

Université de Montréal

**Méthodes de collecte de données pharmacoéconomiques en pratique réelle**

par

Véronique Lambert-Obry

Médicament et Santé des Populations

Faculté de Pharmacie

Thèse présentée à la Faculté des études supérieures en vue de l'obtention du grade de Docteur  
en philosophie (Ph.D.) en Sciences Pharmaceutiques option Médicament et Santé des  
Populations

Octobre, 2022

© Véronique Lambert-Obry, 2022

Université de Montréal  
Faculté des études supérieures et postdoctorales

Cette thèse intitulée :  
Méthodes de collecte de données pharmacoéconomiques en pratique réelle

présentée par  
Véronique Lambert-Obry

a été évaluée par un jury composé des personnes suivantes :

Isabelle Chabot  
Président-rapporteur

Jean Lachaine  
Directeur de recherche

Jason Robert Guertin  
Membre du jury

Laurent Azoulay  
Examineur externe

## RÉSUMÉ

---

Dans un système de santé avec budget limité, l'efficacité des différentes options thérapeutiques doit être comparée afin d'assurer une allocation optimale des ressources. À cette fin, l'analyse coût-utilité (ACU) est privilégiée par les agences d'évaluation des technologies de la santé et requiert des données pharmacoéconomiques telles que les coûts et les scores d'utilité. Toutefois, les données pharmacoéconomiques ne sont pas systématiquement collectées lors des essais contrôlés randomisés (ECR). Lorsqu'elles le sont, elles sont soumises aux limites des ECR affectant ainsi la validité externe des résultats. Cela a mené à l'émergence des études en pratique réelle (*real-world evidence studies*) comme évidence complémentaire aux ECR aux fins de remboursement. Les données en pratique réelle peuvent être des données primaires (recueillies spécifiquement pour répondre à un objectif particulier) ou des données secondaires (données déjà collectées pour répondre à d'autres objectifs). D'importantes données pharmacoéconomiques, telles que l'utilité, la perte de productivité et les coûts non médicaux, étant généralement manquantes dans les banques de données traditionnelles, les chercheurs en pharmacoéconomie doivent parfois générer leurs propres données. Néanmoins, il n'existe pas de lignes directrices spécifiques pour guider les chercheurs désirant générer des données pharmacoéconomiques en pratique réelle.

Le but de cette thèse était d'étudier et contribuer à l'amélioration des pratiques méthodologiques pour la génération de données pharmacoéconomiques en pratique réelle. Précisément, l'objectif du volet 1 était donc de caractériser les pratiques méthodologiques des études collectant des données de pharmacoéconomie et d'évaluer la qualité des résultats obtenus. Une revue systématique de la littérature a permis de confirmer certaines craintes face aux études de pratique réelle, c'est-à-dire le manque de transparence et de crédibilité. Le volet 1 a mis de l'avant le besoin d'améliorer les méthodologies actuelles qui ne produisent pas toujours des résultats robustes, pouvant ainsi affecter la validité et la pertinence des évaluations économiques. Un cadre de recherche spécifique à la collecte de données pharmacoéconomiques en pratique réelle pourrait augmenter la qualité des résultats générés. L'objectif du volet 2 était donc d'identifier les

lignes directrices spécifiques à la génération de données d'utilité. Ensuite, d'incorporer les bonnes pratiques méthodologiques dans une boîte à outils destinée spécifiquement aux investigateurs cherchant à générer des valeurs d'utilité en pratique réelle. Bien que cette thèse englobe plus largement l'ensemble des données pharmacoéconomiques, telles que les coûts et l'utilité, la boîte à outils se concentre sur les scores d'utilité uniquement. Il s'agit d'un premier pas vers des lignes directrices officielles fournissant des bonnes pratiques méthodologiques pour la conception et conduite d'études en pharmacoéconomie. Pour ce qui est du volet 3, l'objectif était de mettre en pratique la boîte à outils en rédigeant un protocole de recherche de qualité, pertinent et applicable à la réalité des chercheurs en pratique réelle. Précisément, d'utiliser les recommandations de la boîte à outils afin de mettre en place un devis d'étude et des procédures d'étude qui limitent le risque de biais. Cette thèse contribue à l'avancement des bonnes pratiques et sensibilise la communauté de chercheurs à viser un niveau hiérarchique d'évidence plus élevé. La standardisation des méthodologies pour générer des données pharmacoéconomiques en pratique réelle est un besoin qui demeure non comblé.

**Mots-clés** : année de vie ajustée par la qualité, qualité de vie reliée à la santé, utilité, coûts, productivité, pharmacoéconomie, économie de la santé, pratique réelle, recherche observationnelle, biais.

## ABSTRACT

---

In a resource-constrained healthcare system, the effectiveness of various therapeutic interventions must be compared by the means of economic evaluations to ensure optimal resource allocation. The cost-utility analysis (CUA) is favored by health technology assessment (HTA) agencies, and requires health economics and outcomes research (HEOR) data such as costs and utilities. Although they have gained prominence, HEOR data are not systematically collected in randomized controlled trials (RCTs). Even if they were, RCTs are well-known to have inherent limitations leading to a low external validity. Consequently, real-world evidence (RWE) studies have been used as complementary evidence in reimbursement decision-making. Real-world data can be primary data (specifically collected to achieve a certain objective) or secondary data (data that have already been collected for other purposes). However, important HEOR data, such as utility scores, productivity and non-medical costs, are not routinely collected in traditional databases. Therefore, HEOR investigators may need to generate their own pharmacoeconomic data. Nevertheless, there is no single guideline for investigators seeking guidance on methodological steps to estimate HEOR data in a real-world setting.

The goal of this thesis was to study and improve good practices for HEOR data generation in a real-world setting. Specifically, the objective of the first article was to characterize methodological practices for collecting HEOR data as well as critically appraise the quality of the results. The systematic literature review confirmed concerns about RWE studies, namely the lack of transparency and credibility. The first article highlighted the need for good practices as current methodologies may not generate robust estimates, thereby affecting the validity and relevance of economic evaluations. A research framework specifically designed for HEOR data collection in a real-world setting could help improve the quality of results. Thus, the objective of the second article was to review current recommendations for health state utility (HSU) generation, and incorporate methodological standards into a single toolbox intended for investigators seeking to collect HSU in a real-world setting. Although this thesis broadly encompasses HEOR data, such as costs and utilities, the toolbox focuses on utility scores only. This is a first step toward official guidelines providing good practice recommendations for designing and conducting HEOR studies. As for the third article, the objective was to use the toolbox to develop a study protocol that is

relevant, realistic and of good methodological quality. The recommendations of the toolbox were followed to design the study and limit bias. This thesis provides good practices to help generate high quality HEOR estimates, and motivate investigators to improve the level of evidence to be generated. Standardization of RWE studies collecting HSU has yet to come.

