

EVALUATION D'IMPACT EN PRATIQUE

GLOSSAIRE

Activité : Toute action entreprise ou travail réalisé dont les contributions telles que des fonds, l'assistance technique et autres types de ressources sont mobilisées pour produire des résultats spécifiques

Hypothèse Alternative : En évaluation d'impact, l'hypothèse alternative est généralement l'hypothèse selon laquelle l'hypothèse nulle est fautive ; en d'autres mots, l'intervention n'a pas un impact sur les résultats.

Attrition : Lorsqu'il y a déperdition de certaines unités de l'échantillon entre deux collectes de données, par exemple parce que les migrants n'ont pas été suivis. L'attrition est un cas d'absence de réponse d'unités. L'attrition peut causer des erreurs systématiques dans les évaluations d'impact si elle n'est pas corrélée au statut du traitement.

Conditions de base : Préalable à l'intervention, ex-ante. La situation avant une intervention, à partir de laquelle les progrès peuvent être mesurés, ou à partir de laquelle on peut faire des comparaisons. Les données des conditions de base sont recueillies avant la mise en œuvre d'un programme ou d'une politique en vue d'évaluer la situation ex ante.

Comparaison avant-après : Appelée aussi pré-post comparaison ou comparaison réflexive, une comparaison avant-après tente d'établir l'impact d'un programme en suivant les changements au niveau des résultats pour les bénéficiaires du programme au fil du temps et en utilisant des mesures avant et après la mise en œuvre du programme.

Erreur systématique (biais) : Le biais ou erreur systématique d'un estimateur est la différence entre l'attente de l'estimateur et la vraie valeur du paramètre estimé. En évaluation d'impact, c'est la différence entre l'impact que vous calculez et le vrai impact du programme.

Données de recensement : Les données qui couvrent toutes les unités de la population d'intérêt (univers). *Contraste avec* : données d'enquête.

Grappe : Une grappe est un groupe d'unités qui sont d'une façon ou d'une autre, similaires. Par exemple dans un échantillonnage d'écoliers, les écoliers d'une même école appartiendront à une grappe, parce qu'ils partagent les mêmes installations scolaires et les mêmes enseignants, et vivent dans le même voisinage.

Echantillon en grappes : Un échantillon obtenu à travers une sélection aléatoire de grappes, après lequel soit toutes les unités de grappes sélectionnées constituent l'échantillon, ou des unités contenues dans chaque grappe sont sélectionnées au hasard. Chaque grappe dispose d'une probabilité bien définie d'être choisie, et les unités contenues dans chaque grappe ont la possibilité bien définie d'être sélectionnées.

Groupe témoin, appelé aussi groupe de contrôle ou de référence : Un groupe témoin valable aura les mêmes caractéristiques que le groupe de bénéficiaires du programme (groupe étudié ou expérimental), à l'exception du fait que les unités contenues dans le groupe témoin ne bénéficient du programme. Les groupes témoins sont utilisés pour estimer le contrefactuel.

Contrefactuel : Le contrefactuel est une estimation de ce que le résultat (y) aurait été pour un participant de programme en l'absence du programme (P). Par définition, le contrefactuel ne peut être observé. Par conséquent, il doit être estimé à travers l'utilisation de groupes témoins.

Coûts-avantages : Les calculs ex-ante du total des coûts et avantages attendus sont utilisés pour estimer et évaluer les propositions de projets. Les coûts-avantages peuvent être calculés ex-poste dans les évaluations d'impact si les avantages peuvent être quantifiés en termes monétaires et si des données sur les coûts sont disponibles.

Rapport coût-efficacité : Consiste à comparer des interventions identiques sur la base du coût et de l'efficacité. Par exemple, les évaluations d'impact sur divers programmes d'éducation permettent aux décideurs politiques de prendre des décisions plus informées sur le type d'intervention susceptible d'atteindre les objectifs souhaités, compte tenu de leur contexte et contraintes particuliers.

Méthode des Doubles Différences ou différence-en-différences ou D-en-D. La méthode des doubles différences estime le contrefactuel pour le *changement* au niveau du résultat pour le groupe expérimental en prenant le *changement* au niveau du résultat pour le groupe témoin. Cette méthode nous permet de prendre en compte toutes les différences existant entre le groupe expérimental et le groupe témoin, qui sont consistants dans le temps. Les deux différences sont donc avant et après, et entre le groupe expérimental et le groupe témoin

Effet : Changement prévu ou accidentel dû directement ou indirectement à une intervention

Estimateur : En statistiques, un estimateur est une statistique (une fonction des observations de données-échantillons) qui est utilisée pour estimer un paramètre ; une **estimation** est le résultat provenant de l'application réelle de la fonction à des données-échantillons.

