

CRCHUM

CENTRE DE RECHERCHE
Centre hospitalier
de l'Université de Montréal



GUIDE PRATIQUE : ÉVALUATION EN MILIEUX RÉELS D'INNOVATIONS NUMÉRIQUES EN SANTÉ



Alexis Metral (CHUM) en collaboration avec David Nguyen-Tri (CHUM), sous la direction de Lise Gauvin
PhD FCAHS (CRCHUM, École de santé publique, Université de Montréal)

Table des matières

GLOSSAIRE.....	2
I. INTRODUCTION À LA PLATEFORME D'ÉVALUATION	5
MISE EN CONTEXTE ET INTRODUCTION SUR LES INNOVATIONS NUMÉRIQUES ET L'ÉVALUATION.....	5
INTRODUCTION À LA PLATEFORME D'ÉVALUATION	6
II. MÉTHODOLOGIES D'ÉVALUATION.....	9
LES QUATRE PHASES À L'ÉVALUATION D'INTERVENTIONS COMPLEXES (OU « D'INNOVATIONS NUMÉRIQUES »).....	10
PHASE 1 : Description précise d'une intervention.....	10
PHASE 2 : Évaluabilité	10
PHASE 3 : Évaluation	11
PHASE 4 : Implantation	15
III. OUTILS ET RESSOURCES EXTERNES.....	17
IV. CAS ILLUSTRATIFS	17
V. LES QUATRE PERSPECTIVES DE RECHERCHE À ADOPTER POUR DES INTERVENTIONS COMPLEXES.....	18
VI. RÉFÉRENCES.....	20

GLOSSAIRE

Nous proposons quelques définitions de termes afin de faciliter la lecture par la suite. Le tableau ci-dessous regroupe ces définitions.

Tableau 1. Définitions de notions importantes en évaluation en milieux réels des IN en santé.

Terme	Définition
Algorithmes « adaptatifs » (ou « d'apprentissage continu »)	Des algorithmes qui modifient leur comportement à l'aide d'un processus d'apprentissage défini. L'adaptation ou les changements de l'algorithme sont mis en œuvre de telle sorte que pour un ensemble donné d'entrées, la sortie peut être différente avant et après la mise en œuvre des changements. Ces changements d'algorithme sont généralement mis en œuvre et validés au moyen d'un processus bien défini et éventuellement entièrement automatisé, qui vise à améliorer les performances sur la base de l'analyse de données nouvelles ou supplémentaires (FDA, 2019).
Algorithmes « fixes »	Des algorithmes qui donnent le même résultat à chaque fois que la même donnée est fournie. En tant que tel, ce type d'algorithme applique une fonction fixe (par exemple, une table de consultation statique, un arbre décisionnel ou un classificateur complexe) à un ensemble donné d'entrées. Ces algorithmes peuvent utiliser des processus manuels pour les mises à jour et la validation (Food & Drug Administration, 2019; https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf).
Entreprise	Toute entité commerciale qui prévoit de vendre ou qui vend une IN de santé à un prestataire de soins de santé. L'entreprise peut être la même que le développeur qui a créé l'IN de santé, ou une autre organisation qui tente de promouvoir l'utilisation de la technologie dans le système de santé et de soins (NICE 2022, https://www.nice.org.uk/corporate/e.cd7).

Terme	Définition
Évaluation	L'appréciation systématique de la conception, de la mise en œuvre et/ou des résultats d'une initiative pour des fins d'apprentissage, de prise de décision et/ou d'avancement des connaissances (https://evaluationontario.ca/about-evaluation).
Innovation	Toute amélioration des pratiques, des technologies ou des façons de faire ayant pour but de créer de la valeur et in fine améliorer la santé et le mieux-être des populations et des intervenants, le tout intégré dans les pratiques de façon durable et généralisée (https://www.chumontreal.qc.ca/innovation/innovation-chum https://www.chumontreal.qc.ca/innovation/innovation-en-sante).
Innovation numérique	Tout produit ou toute technologie numérique, qui mobilise des données et qui s'appuie sur l'intelligence artificielle et une vaste gamme d'algorithmes en santé computationnelle.
Intelligence artificielle	Domaine d'étude ayant pour objet la reproduction artificielle des facultés cognitives de l'intelligence humaine dans le but de créer des systèmes ou des machines capables d'exécuter des fonctions relevant normalement de celle-ci (Office québécois de la langue française; https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/8385376/intelligence-artificielle).
Intervention complexe	Une intervention qui a plusieurs composantes; vise plusieurs pratiques, gestes ou comportements; requiert beaucoup d'expertise de la part des gens qui implantent l'intervention; requiert parfois de l'expertise chez celles et ceux qui la reçoivent, doit être implantée parmi plusieurs groupes, dans différents milieux et à différents niveaux; et/ou qui comporte plus ou moins de flexibilité dans son implantation (Skivington et al., 2021a).
Intervention en milieu réel	Toutes améliorations des pratiques, des technologies ou des façons de faire ayant pour but de créer de la valeur (https://www.chumontreal.qc.ca/innovation/innovation-en-sante).

Terme	Définition
Système de santé apprenant	Système de santé dans lequel les données de santé issues de la prestation de soins et de la recherche sont analysées de manière à faire progresser les connaissances, lesquelles sont ensuite intégrées dans les pratiques à des fins d'amélioration continue (Office québécois de la langue française; https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/26570745/systeme-de-sante-apprenant).