**Keywords:** Quality-adjusted life years, health-related quality of life, utility, costs, productivity, pharmacoeconomics, health economics, real-world evidence, observational research, bias.

## TABLE DES MATIÈRES

---

<b>RÉSUMÉ</b> .....	<b>3</b>
<b>ABSTRACT</b> .....	<b>5</b>
<b>TABLE DES MATIÈRES</b> .....	<b>7</b>
<b>LISTE DES TABLEAUX</b> .....	<b>12</b>
<b>LISTE DES FIGURES</b> .....	<b>13</b>
<b>LISTE DES ABRÉVIATIONS</b> .....	<b>14</b>
<b>REMERCIEMENTS</b> .....	<b>16</b>
<b>CHAPITRE 1 : INTRODUCTION</b> .....	<b>17</b>
<b>CHAPITRE 2 : ÉVALUATIONS ÉCONOMIQUES</b> .....	<b>21</b>
<b>2.1 Évaluations économiques</b> .....	<b>21</b>
<b>CHAPITRE 3 : BIAIS DANS LES ÉTUDES EN PRATIQUE RÉELLE</b> .....	<b>41</b>
<b>3.1 Biais de sélection</b> .....	<b>41</b>
<b>3.2 Biais d'information</b> .....	<b>45</b>
<b>3.3 Biais de confusion</b> .....	<b>47</b>
<b>CHAPITRE 4 : EXEMPLE DE CAS – DIABÈTE DE TYPE 2</b> .....	<b>55</b>
<b>4.1 Prise en charge du diabète de type 2</b> .....	<b>55</b>
<b>4.2 Modélisation économique en diabète de type 2</b> .....	<b>59</b>
<b>4.3 Impact de l'hypoglycémie sur l'utilité et la productivité</b> .....	<b>64</b>
<b>4.3.1 Impact de l'hypoglycémie sur les scores d'utilité : état des connaissances</b> .....	<b>64</b>

4.3.2	Impact de l'hypoglycémie sur les scores d'utilité : facteurs confondants.....	70
4.3.3	Impact de l'hypoglycémie sur la productivité : état des connaissances .....	73
4.3.4	Impact de l'hypoglycémie sur la productivité : facteurs confondants.....	78
<b>CHAPITRE 5 : OBJECTIFS.....</b>		<b>81</b>
<b>5.1</b>	<b>Article 1 – Revue des études en pratique réelle en diabète de type 2 : manque de bonnes pratiques</b>	<b>82</b>
5.1.1	Objectif général .....	82
5.1.2	Objectifs spécifiques .....	82
<b>5.2</b>	<b>Article 2 – Études en pratique réelle : une boîte à outils pour la collecte de données d'utilité</b>	<b>83</b>
5.2.1	Objectif général .....	83
5.2.2	Objectifs spécifiques .....	83
<b>5.3</b>	<b>Article 3 – Impact de l'hypoglycémie sur la perte de productivité et l'utilité des patients atteints de diabète de type 2 et traités à l'insuline en pratique réelle canadienne : un protocole d'étude</b>	<b>84</b>
5.3.1	Objectif général .....	84
5.3.2	Objectifs spécifiques .....	84
<b>CHAPITRE 6 : MÉTHODOLOGIE.....</b>		<b>85</b>
<b>6.1</b>	<b>Article 1 – Revue des études en pratique réelle en diabète de type 2 : manque de bonnes pratiques</b>	<b>85</b>
6.1.1	Revue de la littérature et sélection des études .....	85
6.1.2	Extraction des données .....	87
6.1.3	Évaluation de la qualité .....	88



**6.2 Article 2 – Études en pratique réelle : une boîte à outils pour la collecte de données d'utilité 90**

**6.3 Article 3 – Impact de l'hypoglycémie sur la perte de productivité et l'utilité des patients atteints de diabète de type de 2 et traités à l'insuline en pratique réelle canadienne : un protocole d'étude 101**

6.3.1	Devis et population à l'étude .....	102
6.3.2	Mesure de l'exposition : hypoglycémie .....	103
6.3.3	Mesure des issues pharmacoéconomiques : utilité et perte de productivité .....	104

**CHAPITRE 7 : ARTICLE 1 – revue des études de pratique réelle en diabète de type 2 : manque de bonnes pratiques ..... 106**

<b>7.1</b>	<b>Abstract .....</b>	<b>106</b>
<b>7.2</b>	<b>Introduction .....</b>	<b>107</b>
<b>7.3</b>	<b>Methods .....</b>	<b>108</b>
7.3.1	Literature Search and Study Selection .....	108
7.3.2	Data Extraction.....	110
7.3.3	Quality Assessment .....	110
7.3.4	Results .....	111
7.3.5	Discussion.....	119
7.3.6	Conclusion.....	122
7.3.7	References .....	123
7.3.8	Supplementary material.....	126

**CHAPITRE 8 : ARTICLE 2 – Études en pratique réelle : une boîte à outils pour la collecte de données d'utilité ..... 147**

<b>8.1</b>	<b>Abstract .....</b>	<b>147</b>
<b>8.2</b>	<b>Introduction .....</b>	<b>147</b>
<b>8.3</b>	<b>Methods .....</b>	<b>148</b>
<b>8.4</b>	<b>Results.....</b>	<b>150</b>
8.4.1	Confounding: Study design and analytical techniques.....	152
8.4.2	Selection bias: Study sample .....	154
8.4.3	Selection bias: Missing data .....	155
8.4.4	Information bias: Exposure and outcome measurement .....	156
<b>8.5</b>	<b>Discussion .....</b>	<b>157</b>
<b>8.6</b>	<b>Conclusion .....</b>	<b>158</b>
<b>8.7</b>	<b>Future perspective .....</b>	<b>158</b>
<b>8.8</b>	<b>Executive summary .....</b>	<b>159</b>
<b>8.9</b>	<b>References .....</b>	<b>160</b>

**CHAPITRE 9 : ARTICLE 3 – Impact de l’hypoglycémie sur la perte de productivité et l’utilité des patients atteints du diabète de type 2 et traités à l’insuline en pratique réelle canadienne : un protocole d’étude prospective .....** **166**

<b>9.1</b>	<b>Abstract .....</b>	<b>166</b>
<b>9.2</b>	<b>Introduction .....</b>	<b>167</b>
9.2.1	Current state of knowledge.....	168
9.2.2	Study rationale and relevance.....	169
<b>9.3</b>	<b>Methods .....</b>	<b>170</b>
9.3.1	Research purpose and study design.....	170