Evaluation : Les évaluations sont des estimations périodiques et objectives de projets, programmes ou politiques prévus, en cours ou achevés. Les évaluations sont utilisées pour répondre à des questions spécifiques souvent liées à la conception, la mise en œuvre et ou aux résultats.

Validité externe : La validité externe signifie que l'impact causal découvert dans l'évaluation d'impact peut être généralisé à l'univers de toutes les unités éligibles. Pour qu'une évaluation soit valable, il est nécessaire que l'échantillon d'évaluation soit un échantillon représentatif de l'univers des unités éligibles.

Enquête de suivi : Appelée aussi enquête post-intervention ou enquête ex-poste. C'est une enquête qui est conduite après que le programme a commencé et une fois que les bénéficiaires

ont joui pendant un certain temps déjà, des avantages du programme. Une évaluation d'impact peut comprendre plusieurs enquêtes de suivi.

Effet Hawthorne : L' « effet Hawthorne » survient lorsque le simple fait que vous observez des unités, les fasse réagir différemment.

Hypothèse : Une **hypothèse** (du grec *ὑπόθεσις* ; hypothèses plurielles) est une proposition d'explication pour un phénomène observable. Voir aussi hypothèse Nulle et hypothèse alternative.

Evaluation d'impact : Une évaluation d'impact est une évaluation qui essaie d'établir un lien causal entre un programme ou intervention et un ensemble de résultats. Une évaluation d'impact essaie de répondre à la question de savoir si un programme est responsable des changements au niveau de résultats d'intérêt. *Contraste avec* évaluation de processus.

Indicateur : Un indicateur est un variable qui mesure un phénomène d'intérêt pour l'évaluateur. Le phénomène peut être une entrée, un résultat, une caractéristique ou un attribut.

Entrées : Les ressources financières, humaines et matérielles utilisées pour le développement d'une intervention.

Variable instrumentale : Une variable instrumentale est une variable qui aide à identifier l'impact causal d'un programme, lorsque la participation au programme est partiellement déterminée par les bénéficiaires potentiels. Une variable doit avoir deux caractéristiques pour être une bonne variable instrumentale : (i) elle doit être corrélée avec la participation au programme ; et (ii) elle ne doit pas être corrélée avec les résultats Y (sauf à travers la participation au programme) ou avec des variables non-observées.

Corrélation intra-grappe : La corrélation intra-grappe est la corrélation (ou similarité) dans les résultats ou caractéristiques entre des unités qui appartiennent à la même grappe. Par exemple, des enfants qui fréquentent la même école seront généralement identiques ou corrélés en termes de zone de résidence ou d'origine socio-économique.

Validité interne : On entend par validité interne le fait que l'évaluation d'impact utilise un groupe témoin valable, c'est-à-dire, un groupe témoin qui est une estimation valable du contrefactuel. **Principe de vouloir traiter ou Estimateur ITT :** L'estimateur ITT n'est que la différence directe dans l'indicateur de résultat Y pour le groupe auquel nous offrons un traitement et le même indicateur pour le groupe auquel nous n'offrons pas de traitement. *Contraste avec* Effet du Traitement-sur-le-Traité (estimateur TOT).

Effet John Henry : L' « effet John Henry » survient lorsque des unités de référence travaillent ardemment pour compenser le fait de ne pas recevoir un traitement. Lorsqu'on compare les unités traitées à ces unités de référence « qui travaillent plus ardemment », l'estimation de l'impact du programme sera biaisée : nous estimerons un impact du programme plus petit que le vrai impact que nous trouverions si les unités de référence ne faisaient pas d'efforts additionnels.

Appariement : L'appariement est une méthode d'évaluation non-expérimentale qui utilise de grands ensembles de données et des techniques statistiques lourdes pour construire le meilleur groupe témoin possible pour un groupe expérimental.

Effet minimum désiré : Le changement minimum au niveau des résultats qui justifie l'investissement qui a été fait dans une intervention, qui représente non seulement le coût du programme et le type d'avantages qu'il procure mais aussi le coût de l'option de ne pas avoir investi des fonds dans une intervention alternative. L'effet minimum désiré constitue une entrée pour les calculs de puissance : les échantillons d'évaluation doivent être assez grands pour détecter au moins les effets minima désirés ayant suffisamment de puissance.