GUIDE PRATIQUE : ÉVALUATION EN MILIEUX RÉELS D'INNOVATIONS NUMÉRIQUES EN SANTÉ

I. INTRODUCTION À LA PLATEFORME D'ÉVALUATION

MISE EN CONTEXTE ET INTRODUCTION SUR LES INNOVATIONS NUMÉRIQUES ET L'ÉVALUATION

Chaque année, de nombreuses petites ou moyennes entreprises (PME) et de jeunes pousses (*start-ups*) québécoises contactent le CHUM afin de tester et d'intégrer leurs innovations numériques (IN) en milieu hospitalier ou dans des milieux communautaires où des patientes et patients reçoivent des traitements et des suivis ou encore tentent d'adopter des pratiques préventives. Ces IN se traduisent par tout produit ou toute technologie numérique, qui mobilise des données et qui s'appuie sur l'intelligence artificielle et une vaste gamme d'algorithmes en santé computationnelle. Or, une intégration réussie de ces IN exige une évaluation scientifique rigoureuse de la valeur des innovations pour les patientes et patients, les équipes, les organisations et la société. Plusieurs critères entrent en ligne de compte dans l'évaluation des IN : leur portée, leur impact, leur acceptabilité, leur faisabilité, la possibilité de les pérenniser et plus encore. Ainsi, la plateforme soutenue par le Centre d'innovation et santé de précision (CISP) a été créée afin d'offrir des services de conceptualisation et d'évaluation scientifique et médico-économique d'IN en santé dans des milieux réels de soins et de vie (hôpital, domicile du patient, communauté, etc.) aux PME et jeunes pousses québécoises de même qu'aux équipes de professionnels de la santé.

Il est important de souligner deux notions fondamentales afin de bien comprendre ce que constitue la plateforme et la spécificité des services qu'elle fournit : l'évaluation et les interventions en milieux réels.

Tout d'abord, bien qu'il n'y ait pas unanimité sur sa définition (Poth et al., 2014), l'[évaluation](#) peut se définir comme une appréciation systématique de la conception, de la mise en œuvre et/ou des résultats d'une initiative pour des fins d'apprentissage, de prise de décision et/ou d'avancement des connaissances. Il faut savoir que l'évaluation est un domaine très vaste, et que le terme n'est pas utilisé de manière homogène. L'article

de Skivington et al. (2021a) présente une autre définition de l'évaluation qui est utile et pertinente au travail qui est réalisé par l'équipe du CISP, à savoir :

« Une étude réalisée pour permettre de juger de la valeur d'une intervention. Les études d'évaluation cherchent généralement :

- À déterminer si une intervention « fonctionne » ou non, c'est-à-dire si elle atteint les résultats escomptés.
- À identifier les impacts imprévus ou négatifs,
- À théoriser le fonctionnement de l'intervention (théorie de l'intervention),
- À étudier son interaction avec le contexte dans lequel elle est mise en œuvre, la manière dont elle contribue au changement du système et ;
- À étudier la manière dont les données générées par l'évaluation peuvent être utilisées pour soutenir la prise de décision dans le monde réel (Skivington et al., 2021a). »

Les interventions en milieux réels (aussi appelées « [innovations](#) ») peuvent se traduire par tous changements des pratiques, des technologies ou des façons de faire ayant pour but de créer de la valeur. Cette dernière peut toucher plusieurs parties prenantes et prendre plusieurs formes. Bien que les interventions/innovations puissent toutes avoir un objectif d'amélioration, il est peu probable qu'elles y parviennent toutes et il se peut aussi que ces interventions aient des effets mixtes ou néfastes d'où la nécessité de réaliser des évaluations.

INTRODUCTION À LA PLATEFORME D'ÉVALUATION

La plateforme d'évaluation est ancrée dans le modèle de performance du CHUM. Ce modèle met de l'avant les valeurs sur lesquelles la mission et la réalité du CHUM reposent, et vise à guider les gestionnaires dans la construction, l'évaluation et l'amélioration de leurs activités et projets. Il s'inspire à la fois de celui mis en place par le Ministère de la Santé et des Services Sociaux (MSSS), du *Quadruple Aim* de l'*Institute for Health Improvement* (IHI) ainsi que du Tableau de Bord Prospectif de *Kaplan et Norton* (Bodenheimer & Sinsky, 2014; Sikka et al., 2015). Tel qu'illustré à la Figure 1, il est composé de cinq dimensions différentes et complémentaires, dont quatre sont inspirées du modèle précédent du *Quadruple Aim*. Toutefois, une 5^e et nouvelle dimension autour de l'équité s'est ajoutée à ce dernier, devenant ainsi le *Quintuple Aim* (Itchhaporia, 2021). Selon ces modèles, la création de la valeur se traduit par :

- Une amélioration de l'expérience patient et des résultats des soins et des services;
- Une amélioration de la santé des populations;
- Une amélioration de l'expérience et du mieux-être du personnel et des équipes;
- Une réduction des coûts de la santé et une optimisation des ressources et;
- Une amélioration de l'équité en santé et en services sociaux.

D'ailleurs, tout récemment, l'idée d'une 6^e valeur soit le développement durable au niveau environnemental se rajoute aux réflexions [voir (Alami et al., 2023)]. La plausibilité d'intégrer cette 6^e dimension dans la plateforme de service en évaluation du CHUM sera abordée ultérieurement.

Finalement, au cœur de ce modèle du CHUM, la valeur est également axée sur la génération et la mobilisation de connaissances dans un système ouvert, innovant, apprenant et enseignant.



Figure 1. Modèle de performance du CHUM basé sur le modèle du Quintuple Aim

La plateforme d'évaluation offre des services visant à évaluer les IN ayant un niveau de maturité de 7 ou plus selon l'échelle de maturité d'Innovation Canada, c'est-à-dire des IN qui ont atteint un état opérationnel prévu et qui sont prêtes à être évaluées en milieu réel. En général, ces évaluations s'intéressent à la portée, l'acceptabilité, la sécurité, la faisabilité, l'efficacité, les freins et leviers à l'implantation, ainsi que les considérations éthiques et économiques associées à l'implantation d'une IN en santé.