9.3.2	Study population.....	170
9.3.3	Definition of hypoglycemia.....	171
9.3.4	Measurement of outcomes.....	172
9.3.5	Study procedures .....	173
9.3.6	Ethical considerations.....	174
9.3.7	Statistical analyses.....	175
<b>9.4</b>	<b>Results.....</b>	<b>180</b>
<b>9.5</b>	<b>Discussion .....</b>	<b>180</b>
<b>9.6</b>	<b>References .....</b>	<b>184</b>
<b>CHAPITRE 10 : DISCUSSION.....</b>		<b>192</b>
<b>CHAPITRE 11 : CONTRIBUTIONS ET PERSPECTIVES .....</b>		<b>200</b>
<b>CHAPITRE 12 : CONCLUSION .....</b>		<b>203</b>
<b>BIBLIOGRAPHIE.....</b>		<b>204</b>
<b>Annexe 1. Stratégie de recherche en diabète de type 2.....</b>		<b>220</b>

## **LISTE DES TABLEAUX**

---

Tableau 1. Méthodes de mesure directe de l'utilité .....	27
Tableau 2. Méthodes de mesure indirecte de l'utilité .....	29
Tableau 3. Instruments génériques de mesure de la productivité .....	33
Tableau 4. Évaluation des instruments génériques de mesure de la productivité aux fins d'évaluations économiques.....	37
Tableau 5. Essais cliniques évaluant l'effet d'un contrôle glycémique intensif.....	58
Tableau 6. Modèles économiques fréquemment utilisés en DMT2 .....	62
Tableau 7. Données canadiennes d'utilité associées à l'hypoglycémie.....	66
Tableau 8. Impact de l'hypoglycémie sur l'utilité : source privilégiée par l'ACMTS .....	68
Tableau 9. Facteurs confondants potentiels pour les valeurs d'utilité .....	71
Tableau 10. Données canadiennes de productivité associées à l'hypoglycémie .....	75
Tableau 11. Facteurs confondants potentiels pour les valeurs de perte de productivité.....	79
Tableau 12. Recommandations pour éviter les biais dans la recherche en pratique réelle.....	91
Tableau 13. Recommandations non publiées par la boîte à outils.....	99

## LISTE DES FIGURES

---

Figure 1. Représentation de la relation structure des covariables.....	48
Figure 2. Représentation du confondant dépendant du temps .....	50
Figure 3. Exemple de modèle économique en diabète de type 2.....	60
Figure 4. Sélection des études sur les scores d'utilité associés à l'hypoglycémie.....	65
Figure 5. Sélection des études sur la perte de productivité associée à l'hypoglycémie.....	74
Figure 6. Stratégie de recherche initiale (en vert) et stratégie finale (en bleu).....	97

## LISTE DES ABRÉVIATIONS

---

ABCD : *Alberta's Caring for Diabetes*

ACMTS : Agence canadienne des médicaments et des technologies de la santé

ACU : Analyse coût-utilité

DMT2 : Diabète de type 2

ECR : Essai contrôlé randomisé

EQ-5D-5L : *EuroQol 5 dimensions 5 levels*

GEE : *Generalized estimating equations*

HAT : *Hypoglycemia Assessment Tool*

HLQ : *Health and Labor Questionnaire*

ICER : *Incremental cost-effectiveness ratio* ou rapport coût-efficacité différentiel

iMTA : *Institute for Medical Technology Assessment*

INESSS : Institut national d'excellence en santé et en services sociaux

iPCQ : *iMTA Productivity Cost Questionnaire*

ISPOR : *International Society for Pharmacoeconomics and Outcomes Research*

JBI : *Joanna Briggs Institute*

NICE : *National Institute for Health and Care Excellence*

PRODISQ : *Productivity and Disease Questionnaire*

QALY : *Quality-adjusted life years* ou années de vie ajustée en fonction de la qualité

QQ : *Quantity and Quality*

ROBINS-I : *Risk Of Bias In Non-randomized Studies - of Interventions*

UKPDS : *United Kingdom Prospective Diabetes Study*

WPAI : *Work Productivity and Activity Impairment*

*À ceux qui m'ont soutenu dans cette aventure,*

## REMERCIEMENTS

---

J'aimerais remercier mon directeur de recherche, Jean Lachaine, qui a rendu possible la réalisation de ce projet. Merci pour ton précieux appui et constant soutien. Merci de m'avoir encouragé dans cette aventure qui m'a permis de repousser mes limites.

Je voudrais également remercier les membres de mon comité aviseur, Michelle Savoie et Jean-Philippe Lafrance, pour leur importante collaboration tout au long de mon projet. J'aimerais aussi remercier mes collègues qui m'ont aidé et supporté dans les moments décisifs.

J'aimerais dédier un merci spécial à mon époux, Bruno Payette, qui me permet de me surpasser. J'aimerais également remercier mes deux garçons, Eric et Philip, qui sont une source de motivation pour accomplir de grandes choses. J'aimerais remercier mes parents pour leur soutien inconditionnel.

Merci à tous mes proches pour leurs encouragements.



## CHAPITRE 1 : INTRODUCTION

---

Les agences réglementaires telles que Santé Canada évaluent la valeur thérapeutique des interventions en santé afin d'autoriser leur mise en marché.(1) L'efficacité des nouveaux médicaments est communément démontrée lors d'études cliniques, où l'essai contrôlé randomisé (ECR) permet d'obtenir un haut niveau hiérarchique d'évidence.(2) Quant aux agences d'évaluation des technologies de la santé, notamment l'Institut national d'excellence en santé et en services sociaux (INESSS) et l'Agence canadienne des médicaments et des technologies de la santé (ACMTS), elles s'intéressent non seulement à la valeur thérapeutique des interventions, mais également à leur valeur économique, notamment à leur efficacité.(3, 4) L'efficacité d'une intervention est déterminée par l'entremise des évaluations économiques qui sont essentielles à favoriser l'usage optimal des ressources dans le système de santé canadien. L'efficacité d'un médicament est définie comme étant le rapport de son coût différentiel et de ses résultats cliniques différentiels avec les autres options à même visée thérapeutique.(4)

Il existe plusieurs types d'évaluations économiques, mais l'analyse coût-utilité (ACU) est privilégiée par les agences d'évaluation des technologies de la santé.(3, 4) En plus des données cliniques, l'ACU requiert également des données pharmacoéconomiques, telles que les coûts et les scores d'utilité. Toutefois, les données pharmacoéconomiques ne sont pas systématiquement collectées lors des ECR. Lorsqu'elles le sont, elles sont soumises aux limites des ECR, telles qu'une sélection pointue des patients et un suivi médical strict, affectant ainsi la validité externe des résultats. Cela a favorisé l'émergence des études en pratique réelle (*real-world evidence studies*) comme évidence complémentaire aux ECR afin de mieux guider les décisions relatives au remboursement des médicaments.(5)

La pratique réelle peut être définie de différentes façons.(6) Certaines définitions sont plus restrictives et se limitent aux études qui n'interfèrent aucunement avec l'assignation de traitement, le suivi des patients et la sélection des participants. Quelques définitions se restreignent à

l'utilisation de données existantes uniquement. La définition la plus communément utilisée est beaucoup plus générale et inclut l'ensemble des études qui ne sont pas un ECR traditionnel.(5, 6) Selon cette dernière définition, les données en pratique réelle peuvent être des données primaires (recueillies spécifiquement pour répondre à un objectif particulier) ou des données secondaires (données déjà collectées pour répondre à d'autres objectifs). Les sources de données secondaires telles que les banques de données administratives, les dossiers médicaux électroniques et les registres regorgent de données cliniques essentielles à l'étude de l'efficacité et de l'innocuité des interventions en pratique réelle d'utilisation.(5, 6) Toutefois, d'importantes données pharmacoéconomiques, telles que les scores d'utilité, les pertes de productivité et les coûts non médicaux, sont généralement manquantes dans les sources de données secondaires traditionnelles. Donc, les chercheurs en pharmacoéconomie doivent parfois générer leurs propres données.