Suivi : Le suivi est le processus continu de collecte et d'analyse d'informations en vue d'évaluer le niveau de performance d'un programme ou d'une politique. Il est essentiellement basé sur des données administratives pour suivre la performance par rapport aux résultats attendus, faire des comparaisons par rapport à d'autres programmes et analyser les tendances dans le temps. Généralement, le suivi suit les entrées, les activités et les extrants, quoiqu'occasionnellement, il inclue aussi les résultats. Le suivi est utilisé pour informer la gestion et les décisions de tous les jours.

Non-réponse : Se produit lorsque des données manquent ou sont incomplètes pour des unités échantillonnées. La non-réponse survient lorsque l'information n'est pas disponible pour des unités d'échantillon, c.-à-d. lorsque l'échantillon réel est différent de l'échantillon prévu. L'attrition constitue une forme de non-réponse d'unités. La non-réponse intervient lorsque les données sont incomplètes pour des unités échantillonnées à un moment donné. La non-réponse peut causer des biais (erreurs systématiques) dans les résultats de l'évaluation si ceci est associé à la qualité du traitement.

Hypothèse nulle : Une hypothèse nulle est une hypothèse qui pourrait être falsifiée sur la base des données observées. L'hypothèse nulle généralement propose une position générale ou par défaut. Généralement, en évaluation d'impact, la position par défaut est qu'il n'y a pas de différence entre le traitement et le groupe témoin, ou en d'autres termes, que l'intervention n'a pas d'impact sur les résultats.

Résultat : Les résultats peuvent être intermédiaires ou définitifs. Un résultat est la conclusion de l'intérêt qui s'en suit à travers une combinaison de facteurs liés à la demande et à l'offre. Par exemple, si une intervention entraîne la prestation accrue de services de vaccination, le nombre réel de vaccinations serait un résultat, étant donné qu'il ne dépend pas seulement de la fourniture de vaccins mais aussi du comportement des bénéficiaires visés : se présentent-ils au point de service pour se faire vacciner ? Les résultats finaux/à long terme sont des résultats éloignés ; cette distance peut être interprétée en dimension temporelle (cela prend beaucoup de temps avant qu'on n'obtienne le résultat).

Extrant : Les produits, les biens d'équipement et de service qui sont produits (fournis) directement par une intervention. Les extrants peuvent inclure aussi les changements provenant de l'intervention et qui sont pertinents pour la réalisation des réalisations.

Efficacité : L'efficacité est égale à 1 moins la probabilité d'une erreur de type II, allant de 0 à 1. Les niveaux populaires d'efficacité sont 0,8 et 0,9. Les niveaux élevés d'efficacité sont plus conservatifs et diminuent la vraisemblance d'une erreur de type II. L'efficacité d'une évaluation d'impact est puissance élevée lorsqu'il existe un faible niveau de non détection des impacts réels du programme, c.-à-d., de commettre une erreur de type II.

Calculs de l'efficacité : Les calculs de l'efficacité indiquent la taille d'échantillon requise pour une évaluation pour détecter un certain effet minimum désiré. Les calculs d'efficacité dépendent de paramètres tels que l'efficacité (ou la vraisemblance d'erreur de type II), le niveau de signification, la variance et la corrélation intra-grappe du résultat d'intérêt.

Evaluation de processus : Une évaluation de processus est une évaluation qui tente d'établir le niveau de qualité ou de réussite des processus ou d'un programme : par exemple : l'adéquation de processus administrative, l'acceptabilité des avantages du programme, la clarté de la campagne d'informations, les dynamiques des organes d'exécution, leur politique, leurs mécanismes de prestation de services, leurs pratiques en matière de gestion et divers rapports entre ces éléments. *Contraste avec :* Evaluation d'Impact.

Echantillon aléatoire : La meilleure manière d'éviter un échantillon biaisé ou non représentatif consiste à procéder à un échantillonnage aléatoire. Un échantillon aléatoire est un échantillon probabiliste dans lequel chaque individu de la population objet d'échantillonnage a la même chance (probabilité) d'être sélectionné.