Grâce à la plateforme, les données probantes générées pourront entre autres :

- Permettre aux entreprises et aux milieux preneurs du réseau de la santé et des services sociaux (établissement, département, unité, etc.) de démontrer la valeur ajoutée de l'adoption de leurs IN ;
- Accélérer l'adoption d'IN dans des milieux réels de soins et de vie;
- Orienter les décideurs et gestionnaires dans leur choix d'implanter des IN;
- Permettre à la communauté scientifique d'avancer l'état des connaissances sur l'évaluation des IN et les meilleures façons de les étudier en milieux réels;
- Servir à identifier les impacts positifs et les risques pour la santé et le bien-être des patients et patientes, du personnel, ainsi que les impacts sur les systèmes de santé.

Le travail se fait en étroite collaboration avec les partenaires industriels et une équipe interdisciplinaire regroupant des chercheurs et chercheuses, du personnel de recherche, du personnel clinique ou des gestionnaires par exemple. Des patientes et patients partenaires ou proche-aidantes et aidants dédiés au projet complètent cette équipe. La plateforme repose sur des piliers déjà existants au CHUM :

- La Direction de la recherche et de l'innovation
- Le Carrefour de l'innovation et de l'évaluation en santé du CRCHUM
- L'Unité d'évaluation des technologies et des modes d'intervention (UETMIS)
- La Direction de l'évaluation, de l'évaluation, de la performance et de l'éthique (DQEPE)
- Le Centre d'excellence sur le partenariat avec les patients et le public (CEPPP)
- Le Centre d'opérationnalisation du Coût par parcours de soins et services (CO-CPSS) de la direction de la performance budgétaire.
- La Direction de l'enseignement et de l'académie CHUM (DEAC)

Finalement, la plateforme propose un service de transfert des connaissances générées grâce aux évaluations, ainsi qu'un partage continu des nouveaux outils d'évaluation qui sont découverts ou élaborés, et ce, pour les différentes parties prenantes impliquées dans l'évaluation et l'adoption des IN au CHUM et ailleurs au Québec.

Il est important de noter que les fondements de plusieurs d'idées dans ce guide sont tirés de références clés provenant de la littérature scientifique dans le domaine de l'évaluation, des interventions complexes et des innovations numériques en santé. Voici

quelques ressources disponibles et autres outils pertinents dans le domaine de l'évaluation des IN en santé :

- Cadre normatif de preuves pour les technologies numériques en santé du NICE. Le cadre a été mise à jour en 2022 et est disponible sur le [site web du NICE](#) en version anglaise. Une traduction autorisée est disponible sur le site du CHUM.
- Cadre pour le développement et l'évaluation d'interventions complexes : article de [Skivington et collègues \(2021a\)](#) publié dans le [British Medical Journal](#).

II. MÉTHODOLOGIES D'ÉVALUATION

Dans le cadre d'évaluations en milieux réels des innovations numériques en santé, nous estimons que le cadre théorique de Skivington et al. (2021a, 2021b) fait partie des références clés dans le domaine de l'évaluation d'interventions complexes, laquelle permet d'enrichir notre pratique et de conceptualiser l'évaluation comme un processus continu et itératif. Selon la définition d'interventions complexes formulée par ces auteurs, nous pouvons considérer les innovations numériques comme des interventions complexes. En milieux réels, nous retrouvons toute une dimension de complexité de l'interaction entre une intervention et le contexte dans lequel elle est mise en œuvre, ainsi que la complexité de la structure d'une intervention en soi. Les interventions sont considérées comme complexes en vertu des caractéristiques suivantes :

- ✓ Les propriétés de l'intervention elle-même, telles le nombre de composantes de l'intervention et les interactions entre ces composantes;
- ✓ L'éventail des comportements visés par l'intervention;
- ✓ L'expertise et des compétences requis par ceux qui implantent ou reçoivent l'intervention;
- ✓ Le nombre de groupes, de niveaux organisationnels ou de contextes ciblés par l'intervention;
- ✓ Le niveau de flexibilité ou d'adaptation de l'intervention ou de ses composantes qui est autorisé (c'est-à-dire le degré de dynamisme ou d'adaptation de l'intervention).

Ce cadre théorique suggère, entre autres, qu'il existe quatre perspectives à adopter dans le développement et l'évaluation d'interventions complexes, lesquelles seront davantage discutées dans la section suivante, soit la description précise de l'intervention, l'estimation de l'évaluabilité de l'intervention, le cadre d'évaluation, et les conditions et procédures d'implantation. De plus, Skivington et al. (2021a, 2021b) mettent l'accent sur six éléments fondamentaux dans la pratique (c.-à-d., le contexte, l'implication des parties prenantes, la théorie du programme, les incertitudes, le raffinement de l'intervention, les considérations économiques), et qui sont à prendre en compte de manière itérative dans tout projet d'évaluation et durant tout le processus de l'évaluation.

LES QUATRE PHASES À L'ÉVALUATION D'INTERVENTIONS COMPLEXES (OU « D'INNOVATIONS NUMÉRIQUES »)

PHASE 1 : Description précise d'une intervention

Bien que la plateforme d'évaluation ne participe pas d'office au développement ou à l'identification des IN à déployer par le milieu preneur, nous rappelons que ces deux étapes, qui sont effectuées par d'autres parties prenantes, sont tout de même fondamentales afin de répondre aux besoins de la population et des établissements de santé de façon pertinente et efficace. Parmi les différentes étapes préalables nécessaires, nous soulignons une étape particulièrement importante :

- Développer une perspective sur la nature du problème et de sa résolution. Elle permet de clarifier la nature du problème selon les parties prenantes en utilisant des données existantes et la façon dont la nouvelle IN permet de résoudre le problème. Une fois que le problème et la solution appariée sont définis, il est question de déterminer comment le problème est socialement et géographiquement distribué, incluant qui sont les plus/moins susceptibles de bénéficier de l'intervention. La perspective sur la nature du problème permet également de comprendre les influences immédiates et sous-jacentes qui sont à l'origine du problème spécifié et comment l'IN proposée permet de les résoudre. Elle est aussi pertinente pour le développement que pour l'évaluation de l'intervention, car elle permet de bien comprendre le problème (Wight et al., 2016).