Contrairement aux ECR, la conduite d'études en pratique réelle est beaucoup moins standardisée et plus à risque de biais. En effet, certaines caractéristiques typiques des études en pratique réelle, telles que l'absence de randomisation et un suivi sans double insu, peuvent affecter la validité des résultats obtenus. D'ailleurs, de nombreux ouvrages ont été publiés au cours des dernières années afin d'augmenter la qualité méthodologique et d'accroître la justesse des données générées. Malgré la pertinence des recommandations actuelles, il n'existe aucune ligne directrice spécifique à la génération de données pharmacoéconomiques en pratique réelle. L'absence de standards méthodologiques précis peut laisser place à un manque de rigueur dans la conduite d'études, ayant de ce fait un impact direct sur la crédibilité des résultats obtenus. L'utilisation de données de faible qualité dans une évaluation économique peut biaiser l'interprétation des résultats et mener à une mauvaise conclusion de la part des payeurs. L'impact d'une prise de décision biaisée peut aller dans deux directions opposées, c'est-à-dire mener au remboursement d'une intervention non efficiente ou empêcher le remboursement d'une intervention efficiente. Le remboursement d'une intervention non efficiente résulte en la mauvaise allocation des ressources du système de santé, diminuant le budget destiné aux interventions avec une valeur économique réelle. À l'inverse, une recommandation négative de remboursement limite l'accès des patients à une intervention efficiente, affectant les soins des patients atteints d'une certaine condition

médicale. Dans tous les cas, une prise de décision biaisée a des conséquences négatives sur la qualité du système de soins.

Cette préoccupation a donc mené au premier volet de cette thèse qui vise à caractériser, au moyen d'une revue de la littérature, les pratiques méthodologiques des études en pratique réelle générant des données pharmacoéconomiques. À cet effet, le diabète de type 2 (DMT2) a été choisi comme exemple de champ thérapeutique étant donné sa grande prévalence et son impact significatif sur le système de santé. Au Canada, presque une personne sur trois vit avec le diabète ou le prédiabète, imposant un fardeau en soins de santé estimé à 3,8 milliards de dollars en 2020.(7) Le diabète contribue à 30% des cas d'accidents vasculaires cérébraux, 40% des crises cardiaques, 50% des insuffisances rénales terminales, 70% des amputations non traumatiques des membres inférieurs et est la principale cause de cécité.(7)

À la suite de la revue des méthodologies des études en pratique réelle en pharmacoéconomie, le deuxième volet de cette thèse porte sur l'élaboration d'une boîte à outils visant à améliorer les bonnes pratiques. La boîte à outils est spécifiquement destinée aux chercheurs qui désirent entreprendre une étude générant des données pharmacoéconomiques, plus précisément l'utilité.

À partir de la boîte à outils, le troisième volet consiste en la rédaction d'un protocole de recherche suivant les recommandations méthodologiques. Le projet est réalisé en DMT2, plus spécifiquement sur les effets de l'hypoglycémie chez patients traités à l'insuline. L'hypoglycémie est un fardeau chez ces patients, donc le troisième volet porte sur l'impact des événements hypoglycémiques sur l'utilité et la perte de productivité.

Afin de bien comprendre les différentes notions abordées dans cette thèse, les trois prochains chapitres discutent plus en détail des évaluations économiques (chapitre 2), des biais

dans les études en pratique réelle (chapitre 3) et du DMT2 (chapitre 4). Le chapitre 2 présente les différents types d'évaluations économiques ainsi que leurs composantes. Bien que la présente thèse ne vise pas la conduite d'une évaluation économique en tant que tel, elle se concentre sur la collecte de données pouvant servir à sa réalisation. Ainsi, les données pharmacoéconomiques visées dans le cadre de cette thèse sont décrites plus en détail, notamment comment elles sont définies, mesurées et utilisées dans une évaluation économique. La validité des données pharmacoéconomiques a un impact direct sur la qualité des résultats d'une évaluation économique. Lorsque les données pharmacoéconomiques sont collectées dans des études en pratique réelle, de nombreux biais peuvent affecter la validité interne des résultats. Ceci nous amène donc au chapitre 3 qui décrit les principaux biais pouvant affecter la justesse des données collectées en pratique réelle. Par la suite, afin de se familiariser avec l'exemple de cas, le chapitre 4 présente un bref aperçu du DMT2 et du traitement de contrôle glycémique, avec un focus particulier sur le risque d'hypoglycémie associé au traitement. Les données pharmacoéconomiques associées à l'hypoglycémie ont été identifiées à l'aide d'une revue de la littérature et les données canadiennes actuellement disponibles pour la réalisation d'évaluations économiques sont présentées.

Les notions décrites aux chapitre 2, 3 et 4 servent de fondements théoriques aux projets de recherche conduits dans le cadre de cette thèse. Par la suite, les chapitres 5 et 6 rentrent dans le vif du sujet en décrivant les objectifs de recherche ainsi que les méthodologies employées afin d'y répondre. Les résultats de la thèse sont présentés par articles, tous publiés, aux chapitres 7, 8 et 9. Enfin, les chapitres 10 et 11 discutent de l'interprétation des résultats, des forces et faiblesses ainsi que des contributions et des conclusions de cette thèse.

## CHAPITRE 2 : ÉVALUATIONS ÉCONOMIQUES

---

### 2.1 Évaluations économiques

Dans un système de santé avec budget limité, l'efficacité relative des différentes options thérapeutiques doit être comparée afin d'assurer une allocation optimale des ressources. Les évaluations économiques sont un type d'analyse qui permet de comparer les interventions de la santé en termes de coûts et de conséquences. Ainsi, les évaluations économiques servent d'outil aux décideurs publics afin d'optimiser l'utilisation des ressources dans le système de santé, notamment le système canadien.(3, 4) Au Canada, il y a deux agences d'évaluation des technologies de la santé, notamment l'INESSS qui émet des recommandations pour le Québec et l'ACMTS qui émet des recommandations pour toutes les autres provinces.(3, 4) Les éléments clés d'une évaluation économique sont présentés ci-dessous, comme suit : type d'évaluations, population cible, comparateurs, perspective, horizon temporel, actualisation, modélisation, efficacité clinique, mesure et évaluation de la santé, utilisation et coût des ressources, analyse et incertitude. Les éléments présentés découlent des lignes directrices de l'ACMTS et de l'INESSS, reflétant ainsi une perspective canadienne.