Attribution aléatoire ou plans témoins aléatoires : L'attribution aléatoire est comme étant la méthode la plus robuste pour l'estimation des contrefactuels. Elle est souvent appelée la "Norme d'Excellence" de l'évaluation d'impact. Avec cette méthode, les bénéficiaires sont sélectionnés au hasard pour recevoir une intervention, et chaque bénéficiaire dispose de la même chance de recevoir le programme. Avec des tailles d'échantillon assez grandes, le processus d'attribution aléatoire assure l'équivalence au niveau des caractéristiques observées et non observées, entre les groupes expérimentaux et les groupes témoins. Ce qui prend en compte toute erreur systématique de sélection.

Offre aléatoire : L'offre randomisée est méthode qui permet d'identifier l'impact d'une intervention. Lorsque l'administrateur de programme peut choisir au hasard les unités auxquelles il peut offrir le traitement, à partir des unités éligibles, mais il ne peut obtenir une conformité ; il ne peut forcer aucune unité à participer/accepter le traitement, et il ne peut refuser de laisser une unité participer si l'unité insiste à participer/accepter être traitée. Dans la méthode d'offre aléatoire ou randomisée, l'offre aléatoire de programme est utilisée comme variable instrumentale de la vraie participation au programme.

Promotion aléatoire : La promotion aléatoire est une méthode identique à l'offre aléatoire. Au lieu de sélectionner au hasard, des unités auxquelles le traitement est offert, les unités sont sélectionnées au hasard pour la promotion du traitement. De cette manière, le programme reste ouvert pour chaque unité.

Méthodes de sélection aléatoire : La méthode de sélection aléatoire et nom de groupe utilisé pour diverses méthodes qui utilisent l'attribution au hasard pour identifier le contrefactuel ; parmi ces méthodes, nous avons : l'attribution aléatoire du traitement et la promotion aléatoire.

Régression : En statistiques, l'analyse de régression comprend toutes les techniques de modélage et d'analyse de divers variables, lorsque le point d'attention est la relation entre une variable dépendante et une ou plusieurs variables explicatives. En évaluation d'impact, l'analyse de régression nous permet de comprendre comment la valeur caractéristique de l'indicateur de résultat Y (variable dépendante) change lorsque l'attribution du traitement ou le groupe témoin P (variable explicative) varie, tandis que les caractéristiques des bénéficiaires (autres variables explicatives) demeurent fixes.

Plan Expérimental de Discontinuité de la Régression (RDD) : Le Plan de Discontinuité de la Régression (RDD) est une méthode d'évaluation non-expérimentale. Il convient aux programmes qui utilisent un indice continu pour ordonner les bénéficiaires potentiels, et qui dispose d'un seuil pour les indices pour déterminer si les bénéficiaires potentiels reçoivent le programme ou pas. Le seuil d'indice pour l'éligibilité de programmes fournit un point de division entre le groupe expérimental et le groupe témoin.

Chaîne de résultats : La logique de programme qui explique comment l'objectif de développement doit se réaliser. Il montre le lien qui va des entrées aux résultats en passant par les activités et les extrants.

Echantillon : En statistiques, un échantillon est un sous-ensemble d'une population. Généralement, la population est très vaste, ce qui rend le recensement ou l'énumération complète de toutes les valeurs de la société infaisable ou impossible. Ainsi, les chercheurs peuvent sélectionner un sous-ensemble représentatif de la population (en utilisant un cadre d'échantillonnage) et recueillir des statistiques sur l'échantillon, qui peuvent être utilisés pour faire des inférences ou extrapoler à la population. Ce processus est appelé **échantillonnage**.

Echantillonnage : Processus par lequel des unités sont choisies de la base d'échantillonnage qui est réalisé à partir de la population d'intérêt (univers). Diverses procédures d'échantillonnage alternatives peuvent s'utiliser. Les méthodes d'*échantillonnage probabiliste* sont les plus rigoureuses étant donné qu'elles attribuent une probabilité bien définie à chaque unité sélectionnée. L'échantillonnage aléatoire, l'échantillonnage aléatoire stratifié et l'échantillonnage en grappes constituent des méthodes d'échantillonnage probabiliste. L'échantillonnage non-probabiliste (tel que l'échantillonnage raisonné ou discrétionnaire) peut engendrer des erreurs.

Base d'échantillonnage : La liste la plus complète d'unités provenant de la population d'intérêt (univers) qui être obtenue. Les différences entre la base d'échantillonnage et la population d'intérêt créé une erreur de couverture (échantillonnage). En présence d'erreur de couverture, les résultats provenant de l'échantillon n'ont pas de validité externe pour toute la population d'intérêt.