PHASE 2 : Évaluabilité

Avant même de démarrer l'évaluation de quelconque IN, il est fondamental de rendre compte de son évaluabilité, c'est-à-dire déterminer si l'IN peut faire l'objet d'une évaluation en bonne et due forme. L'étude d'évaluabilité est « un processus rigoureux et systématique qui permet de s'assurer qu'une évaluation satisfaisante du programme est réalisable et de déterminer si celle-ci peut produire des connaissances utiles pour les parties prenantes » (Soura et al., 2016). Ainsi, avant qu'une IN ne soit acheminée à l'équipe du CISP afin de mener une évaluation en milieu réel, l'équipe se doit de vérifier une panoplie de critères s'assurant que le projet est conforme aux exigences de l'établissement et qu'il est possible et pertinent d'entamer les démarches d'évaluation. Pour cette phase, un outil pertinent a été développé par l'équipe de Pascale Lehoux du Département de gestion, d'évaluation et de politique de santé de l'École de santé publique de l'Université de Montréal : [l'Outil In Fieri](#) de l'Innovation Responsable en Santé (IRS). Cet outil est une grille d'évaluation du degré de responsabilité de l'innovation en santé, destiné à être appliqué après le prototypage de l'innovation, suite à des études pilotes (qui démontrent la faisabilité et l'acceptabilité) et/ou à la suite d'études expérimentales ou quasi-expérimentales qui ont déjà été effectuées afin de démontrer l'efficacité de l'innovation, c'est-à-dire lorsque l'innovation a atteint un niveau de maturité de 7 ou plus selon [l'échelle de maturité d'Innovation Canada](#), lorsqu'elle est

prête à être évaluée en milieu réel, et est disponible pour usage dans les endroits où se trouvent les utilisateurs visés.

PHASE 3 : Évaluation

Dans cette phase, il est important de considérer comment maximiser l'utilité des informations pour la prise de décision. Pour cela, il faut identifier les **incertitudes** autour de l'intervention et il est incontournable d'élaborer des **questions de recherches/questions d'évaluation** pertinentes pour résoudre ces incertitudes. Il existe un grand nombre de devis d'étude, et la décision quant au modèle à utiliser doit être prise en fonction des questions de recherche posées, lesquelles doivent être définies sur la base des incertitudes qui ont été identifiées. En fonction du type d'innovation qui est évaluée, les questions d'évaluation vont généralement rejoindre les grands axes du modèle de performance du CHUM. L'objectif étant que l'innovation puisse démontrer la plus grande valeur ajoutée possible, et donc avoir un impact positif autant pour le CHUM et son personnel que pour les patients et la population de manière générale. Pour cela, les questions d'évaluation traitent souvent des catégories d'indicateurs suivantes [cette liste n'est pas exhaustive mais inspirée de Skivington et al. (2021a, 2021b) et des perspectives [RE-AIM](#)] :

- **Efficacité** : quel est l'impact global de l'intervention? Quelles composantes vont « certainement » fonctionner et quelles composantes risquent de ne pas fonctionner?
- **Portée** : qui « embarquera », qui « n'embarquera pas » et que peut-on faire pour que les gens « embarquent » mieux ?
- **Acceptabilité** : est-ce que l'intervention est perçue comme appropriée? Est-ce que certaines composantes « rebutent » les gens?
- **Implantation** : quelles composantes de l'intervention sont « livrées » telles que conceptualisées et lesquelles seront « rejetées », « incomprises » ou « laissées de côté »?
- **Pérennisation** : comment intégrer l'intervention à l'organisation pour assurer sa « survie » à travers le temps?
- **Effets inattendus** : est-ce que l'intervention pourrait avoir des retombées qui ne sont pas prévues, tant favorables de défavorables?

L'une des premières étapes fondamentales à l'élaboration du protocole d'évaluation est la **création d'un modèle logique**. Ce dernier est une représentation visuelle des ressources, activités et résultats (à court, moyen et long-termes) escomptés de l'IN, qui aide à cadrer et à mieux concevoir l'évaluation de l'IN. C'est le premier livrable produit par la plateforme pour chaque projet d'évaluation. C'est une étape fondamentale pour l'évaluation puisque le modèle logique permet de décortiquer et faire ressortir la complexité de certaines IN, avec l'interaction (souvent bidirectionnelle) de la multitude des composantes de celles-ci. Le modèle logique est ainsi très utile au développement des questions d'évaluation. De plus, il constitue également une valeur ajoutée pour les entreprises, qui voient leur IN et leur interaction avec le milieu réel du point de vue du

milieu preneur. Il permet d'illustrer comment l'IN créera de la valeur, et de s'assurer que l'équipe d'évaluation et toutes les parties prenantes ont bien compris l'IN et son fonctionnement.

- Afin de réaliser le modèle logique, voici une ressource interactive facilitante : <https://evaluationplanningtool.org/>. Il existe de nombreux outils semblables.

Identification des mesures spécifiques à utiliser et des données à collecter ou extraire de bases de données existantes pour répondre aux questions évaluatives.

Afin de percevoir plus rapidement les questions à poser et les devis de recherche requis, il existe un guide fort utile pour catégoriser les IN. En continuité avec la définition proposée plus haut (c.-à-d., tout produit ou toute technologie numérique, qui mobilise des données et qui s'appuie sur l'intelligence artificielle et la vaste gamme d'algorithmes en santé computationnelle), il peut s'agir :

- D'applications mobiles;
- De logiciels autonomes;
- D'outils en ligne pour traiter ou diagnostiquer des maladies, prévenir des problèmes de santé ou améliorer l'efficacité du système;
- De programmes pouvant être utilisés afin d'analyser des données provenant de dispositifs médicaux tels que des scanners, des capteurs ou des moniteurs.

