probabilité d'être employé en vertu d'un contrat à durée indéterminée	6, 12, 16 mois à partir des interventions	- données de suivi du FSE pour le groupe traité- données administratives provenant des SPE concernant les personnes enregistrées - données administratives sur les contrats de travail activés dans la région (Comunicazioni Obbligatorie - COB)
Allemagne*	Évaluation du programme pour l'intégration des chômeurs de longue durée en Allemagne sur la période 2014-2020 (2019)	Mesures d'intégration pour les chômeurs de longue durée de 24 mois ou plus (emplois subventionnés) 2015-2017	Double différence (l'intention de traiter a été estimée)	Groupe traité: chômeur de longue durée potentiellement admissibles au cours de la période du programme (à partir de 2015) Groupe témoin: personnes remplissant les critères d'admissibilité avant la mise en œuvre du programme (2010-2012) Groupe traité et groupe témoin: 134 515 personnes pour 237 874 épisodes de chômage (35 % pour le groupe traité et 65 % pour le groupe témoin)	- probabilité d'être employé (emploi avec des contributions sociales)	24 mois à partir des interventions	- données de surveillance administrative provenant du bureau fédéral de l'administration (base de données ZUWES) - Biographies intégrées sur l'emploi (BIE) pour les variables de résultat
Allemagne (Bade-Wurtemberg)*	Deuxième rapport d'évaluation de l'objectif spécifique A1.1 au titre du PO du FSE du Bade-Wurtemberg, pour la période 2014-2020 (2020)	Mesures d'intégration des chômeurs, notamment des chômeurs de longue durée et autres personnes vulnérables mars 2016 - décembre 2017	PSM	Groupe traité: participants identifiés à partir de la base de données BIE Groupe témoin: personnes remplissant les critères d'admissibilité au cours de la même période Groupe traité: 1 578 (sur un total de 1 800) - groupe témoin: s.o.	- probabilité d'être employé (emploi avec des contributions sociales) - probabilité de continuer à toucher des allocations chômage	15 mois à partir des interventions	- Données de surveillance - Biographies intégrées sur l'emploi (BIE) pour les variables de résultat

Remarque: s.o.

Remarque: * = évaluations identifiées dans le Helpdesk

COMMENT PRENDRE CONTACT AVEC L'UNION EUROPÉENNE?

En personne

Dans toute l'Union européenne, des centaines de centres d'information Europe Direct sont à votre disposition. Pour connaître l'adresse du centre le plus proche, consultez la page suivante: https://europa.eu/european-union/contact_fr

Par téléphone ou courrier électronique

Le service Europe Direct répond à vos questions sur l'Union européenne. Vous pouvez prendre contact avec ce service:

- via un numéro gratuit: 00 800 6 7 8 9 10 11 (certains opérateurs facturent cependant ces appels),
- via le numéro standard suivant: +32 22999696, ou
- par courrier électronique via la page: https://europa.eu/european-union/contact_fr

COMMENT TROUVER DES INFORMATIONS SUR L'UNION EUROPÉENNE?

En ligne

Des informations sur l'Union européenne sont disponibles dans toutes les langues officielles de l'Union sur le site web Europa à l'adresse suivante: https://europa.eu/european-union/index_fr

Publications de l'Union européenne

Vous pouvez télécharger ou commander des publications de l'UE gratuites et payantes à l'adresse: <https://op.europa.eu/fr/publications>. Vous pouvez obtenir plusieurs exemplaires de publications gratuites en contactant Europe Direct ou votre centre d'information local (voir https://europa.eu/european-union/index_fr).

Droit de l'Union et documents connexes

Pour accéder aux informations juridiques de l'Union européenne, notamment à l'ensemble du droit de l'Union depuis 1952 dans toutes les versions linguistiques officielles, veuillez consulter EUR-Lex à l'adresse suivante: <http://eur-lex.europa.eu>

Données ouvertes de l'UE

Le portail des données ouvertes de l'Union européenne (<http://data.europa.eu/euodp/fr>) donne accès à des ensembles de données provenant de l'UE. Les données peuvent être téléchargées et réutilisées gratuitement, à des fins commerciales ou non commerciales.



Office des publications
de l'Union européenne