Erreur de sélection : L'erreur de sélection survient lorsque les motifs pour lesquels un individu participe à un programme sont corrélés avec les résultats. Cette erreur se produit généralement lorsque le groupe contrôle est inéligible ou est objet de libre-sélection en dehors du traitement.

Niveau de signification : Le niveau de signification est généralement représenté par le symbole grec α (alpha). Les niveaux de signification populaires sont 5% (0,05), 1% (0,01) et 0,1% (0,001). Si un test de signification donne une valeur-P inférieure au niveau α (alpha), l'hypothèse nulle est rejetée. Ces résultats sont qualifiés de façon informelle, de « statistiquement significatifs ». Plus le niveau de signification est bas, plus fort est l'évidence requise. Le choix du niveau de signification est une tâche arbitraire, mais pour diverses applications, un niveau de 5% est choisi pour la seule raison qu'elle est conventionnelle.

Effet de propagation : Appelé aussi contamination du groupe témoin. L'effet de propagation survient lorsque le groupe témoin est affecté par le traitement administré au groupe expérimental, même si le traitement n'est pas administré directement au groupe témoin. Si l'effet de propagation sur le groupe témoin est négatif (c.-à-d., ils souffrent à cause du programme), alors la différence directe entre les résultats au niveau du groupe expérimental et du groupe témoin produira la surestimation de l'impact du programme. Par contraste, si l'effet de propagation sur le groupe témoin est positif (c.-à-d., ils bénéficient), il y aura alors une sous-estimation de l'impact du programme.

Efficacité statistique : L'efficacité d'un test statistique est la probabilité que le test rejette l'hypothèse nulle lorsque l'hypothèse alternative est vérifiée (c.-à-d., qu'il n'y aura pas d'erreur du type II). Au fur et à mesure que l'efficacité augmente, la probabilité d'une erreur de type II décroît. La probabilité d'une erreur de type II est aussi appelée le taux faux négatif (β). Par conséquent, l'efficacité est égale à $1 - \beta$.

Echantillonnage stratifié : Il s'obtient en divisant la population d'intérêt (base d'échantillonnage) en groupes (par exemple, male et femelle), et en procédant ensuite à un échantillonnage aléatoire au sein de chaque groupe. Un échantillon stratifié est un échantillon probabiliste ; chaque unité au sein de chaque groupe (ou strate) jouit de la même probabilité d'être choisie.

Données d'enquête : Données couvrant un échantillon de la population d'intérêt. *Contraste avec :* données de recensement.

(Effet du) Traitement-sur-le-Traité : Appelé aussi estimateur TOT. L'effet du traitement-sur-le traité est l'impact du traitement sur ces unités qui ont déjà bénéficié du traitement. *Contraste avec :* Principe de Vouloir Traiter (estimateur ITT).

Groupe expérimental : Appelé aussi groupe de l'intervention ou groupe traité. Le groupe expérimental est le groupe d'unités qui bénéficie de l'intervention par rapport au groupe témoin qui n'en bénéficie pas.

Erreur de type I : Erreur commise en rejetant une hypothèse nulle même lorsque l'hypothèse nulle en réalité tient. Dans le contexte de l'évaluation d'impact, une erreur de type I est commise lorsqu'une évaluation conclut qu'un programme a eu un impact (c.-à-d., l'hypothèse de « non

impact » est rejetée), même si en réalité le programme n'a pas eu d'impact (c.-à-d., l'hypothèse nulle ne tient pas). Le niveau de signification détermine la probabilité de commettre une erreur de type I.

Erreur de type II. Erreur commise en acceptant (en ne rejetant pas) l'hypothèse nulle, même lorsque l'hypothèse nulle ne tient pas. Dans le contexte de l'évaluation d'impact, une erreur de type II est commise lorsqu'une évaluation conclut qu'un programme n'a pas eu d'impact (c.-à-d., l'impact l'hypothèse nulle de non impact n'est pas rejetée) même si le programme a eu un impact (c.-à-d., l'impact l'hypothèse nulle ne tient pas). La probabilité de commettre une erreur de type II est i moins le niveau d'efficacité.

Univers : Le groupe d'unités qui sont éligibles pour recevoir une intervention ou un traitement. L'univers est parfois appelé population d'intérêt.

Variable : En terminologie statistique, une variable est un symbole qui représente un valeur qui peut varier.