Toutefois, il est important de noter que le cadre normatif de preuves du NICE n'est pas destiné à être utilisé afin d'évaluer les types de TNS suivants :

- Les logiciels qui font partie intégrante d'un dispositif médical ou d'un diagnostic in vitro (DIV) ou qui y sont incorporés, également appelés "*software in a medical device*" (SiMD);
- Les TNS conçues pour la formation des professionnels de la santé ou des soins (tels que la réalité virtuelle [RV] ou la réalité augmentée [RA] pour la formation chirurgicale);
- Les TNS qui facilitent la collecte de données dans le cadre d'études de recherche (NICE, 2022).

En outre, le [guide](#) du NICE propose de catégoriser les IN en fonction des niveaux de risque, A, B, ou C. Le risque le plus faible (Niveau A), est présent pour des IN qui visent à réduire les coûts ou le temps de travail du personnel ou à améliorer l'efficacité du système en place. Un risque un peu plus élevé (Niveau B) est observable lorsqu'il s'agit d'IN visant à aider les patients ou les citoyens à gérer leur propre santé ou bien être. Finalement, des risques élevés (Niveau C) sont présents lorsque les IN visent à traiter ou diagnostiquer une affection précise ou à guider les choix en matière de traitement, de diagnostic et de soins.

Une fois le niveau de risque identifié, il est ensuite nécessaire d'identifier lesquelles parmi 21 normes associées au [cadre](#) s'appliquent relativement aux facteurs de conception (9 normes), la description de la valeur (4 normes), la démonstration de performance (3 normes), l'apport de la valeur ajoutée (2 normes) et les considérations relatives au

déploiement initial (3 normes). La classification d'une IN en fonction de son niveau de risque permet d'anticiper la nature des données probantes qui seront requises pour éventuellement l'adopter.

L'identification du devis de l'étude se fait toujours en fonction des questions d'évaluation qui ont été posées, du niveau de risque (A, B, C) et des normes pertinentes. L'approche méthodologique globale peut être quantitative, qualitative et/ou mixte. Parmi les différents types de devis (voir aussi Skivington et al. 2021b), il en existe 5 principaux qui sont fréquemment utilisés en évaluation, et dont le choix dépend en grande partie des questions d'évaluation:

- Expérimental
- Quasi-expérimental
- Observationnel
- Qualitatif
- Mixte

Un descriptif des différents devis d'étude est disponible en ligne sur le [site de Sage](#). Une fois le devis choisi, il faut identifier des **indicateurs** utilisés pour répondre aux objectifs/questions d'évaluations fixés, c'est-à-dire, **choisir des mesures de résultats pertinents pour le changement souhaité**. Les indicateurs doivent être choisis en fonction des résultats ou du changement qui sont importants pour les parties prenantes, et en fonction d'une bonne compréhension de l'intervention. De plus, les mesures de résultats ne devraient pas s'en tenir qu'à saisir les changements chez les individus, mais peuvent également porter sur des mesures de procédés ou de changements dans le système. En effet, il existe entre autres deux types d'évaluation possibles : une **évaluation des résultats**, et/ou **une évaluation de processus/de la stratégie d'implantation**. Cette dernière vise à comprendre le fonctionnement de l'intervention en examinant l'implantation, les mécanismes d'impact et les facteurs contextuels (Moore et al., 2015).

Par ailleurs, lorsqu'on fait référence à la stratégie d'implantation, il peut être nécessaire **de mesurer le niveau de littéracie numérique** des utilisateurs. Des outils (traduits en français) ont été mis en place afin d'évaluer cette composante. En voici quelques-uns :

- **mHealth App Usability Questionnaire (MAUQ)** – questionnaire qui permet de mesurer la convivialité/facilité d'utilisation des applications mobiles de santé (bonnes fiabilité et la validité).
- **eHealth Literacy Assessment toolkit (eHLA)** – boîte à outils qui contient **7 échelles (composées de 5 à 23 items chacune)** qui mesurent non seulement les connaissances et les compétences des individus en matière de littéracie numérique, **mais aussi** leur niveau de littéracie en santé. Les outils de 1 à 4 mesurent la littéracie en santé, et ceux de 5 à 7 mesurent la littéracie numérique.
- **Digital Health Literacy Instrument (DHIA)** – instrument de mesure composé de **21 questions** qui permet l'auto-évaluation de la littéracie numérique en matière de santé à l'aide de plusieurs sous-échelles. Ces sous-échelles mesurent les compétences opérationnelles, les compétences de navigation et la recherche

d'informations par les utilisateurs, et aussi l'évaluation de la fiabilité, la détermination de la pertinence, l'ajout de contenu autogénéré (généré par l'utilisateur) et la protection de la vie privée.

Une fois le projet d'évaluation rendu à cette étape, il est question de vérifier sa faisabilité, à savoir si l'évaluation est viable sous tous ses différents angles. Généralement, il est pertinent de décortiquer cette faisabilité en deux grandes parties :

- La faisabilité reliée au devis d'évaluation :
 - L'obtention de l'approbation du comité d'éthique à la recherche;
 - La faisabilité du recrutement et de la rétention des participants;
 - La capacité d'atteindre une taille d'échantillon requise;
 - La faisabilité de colliger les données souhaitées et de réaliser le suivi longitudinal requis;
 - La capacité et le temps de collecte et d'analyse des données et;
 - La prise en compte des résultats non-escomptés.
- La fidélité de l'implantation de l'IN-même :
 - Les contenu et déroulement optimaux et acceptables de l'implantation de l'intervention;
 - Le niveau d'adhésion à l'intervention/acceptabilité de l'intervention;
 - La probabilité d'un bon rapport coût-efficacité et;
 - La capacité des prestataires de mettre en œuvre l'intervention tel que prévu.