#### Type d'évaluations

Il existe plusieurs types d'évaluations économiques, impliquant diverses méthodologies et répondant à différents besoins.(3, 4) L'analyse de minimisation de coûts considère que les interventions comparées ont des effets équivalents. Donc, seuls les coûts influencent le choix des décideurs, ainsi l'intervention la moins dispendieuse sera favorisée. Cependant, la notion d'équivalence entre les effets des interventions doit tenir de l'ensemble des différents aspects, par exemple la fréquence d'événements indésirables ou l'impact sur la qualité de vie, et ne doit pas se limiter uniquement à l'équivalence d'efficacité. À titre d'exemple, l'analyse de minimisation de coûts peut être utilisée lorsque l'intervention est une drogue qui s'ajoute à une classe thérapeutique dans laquelle un médicament similaire est déjà remboursée pour la même indication.(3) Les biosimilaires sont notamment propices à l'analyse de minimisation de coûts. Il faut noter par ailleurs

que, dérogeant au concept théorique d'équivalence des traitements, à de nombreuses reprises l'INESSS et l'ACMTS ont eu recours à l'analyse de minimisation des coûts lorsque les seules données cliniques pour un nouveau médicament s'appuyaient sur une étude de non-infériorité.

L'analyse coût-conséquence ne présente pas de résultats agrégés et consiste plutôt en une description des effets et des coûts.(3, 4) Elle permet de présenter les retombées à plusieurs niveaux, mais l'impact global d'une intervention est plus difficile à interpréter. L'analyse coût-conséquence peut être utile pour les interventions en santé publique ou pour présenter les résultats sous différents angles, par exemple pour une intervention qui aurait des répercussions sur différents secteurs de la société, tels que l'éducation, les services sociaux, la criminalité.(3) On a aussi parfois recours à l'analyse cout-conséquence lorsque les traitements comparés entraînent des effets sur différentes mesures d'efficacité et qu'il est difficile de retenir une seule mesure d'efficacité principale et commune aux traitements comparés. Toutefois, cette analyse demeure un complément d'information et n'est pas à privilégier comme analyse principale.(3, 4)

Bien que prometteuse à ces débuts, l'analyse cout-avantage s'est avéré difficile à considérer. L'analyse coût-avantage est une approche dans laquelle les coûts et les effets sont exprimés en unités monétaires.(3) La valorisation des résultats s'effectue selon la méthode de la volonté de payer par évaluation contingente ou analyse conjointe. La difficulté de ce type d'analyse réside dans la détermination de la valeur financière des effets.(3) Un exemple d'utilisation de l'analyse coût-avantage est l'évaluation des programmes de prévention.

L'analyse coût-efficacité permet quant à elle de comparer les traitements d'intérêt sur la base de leur impact respectif sur une conséquence clinique principale et commune. Cette conséquence est définie en termes d'unités naturelles (p. ex. année de vie sauvée, événement clinique ou indésirable évité). Un seul effet commun est utilisé pour comparer les interventions étudiées, n'offrant ainsi qu'un aperçu partiel des bénéfices.(3, 4) À titre d'exemple, l'analyse coût-efficacité peut être intéressante pour l'évaluation de dispositifs médicaux, où l'effcience d'une

procédure médicale visant à augmenter la fertilité pourrait être exprimée en termes de naissances vivantes. Une limite importante de l'analyse coût-efficacité est l'absence d'un dénominateur commun, empêchant les comparaisons à plus grande échelle dans l'ensemble des champs thérapeutiques.(3, 4) Ainsi, en dehors du contexte d'évaluation, il est difficile d'interpréter et de comparer la valeur du résultat obtenu.

Le type d'évaluation économique privilégié par les agences d'évaluation des technologies est l'ACU, car elle considère les effets à l'aide d'un résultat générique comparable dans l'ensemble des domaines et des interventions.(3, 4) L'ACU présente un dénominateur commun exprimé en termes d'années de vie ajustées en fonction de la qualité (*Quality-Adjusted Life Years* [QALY]). Le nombre de QALY est calculé en pondérant le nombre d'années de vie dans un état de santé donné par l'appréciation de la qualité qui prévaut dans cet état de santé, plus précisément l'utilité. L'utilité sera décrite plus en détail dans la section Mesure et évaluation de la santé. En plus d'être un dénominateur commun, l'avantage du QALY est qu'il permet la capture simultanée des effets sur la morbidité et la mortalité.(3, 4) À noter, il existe des alternatives aux QALY, telles que les *Disability-Adjusted Life Years* (DALY) et les *Healthy Years Equivalent* (HYE), mais ces méthodes ne sont pas privilégiées par l'ACMETS et l'INESSS, et ne font donc pas parties des lignes directrices.(3, 4)

### Population cible

Le ratio coût-efficacité d'une intervention peut varier d'une population à l'autre et donc l'évaluation économique est spécifique à une population cible.(3) La population cible englobe l'ensemble des patients à qui les interventions étudiées sont destinées. La population doit être définie de façon précise en décrivant l'âge de la population, le sexe, la sévérité de la maladie, les comorbidités, etc. Il est possible que l'évaluation économique soit stratifiée par sous-groupe en cas d'hétérogénéité. Également pour certaines interventions de vaste portée, telles que la vaccination, la population générale peut constituer la population cible.(3)

## Comparateurs

Les comparateurs représentent les interventions pertinentes au problème de décision.(3, 4) Dans l'évaluation d'une nouvelle technologie pour fins de remboursement, les comparateurs seront les interventions susceptibles d'être remplacées. Lorsque le problème de décision porte sur la pratique la plus efficiente pour une population donnée, les comparateurs seront toutes les options thérapeutiques actuelles. À noter, les comparateurs sont habituellement des options qui sont déjà financées.(3, 4) Par exemple, les comparateurs acceptés par l'INESSS sont les traitements inscrits sur la liste des médicaments constituant le standard de soin ancré dans la pratique clinique du Québec, alors que ceux retenus par l'ACMTS incluent l'ensemble des traitements et soins courants des différentes provinces.