La dernière étape fondamentale dans la réalisation d'une évaluation rigoureuse et pertinente est **l'identification des analyses statistiques et/ou qualitatives qui seront requises pour répondre aux questions d'évaluation**. Pour ce faire, le recours à des équipes expertes est recommandée. Au CHUM, le CISP a recours aux services de [CITADEL](#).

Finalement, une fois l'analyse des données collectées dans le cadre du projet d'évaluation complétée, il est nécessaire de **rédiger le rapport d'évaluation**. Utiliser, selon le projet d'évaluation et le type d'IN évaluée, la ou les cinq listes de vérification suivantes, disponibles dans la banque d'outils du CISP :

Tableau 2. Grille de vérification utiles en évaluation en milieux réels des IN en santé. L'intégralité de ces outils a été traduite en français. Ceux-ci sont disponibles dans la boîte à outils du CISP.

Grille de vérification en support à l'évaluation	Description de la grille	Cadre d'usage précis
DECIDE-AI	Liste de vérification comprenant les éléments clés qui devraient être rapportés dans des études cliniques de stade « précoce », pour des systèmes basés sur l'IA dans les soins de santé. La liste contient 27 items de rapport dont 17 sont spécifiques à l'IA.	Pour le rapport d'évaluation d'innovations <u>qui comprennent de l'IA.</u>
StaRI Standards for Reporting Implementation	Liste de vérification qui aide à distinguer deux aspects connexes d'une étude d'implantation : la stratégie d'implantation et l'intervention qui est mise en œuvre.	Pour le rapport d'évaluation d'innovations qui prennent en compte la stratégie d'implantation.
TIDieR Template for Intervention Description and Replication	Liste de vérification qui aide à décrire les interventions de manière suffisamment détaillée pour permettre de produire des rapports suffisamment détaillés pour que l'intervention puisse être répliquée ailleurs tout en respectant les conditions de succès.	Pour <u>tous</u> les rapports d'évaluation d'innovations.
iCheck-DH	Lignes directrices et liste de vérification pour le rapport d'implantations en santé numérique.	Pour <u>tous</u> les rapports d'évaluation d'innovations.
CHEERS Consolidated Health Economic Evaluation Reporting Standards	Liste de vérification qui soutient l'évaluation du volet économique d'une intervention.	Pour le rapport d'évaluation d'innovations qui ont un volet coût-efficacité/coût-efficience.

PHASE 4 : Implantation

Finalement, suite aux évaluations qui sont menées par Éval-IN^{CHUM}, il reste une étape subséquente fondamentale à la pérennisation des IN dans un système de santé, à savoir, leur mise à l'échelle. L'Organisation Mondiale de la santé définit l'implantation de la façon suivante : « des efforts délibérés afin d'accroître l'impact des innovations de santé testées avec succès, dans le but d'en faire bénéficier un plus grand nombre de personnes, et de favoriser l'élaboration de politiques et de programmes sur une base durable » (Meuse-Rhin & Organization, 2016).

Dans le domaine des soins de santé, la science de l'implantation est un domaine de recherche important et en constant développement, qui entreprend des études visant à

développer et évaluer des stratégies de mise en œuvre d'interventions qui sont efficaces, qui tiennent compte des réalités locales, qui peuvent être mises en œuvre à grande échelle et qui sont potentiellement durables (Damschroder et al., 2009). Ainsi, une grande partie de la recherche sur la science de l'implantation implique donc le développement et l'évaluation d'interventions complexes afin de maximiser l'implantation efficiente d'interventions, dans la pratique et/ou la politique, qui ont déjà démontré leur efficacité ou leur efficience (Skivington et al., 2021a).

LES SIX CONDITIONS FONDAMENTALES À METTRE EN PLACE DE MANIÈRE ITÉRATIVE POUR ASSURER LE SUCCÈS D'UNE ÉVALUATION

a. Considérer le contexte

- Il existe une relation bidirectionnelle entre le contexte et l'intervention elle-même. Le contexte dans lequel s'implante une intervention se superpose à sa complexité car il peut influencer la faisabilité et l'impact d'une intervention, et l'intervention peut à son tour transformer le contexte (Skivington et al., 2021a).

b. Impliquer les parties-prenantes

- Impliquer dès le départ toutes les personnes/acteurs impliqués dans l'implantation des changements ou qui seront affectés par ceux-ci.
- Encourager la concertation de ces personnes/acteurs tout au long du processus, de la construction du protocole à la rédaction finale du rapport d'évaluation.
- Les parties prenantes incluent généralement :
 - Les personnes qui auront à changer leurs pratiques ou adopter de nouvelles technologies;
 - Les personnes qui bénéficieront de l'innovation (patients, citoyen, populations, professionnels de la santé, décideurs, etc.);
 - Les personnes qui font la conception et l'implantation de l'innovation et;
 - Les personnes qui évalueront l'innovation (Skivington et al., 2021a).

c. Développer et raffiner la théorie du programme

- La théorie du programme décrit comment une intervention est censée mener aux effets escomptés et sous quelles conditions. Elle comprend généralement : (1) les composantes clés de l'intervention et comment celles-ci interagissent ensemble; (2) les mécanismes de l'intervention; (3) les paramètres du contexte dont on anticipe une influence sur ces mécanismes; et (4) comment ces mécanismes vont influencer le contexte (Funnell & Rogers, 2011).
- Grâce au modèle logique causal développé dans les premières étapes du processus d'évaluation (incluant les intrants, activités, extrants, résultats et impacts, et d'autres composantes de la théorie du programme incluant les mécanismes par lesquels l'intervention est censée générer les résultats, ainsi que les facteurs contextuels et autres conditions préalables sur lesquelles elle repose), il est plus

facile de raffiner davantage la théorie du programme en fonction des avancées dans le projet.