## Perspective

La perspective est le point de vue adopté dans une évaluation économique et détermine le type de coûts et de conséquences qui feront partie de l'analyse.(3, 4) Les différentes perspectives possibles incluent entre autres le système public de soins, un payeur privé (p.ex. assureur privé), un payeur public (p.ex. un certain ministère) et la société (incluant les patients eux-mêmes). Au Canada, l'ACMTS exige que l'analyse de référence utilise la perspective du système public de soins : les coûts à considérer sont ceux couverts par le payeur public canadien, tandis que les résultats à examiner sont les effets sur la santé importants pour les patients et leurs aidants.(3) Au Québec, l'INESSS recommande de présenter l'évaluation économique selon la perspective du système de soins et la perspective sociétale qui considère l'ensemble des coûts engendrés indépendamment du parti qui les assume.(4) Le choix de la perspective peut avoir des conséquences importantes sur les résultats obtenus. Ainsi selon une perspective donnée un traitement peut s'avérer cout-efficace, alors qu'il ne le serait pas selon une autre perspective.



## Horizon temporel

L'horizon temporel consiste en la période de temps sur laquelle les coûts et les conséquences sont évalués.(3) L'horizon doit être assez long pour capter l'ensemble des différences entre les interventions étudiées. Pour les maladies chroniques, un horizon couvrant la vie entière du patient est habituellement utilisé. Un horizon plus court est utilisé si les différences significatives entre les interventions ne sont plus observées au-delà de cette période (convergence du cheminement clinique) ou si la maladie est brève (p.ex. maladie aigue). Il est possible de tester différents horizons en analyse de scénarios.(3) Un enjeu dans le choix de l'horizon temporel approprié réside dans la disponibilité de données permettant de supporter un tel horizon temporel.

## Actualisation

Lorsque l'évaluation économique porte sur un horizon temporel au-delà d'un an, l'actualisation permet de tenir compte des préférences de la société pour le temps présent. Les coûts et les conséquences doivent être actualisés à un taux de référence de 1.5%.(3, 4) Des analyses complémentaires à 0% et 3% sont recommandées afin de tenir compte de l'incertitude. Une analyse complémentaire avec un taux de 5% peut également être présentée en addition aux autres analyses.(3, 4) Il est à noter qu'avant les lignes directrices de l'ACMTS de 2017, le Canada avait un taux d'actualisation de 5,0% parmi les plus élevés au monde et maintenant avec un taux de 1,5%, le Canada fait partie des plus bas taux d'actualisation.

## Modélisation

La modélisation comprend plusieurs étapes importantes débutant par la conceptualisation du modèle, c'est-à-dire concevoir la structure en fonction d'états de santé et de leurs interrelations pour représenter l'évolution clinique d'une maladie étudiée.(3) Il y a de nombreuses techniques de modélisation décisionnelle telles que l'arbre décisionnel, le modèle d'une cohorte en transition entre des états de santé (p.ex. modèle de Markov), le modèle d'un patient en transition entre des

états de santé (p.ex. microsimulation), les modèles « partition survival », la simulation d'évènements discrets, etc.(3) Lorsque la population à l'étude est modélisée en cohorte, un groupe homogène d'individus présentant des caractéristiques communes évoluent ensemble à travers le modèle. Quant à la microsimulation, elle modélise chacun des individus selon ses propres caractéristiques. Afin de guider les transitions à travers les états du modèle, des données sur l'histoire naturelle de la maladie et sur l'efficacité clinique doivent être incorporées dans la modélisation. Le modèle doit également être validé en s'assurant qu'il se comporte comme prévu lorsque mis à l'épreuve.(3) En général, pour l'évaluation de condition chronique, il est commun d'utiliser un modèle de Markov.

### Efficacité clinique

L'efficacité réfère à l'ensemble des bénéfices cliniques et des effets indésirables influençant à la fois les conséquences et les coûts. Traditionnellement, l'efficacité des interventions dans la modélisation économique provient des ECR.(3, 4) D'autres sources telles que les études observationnelles peuvent également aider à répondre aux besoins de l'évaluation économique en fonction du problème de décision.(3) La crédibilité de l'estimation des paramètres, c'est-à-dire l'absence de biais apparent, est à prendre en considération.(3) Les agences d'évaluation des technologies de la santé privilégient les résultats finaux aux indicateurs de substitution et aux biomarqueurs qui consistent en des résultats intermédiaires.(3)

### Mesure et évaluation du QALY dans l'ACU

L'ACU, le type d'évaluation recommandé par les agences, s'exprime par un dénominateur commun exprimé en termes de QALY.(3, 4) Le nombre de QALY est calculé en pondérant le nombre d'années de vie dans un état de santé donné par l'appréciation de la qualité qui prévaut dans cet état de santé, plus précisément l'utilité. L'utilité est une mesure de préférence associée à un état de santé qui prend une valeur entre 0 et 1. La valeur 1 représente la parfaite santé, la valeur 0 représente la mort et où les valeurs négatives représentent les états pires que la mort.(3) À noter,

la mesure associée à un état de santé peut être exprimée en termes d'utilité ou de décrétement d'utilité qui indique un décrétement de l'état de santé parfait. Le décrétement d'utilité sera présenté sous forme de valeur négative représentant l'impact d'un symptôme ou d'une complication. Dans tous les cas, il est important de présenter les paramètres d'incertitude (p.ex. intervalles de confiance) entourant la valeur d'utilité afin de pouvoir mesurer l'impact de cette incertitude sur l'ensemble de l'évaluation économique.(3) L'utilité peut être estimée via des méthodes de mesure directes ou indirectes. Les mesures directes traduisent les préférences en une échelle d'utilité et les méthodes les plus courantes sont présentés au Tableau 1.(8)

**Tableau 1. Méthodes de mesure directe de l'utilité**

Mesure directe	Description
Échelle analogique visuelle	Positionnement d'un état de santé sur une échelle dont les extrémités représentent la parfaite santé et la mort
Pari standard	Probabilité ( $p$ ) d'indifférence entre l'alternative A (risque entre $p$ de vivre en parfaite santé pour $t$ années et $1-p$ de mort immédiate) et alternative B (vivre avec l'état de santé pour $t$ années)
Arbitrage temporel	Temps ( $x$ ) d'indifférence entre l'alternative A (vivre avec l'état de santé pour $t$ années) et alternative B (vivre en parfait santé pour $x < t$ années suivi de la mort)

Adapté de Brazier et al.(8)