d. Identifier les incertitudes clés

- Il est question ici d'identifier les incertitudes qui existent compte tenu de ce qu'on sait déjà et de ce que la théorie du programme, l'équipe d'évaluation et les parties prenantes considèrent comme les plus importantes. Ces jugements éclairent la formulation et la hiérarchisation des questions d'évaluation, et devraient permettre d'obtenir des données plus utiles pour la prise de décision (Skivington et al., 2021a).

e. Raffiner l'intervention

- Il s'agit du processus de "raffinement" ou de modification de l'intervention une fois qu'une version préliminaire (prototype) a été développée (p. ex., raffinement du protocole d'évaluation suite à l'ajustement des questions d'évaluation). Il est suggéré que le perfectionnement continu d'une intervention, en accord avec la théorie du programme, peut améliorer l'applicabilité potentielle de l'intervention (Skivington et al., 2021a).

f. Considérations économiques

- Il est important de prendre en compte les conséquences comparatives des interventions sur les ressources et les résultats pour les personnes et les organisations concernées. L'accent est davantage mis sur ces aspects car ils sont considérés comme un élément clé à toutes les phases d'un projet de recherche (et donc d'évaluation), plutôt que comme un simple ensemble de méthodes d'évaluation du rapport coût-efficacité (Skivington et al., 2021a).



Tous ces 6 éléments doivent être pris en compte dans chacune des phases de l'évaluation (section II), et ce de manière **itérative**.

III. OUTILS ET RESSOURCES EXTERNES

Voici une liste d'outils supplémentaires qui sont également pertinents pour supporter l'évaluation en milieu réel :

- Le fichier Excel pour évaluer l'impact budgétaire de l'IN (NICE)
- Le fichier Excel qui présente des exemples de classifications d'IN (NICE)
- **Le Site ICHOM** (pour identifier des mesures des retombées de santé centrées sur le patient) : <https://www.ichom.org/patient-centered-outcome-measures/>
- **Site RE-AIM** : <https://re-aim.org/assessments/>

IV. CAS ILLUSTRATIFS

La page web du NICE, qui décrit le cadre normatif du NICE pour les technologies numériques en santé, inclut un fichier Excel qui offre de nombreux exemples.

<https://www.nice.org.uk/corporate/ecd7/resources>

V. LES QUATRE PERSPECTIVES DE RECHERCHE À ADOPTER POUR DES INTERVENTIONS COMPLEXES

Jusqu'à maintenant, les pratiques en évaluation étaient surtout centrées sur les questions d'efficacité, c'est-à-dire sur le niveau de résultats escomptés produits par les interventions dans un contexte expérimental ou idéal. Pourtant, il est primordial de considérer et d'utiliser diverses perspectives de recherche, ainsi que d'adopter une vision pluraliste et pragmatique dans le choix des questions de recherche et des méthodes utilisées (Skivington et al., 2021a), notamment en évaluation en milieux réels où les conditions et le contexte ne sont pas toujours idéaux. À cette (a) perspective d'efficacité devrait s'ajouter une (b) perspective d'efficacité, une (c) perspective rattachée à la théorie du programme, ainsi qu'une (d) perspective du système dans lequel l'intervention est évaluée.

1. Perspective d'efficacité – Dans quelle mesure l'intervention produit-elle le(s) résultat(s) escompté(s) dans un contexte expérimental ou idéal?

En pratique, les projets d'évaluation qui adoptent une perspective d'efficacité ou d'efficacité uniquement, visent notamment à obtenir des estimations impartiales de l'effet moyen des interventions sur des résultats prédéterminés (Skivington et al., 2021a). La recherche qui s'inscrit dans cette perspective a pour objectif de tester des hypothèses sur les mécanismes d'action des interventions (Zwarenstein, 2017). Elle vise entre autres une validité interne élevée, en adoptant une approche explicative pour tester des hypothèses causales sur le(s) résultat(s) généré(s) par l'intervention. Les projets qui s'inscrivent dans une perspective d'efficacité évaluent généralement les interventions dans des conditions idéales et contrôlées et au sein d'un groupe homogène de participants. Ainsi, les résultats d'une étude d'efficacité indiquent si une intervention peut fonctionner, ou ne pas fonctionner, dans des conditions idéales. En règle générale, les études d'efficacité contrôlent les variations contextuelles, de sorte qu'elles ne contribuent pas à mettre de l'avant l'influence du contexte sur l'intervention, ce qui est fondamental lors de l'évaluation d'innovations en milieu réel.

2. Perspective d'efficacité – Dans quelle mesure l'intervention produit-elle le(s) résultat(s) escompté(s) dans un contexte réel?

Les études qui adoptent une perspective d'efficacité visent à répondre à des questions pragmatiques sur les effets produits par les interventions dans la pratique courante afin d'éclairer directement le choix entre les options d'intervention (Zwarenstein, 2017). Ces études cherchent à tester une intervention dans des échantillons et des contextes représentatifs de ceux dans lesquels l'intervention serait mise en œuvre dans la pratique quotidienne, c'est-à-dire dans un milieu réel, généralement avec un résultat en matière de santé ou de services de santé. Une certaine flexibilité dans la mise en œuvre des interventions et l'adhésion à celles-ci peut être autorisée afin de permettre des variations dans la manière, le lieu et les personnes qui mettent en œuvre et/ou reçoivent les interventions (Skivington et al., 2021a).

3. Perspective réaliste sur l'intervention – Qu'est-ce qui fonctionne dans quelles circonstances et comment?

L'idée principale derrière l'adoption d'une perspective de la théorie du programme est de fournir des données probantes sur les processus par lesquels les interventions conduisent à un changement des résultats, ainsi que sur les conditions préalables qui peuvent être requises pour que ce changement se produise, ce qui permet ainsi d'explorer comment et pourquoi elles entraînent un changement (Skivington et al., 2021a). Cette perspective permet d'explorer les interventions dans la pratique, en tenant compte du contexte, et prend souvent en considération plus qu'une hypothèse sur la façon dont l'intervention peut fonctionner. Ainsi, adopter cette approche à l'évaluation des innovations vise à élargir le champ de l'évaluation pour comprendre comment une innovation peut fonctionner, et comment cela peut varier dans différents contextes ou pour différents individus (Astbury & Leeuw, 2010). En effet, les innovations sont évaluées dans le cadre d'un processus continu d'élaboration, d'essai et d'ajustement de la théorie du programme. Ainsi, les évaluations menées dans cette perspective peuvent permettre de comprendre comment les mécanismes et le contexte interagissent ensemble, en fournissant des preuves qui peuvent permettre aux innovations d'être potentiellement appliquées dans d'autres contextes (Skivington et al., 2021a).

4. Perspective de système – Comment le système et l'intervention s'adaptent l'un à l'autre?

Cette perspective suggère que les interventions peuvent être mieux comprises en examinant le(s) système(s) dans le(s)quel(s) elles s'inscrivent, ou les systèmes qu'elles visent à modifier. L'idée principale derrière l'adoption d'une perspective de système est de considérer les relations entre l'intervention et le contexte, de s'engager dans multiples perspectives, et d'identifier les limites du système (Williams, 2015). Cette perspective encourage donc les chercheurs et évaluateurs à considérer comment l'intervention peut être influencée par de nombreux éléments du système, et également comment elle peut avoir un impact sur ces éléments à son tour, et ce, sur une période de temps prolongée. Toutefois, il n'est généralement pas possible de tenir compte de l'ensemble d'un système, qui est souvent ouvert. Par conséquent, afin de rendre une évaluation réalisable, il est nécessaire d'en limiter la portée en déterminant les limites du système, et en restreignant l'éventail des mécanismes ou des explications potentiels qui sont étudiés (Skivington et al., 2021a).

VI. RÉFÉRENCES

- Alami, H., Lehoux, P., Miller, F. A., Shaw, S. E., & Fortin, J. P. (2023). An urgent call for the environmental sustainability of health systems: A 'sextuple aim' to care for patients, costs, providers, population equity and the planet. *Int J Health Plann Manage*, 38(2), 289-295. <https://doi.org/10.1002/hpm.3616>
- Astbury, B., & Leeuw, F. L. (2010). Unpacking black boxes: mechanisms and theory building in evaluation. *American journal of evaluation*, 31(3), 363-381.
- Bodenheimer, T., & Sinsky, C. (2014). From triple to quadruple aim: care of the patient requires care of the provider. *Ann Fam Med*, 12(6), 573-576. <https://doi.org/10.1370/afm.1713>
- Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation Science*, 4(1), 50. <https://doi.org/10.1186/1748-5908-4-50>
- FDA. (2019). Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD). In.
- Funnell, S. C., & Rogers, P. J. (2011). *Purposeful program theory: Effective use of theories of change and logic models*. John Wiley & Sons.
- Itchhaporia, D. (2021). The Evolution of the Quintuple Aim: Health Equity, Health Outcomes, and the Economy. *J Am Coll Cardiol*, 78(22), 2262-2264. <https://doi.org/10.1016/j.jacc.2021.10.018>
- Meuse-Rhin, E., & Organization, W. H. (2016). Scaling up projects and initiatives for better health: from concepts to practice.
- Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D., & Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *Bmj*, 350, h1258. <https://doi.org/10.1136/bmj.h1258>
- NICE. (2022). Evidence standards framework for digital health technologies. In.
- Poth, C., Lamarche, M. K., Yapp, A., Sulla, E., & Chisamore, C. (2014). Toward a Definition of Evaluation Within the Canadian Context: Who Knew This Would Be So Difficult? *Canadian Journal of Program Evaluation*, 29(1), 87-103. <https://doi.org/10.3138/cjpe.29.1.87>
- Sikka, R., Morath, J. M., & Leape, L. (2015). The Quadruple Aim: care, health, cost and meaning in work. *BMJ Quality & Safety*, 24(10), 608-610. <https://doi.org/10.1136/bmjqs-2015-004160>
- Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., Boyd, K. A., Craig, N., French, D. P., McIntosh, E., Petticrew, M., Rycroft-Malone, J., White, M., & Moore, L. (2021a). Framework for the development and evaluation of complex interventions: gap analysis, workshop and consultation-informed update. *Health Technol Assess*, 25(57), 1-132. <https://doi.org/10.3310/hta25570>
- Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., Boyd, K. A., Craig, N., French, D. P., McIntosh, E., Petticrew, M., Rycroft-Malone, J., White, M.,

- & Moore, L. (2021b). A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *BMJ*, 374, n2061. <https://doi.org/10.1136/bmj.n2061>
- Soura, B., Dagenais, C., Bastien, R., Fallu, J.-S., & Janosz, M. (2016). L'étude d'évaluabilité : Utilité et pertinence pour l'évaluation de programme [Evaluability Assessment: Usefulness and Relevance for Program Evaluation]. *Canadian Journal of Program Evaluation*, 31, 18-33. <https://doi.org/10.3138/cjpe.207>
- Wight, D., Wimbush, E., Jepson, R., & Doi, L. (2016). Six steps in quality intervention development (6SQulD). *Journal of Epidemiology and Community Health*, 70(5), 520-525. <https://doi.org/10.1136/jech-2015-205952>
- Williams, B. (2015). Prosaic or profound? The adoption of systems ideas by impact evaluation. *IDS Bulletin*, 46(1), 7-16.
- Zwarenstein, M. (2017). 'Pragmatic' and 'explanatory' attitudes to randomised trials. *Journal of the Royal Society of Medicine*, 110(5), 208-218. <https://doi.org/10.1177/0141076817706303>