L'échelle analogique visuelle est sujette au biais d'extrémité (les répondants ont tendance à ne pas utiliser les extrémités de l'échelle) et au biais d'étendue (les répondants ont tendance à utiliser l'ensemble de l'échelle quels que soient les états de santé évalués).(9) Le pari standard et l'arbitrage temporel demandent aux répondants de valoriser un état de santé par rapport à la mort, amenant une complexité au niveau de la conception et mise en application.(3) La limite principale des mesures directes réside dans le fait que les répondants doivent valoriser différents scénarios, ce qui peut mener à des résultats biaisés et non représentatifs, particulièrement si les répondants

ne valorisent leur propre état de santé.(3, 10) En effet, les mesures directes peuvent inclure différents types de répondants, tels que des patients qui valorisent leur propre état de santé ou bien des patients ou des membres de la population générale qui valorisent des états de santé fictifs. Ces états de santé fictifs sont communément appelés vignettes, où les vignettes sont des descriptions dites « sur mesure » de certains états de santé d'un modèle économique.(10) Le fait que les états de santé puissent être spécifiquement adaptés au modèle économique est un avantage majeur.(10) Puisque les répondants ne valorisent pas leur propre état de santé, mais plutôt une description de celui-ci, cette méthode facilite le recrutement des répondants, particulièrement lorsque les états de santé évalués sont rares. À l'inverse, l'implication active de cliniciens est nécessaire afin d'élaborer les différentes vignettes.(10) Somme toute, les études de vignettes ne sont pas privilégiées par les agences d'évaluation des technologies de la santé, car la validité des résultats dépend largement de la validité de la description des états de santé.(3, 10)

Quant aux mesures indirectes de l'utilité, elles requièrent une échelle de conversion.(3, 10) Cette conversion se fait à partir d'ensembles de préférences établis ou algorithmes qui ont préalablement été générés par une mesure directe. Les systèmes de classification des états de santé à attributs multiples (mesures indirectes) peuvent être classés comme génériques ou spécifiques, c'est-à-dire conçus pour une maladie en particulier. Les mesures indirectes avec système de classification spécifique peuvent être utilisées dans l'étude de certaines maladies pour lesquelles la capacité des instruments génériques à mesurer des changements entre les états de santé est limitée (p.ex. démence, certains problèmes de vision ou d'audition).(8) Cependant, ce type de mesures restreint la comparabilité entre les différentes conditions de santé.(3, 8, 10) Ainsi, les méthodes de mesures indirectes avec système de classification générique (p.ex. EQ-5D) sont privilégiées par les agences d'évaluation des technologies canadiennes.(3) Le Tableau 2 présente une liste des questionnaires génériques de préférence à attributs multiples qui sont le plus communément utilisés selon une revue par Brazier et al.(8)

**Tableau 2. Méthodes de mesure indirecte de l'utilité**

Méthode	Description	Nombre de niveaux de sévérité	Nombre de combinaisons d'états de santé
15D (11)	15 dimensions : respiration, dépression, inconfort, détresse, alimentation, élimination, ouïe, fonction mentale, mobilité, activité sexuelle, sommeil, parole, activités courantes, vitalité visuelle	4 - 5	31 millions
AQoL-8D (12)	8 dimensions : adaptation, bonheur, mode de vie autonome, santé mentale, douleur, relation, estime de soi, sens	4 - 6	2,37 x 10 <sup>23</sup>
EQ-5D-3L (13)	5 dimensions : anxiété / dépression, mobilité, douleur / inconfort, autonomie, activités courantes	3	243
EQ-5D-5L (14)	5 dimensions : anxiété / dépression, mobilité, douleur / inconfort, autonomie, activités courantes	5	3 125
HUI3 (15)	8 dimensions : vision, ouïe, parole, mobilité, dextérité, émotion, cognition, douleur / inconfort	5 - 6	972 000
SF-6D (16)	6 dimensions : énergie, santé mentale, douleur / inconfort, fonction physique, limitation, fonction sociale	4 - 6	18 000 (SF-36 v1) 18 750 (SF-36 v2) 7500 (SF-12)
QWB-SA (17)	3 dimensions : mobilité, activité physique, fonction sociale + 68 symptômes	2	945

15D : 15-dimensional; AQoL-8D : Assessment of Quality of Life 8 dimensions; EQ-5D : EuroQol 5 dimensions; HUI : Health Utility Index version 3; SF-6D : Short Form 6 dimensions; QWB-SA : Quality of Well-being.

Adapté de Brazier et al.(8)

Lors de la sélection d'un questionnaire générique de préférence à attributs multiples aux fins d'évaluations économiques, les exigences des agences d'évaluation des technologies de la santé doivent être prises en considération, ainsi que les propriétés psychométriques des questionnaires dans la condition de santé d'intérêt.(8) Les propriétés psychométriques incluent entre autres la praticité, la fiabilité, la validité et la réactivité. La praticité, qui fait référence au temps de complétion par les patients, a été démontrée pour l'ensemble des questionnaires du Tableau 2, bien que certains soient plus longs à compléter que d'autres.(8) L'ensemble des questionnaires ont également démontré leur fiabilité (capacité à reproduire la même valeur) à l'aide de la méthode test-retest.(8) La validité de contenu fait référence au degré auquel les différentes dimensions d'une condition médicale sont couvertes par un questionnaire. Afin d'être considéré comme générique, un questionnaire doit couvrir au minimum la santé physique, mentale et sociale ainsi les symptômes somatiques (p.ex. la douleur).(8) L'ensemble des questionnaires couvrent ces dimensions. Toutefois, certains se concentrent plus sur les aspects physiques alors que d'autres couvrent plus en détail les aspects psychologiques. À cet effet, la validité de construction fait référence au degré auquel une mesure reflète les différences connues, où la réactivité, qui est liée à la validité de construction, est la capacité à mesurer un changement dans un état de santé. D'un point de vue pratique, il est aussi important de considérer l'accès aux questionnaires, par exemple les droits d'auteurs, les frais d'utilisation et également les langues pour lesquelles la traduction du questionnaire a été validée.

Bien qu'il n'y ait pas de standard universel, le questionnaire EQ-5D est l'instrument générique le plus utilisé,(8) et est favorisé par plusieurs agences d'évaluation des technologies de la santé à travers le monde.(10) En ce qui a trait au DMT2 spécifiquement, l'EQ-5D a été utilisé à maintes reprises dans l'étude de cette condition et génère des valeurs d'utilité pertinentes à la modélisation économique.(18) À noter, les valeurs d'utilité associées au DMT2 seront décrites au chapitre 4. Considérant que l'EQ-5D est un instrument validé largement répandu, et aux fins de comparabilité en modélisation économique en DMT2, ce questionnaire a été sélectionné pour la collecte des mesures d'utilité dans l'article 3 de cette thèse.

Le questionnaire EQ-5D couvre cinq dimensions, soient l'anxiété / la dépression, la mobilité, la douleur / l'inconfort, l'autonomie et les activités courantes. Dans le questionnaire original, l'EQ-5D-3L, chacune des cinq dimensions se divisent en trois niveaux (pas de problème, problèmes modérés et problèmes extrêmes).(13) Toutefois, l'EQ-5D-3L présente deux limites principales, soient l'effet « plafond » (trop de patients présentent un état de santé parfait) et un nombre de niveaux de sévérité insuffisant pour détecter de faibles changements dans un état de santé.(8) Ces faiblesses ont mené au développement de l'EQ-5D-5L, dans lequel chacune des cinq dimensions se divisent en cinq niveaux (pas de problème, problèmes légers, problèmes modérés, problèmes sévères et problèmes extrêmes).(14) S'agissant d'un questionnaire générique de préférence à attributs multiples, les états de santé du EQ-5D-5L doivent être convertis à l'aide d'une échelle de conversion en un score index sommaire. Un score index d'utilité de 1 représente la parfaite santé, 0 représente la mort et les valeurs négatives représentent les états de santé pires que la mort. Une échelle de conversion au EQ-5D-5L a été développée au Canada et sera utilisée dans l'article 3.(19) Le questionnaire inclut également une échelle analogue visuelle représentée par une échelle verticale sur laquelle le patient peut indiquer son état de santé quelque part entre les extrémités qui représentent le meilleur état de santé imaginable et le pire état de santé imaginable. Cette échelle analogue visuelle est une mesure directe de l'utilité, tel que précédemment décrit. Les scores du EQ-5D-5L représentent l'état de santé du répondant la journée de complétion du questionnaire.(14)

### Utilisation et coût des ressources

Il y a trois étapes principales à considérer, soit le relevé, la mesure et l'évaluation des ressources.(3) Le relevé des ressources consiste en l'identification de l'ensemble des ressources pertinentes engendrées dans le cheminement clinique de la maladie d'intérêt. La mesure des ressources correspond à la quantification des ressources utilisées. Puis, l'évaluation des ressources consiste en l'attribution d'un coût aux ressources consommées. Il existe différents types de coûts, notamment les coûts directs et indirects, et la perspective adoptée déterminera lesquels inclure dans l'évaluation économique.(3) Les coûts médicaux directs correspondent aux coûts de soins de santé, par exemple les frais associés à l'utilisation de médicaments, services hospitaliers, visites

médicales, etc.(3) Les coûts médicaux directs peuvent être estimés à l'aide de différentes sources de données, telles que les dossiers médicaux électroniques, les registres et les données provenant des régimes d'assurance. Les données médico-administratives de la Régie d'assurance-maladie du Québec en sont un exemple. Les banques de données administratives peuvent être des sources d'information grandement utiles pour estimer l'utilisation des ressources aux fins d'évaluations économiques.(3, 20) Toutefois, il convient de reconnaître les limites associées à ce type de sources de données. La principale limite réside dans le fait que les données ont été collectés dans un but précis, par exemple la facturation, et par conséquent ne sont pas nécessairement adaptés à différents besoins de recherche. À titre d'exemple, lors de l'analyse de réclamations d'assurance, il est possible que le diagnostic d'un patient manque de spécificité et ne reflète pas l'ensemble de l'état de santé du patient.(5) Plusieurs données cliniques peuvent également manquer de précision, voire être complètement absentes.(5) Il peut également y avoir des erreurs de codification.(5) De plus, l'accès aux banques de données administratives n'est pas automatique et instantané.(20)

Un portrait des coûts médicaux directs, quoique moins précis, peut également être obtenu en consultant directement les patients ou les médecins à l'aide de questionnaires ou d'entrevues par exemple. Puisque les données collectées reposent sur la mémoire de l'individu, la qualité peut en être affectée.(5) Les données peuvent également manquer d'objectivité et ne pas refléter la réalité avec exactitude.(5) Un des avantages de consulter les patients est la collecte des coûts non-médicaux directs, qui sont les coûts assumés par le patient (p.ex. transport, débours personnels).(3) L'option d'utiliser des données secondaires n'est habituellement pas possible, ainsi le patient devient la source principale d'information. Également, la perte de productivité des patients est un aspect important à considérer lors d'une analyse économique menée dans la perspective sociale.(3, 4) Cette perte de productivité englobe divers aspects tels que l'absentéisme (temps de travail rémunéré perdu), le présentéisme (productivité au travail réduite) et le temps de travail non rémunéré perdu (capacité à réaliser les activités courantes).(21)

Lors de la réalisation d'une évaluation économique, les données de perte de productivité doivent être converties en unité monétaire, c'est-à-dire que l'unité de temps devient une unité de



coût. Pour ce faire, il existe plusieurs méthodes.(21) Les deux principales méthodes pour la conversion du temps de travail rémunéré perdu sont celles du capital humain et du coût de friction. La méthode du capital humain multiplie simplement le temps perdu par le taux de rémunération. La méthode du coût de friction quant à elle se limite à la période de friction, c'est-à-dire la période nécessaire à une organisation pour remplacer l'individu et retrouver son niveau de productivité initial. Pour déterminer la valeur du temps de travail non rémunéré perdu, les principales méthodes incluent la méthode du coût de renonciation et du coût de remplacement. Le coût de renonciation représente le manque à gagner, où la valeur du temps équivaut à celle d'une autre fonction (p.ex. travail rémunéré). Le coût de remplacement considère plutôt le coût d'acquisition de la ressource sur le marché (p.ex. prix d'un service d'entretien ménager).(21) Afin que les données de perte de productivité soient applicables à différentes évaluations économiques, il est important que les données brutes (en unité de temps avant la conversion en unité monétaire) soient accessibles. Ceci prévaut également pour les coûts directs, où l'accès aux données d'utilisation des ressources de la santé favorise la transférabilité des évaluations économiques à travers différentes juridictions. Ainsi, il est recommandé de publier les données brutes afin de permettre leur réutilisation dans différents scénarios adaptés à différents besoins.(22)

En 2015, une revue par Tang et al.(21) a identifié les questionnaires génériques de mesure de la productivité les plus fréquemment utilisés et un sommaire est présenté au Tableau 3.

**Tableau 3. Instruments génériques de mesure de la productivité**

<b>Instrument</b>	<b>Période de rappel</b>	<b>Forces</b>	<b>Limites</b>
HLQ (23)	2 semaines	<ul style="list-style-type: none"> <li>• Inclut absentéisme, présentéisme et travail non rémunéré</li> <li>• Compatible avec 2 méthodes de conversion du temps de travail rémunéré</li> <li>• Permet la conversion du travail non rémunéré en unités monétaires</li> </ul>	<ul style="list-style-type: none"> <li>• Évaluation du présentéisme est complexe pour les patients</li> <li>• Complétion difficile menant à de faibles taux de réponse</li> </ul>

<b>Instrument</b>	<b>Période de rappel</b>	<b>Forces</b>	<b>Limites</b>
HPQ (24)	Versions : 7 jours; 4 semaines	<ul style="list-style-type: none"> <li>Validité démontrée par comparaison à indicateurs directs (p.ex. chèques de paie)</li> <li>Plus d'une version disponible selon les besoins</li> </ul>	<ul style="list-style-type: none"> <li>N'inclut pas le travail non rémunéré</li> <li>Inclut une notion de présentéisme relatif dont l'applicabilité manque de clarté</li> </ul>
HRPQ-D (25)	Journalier	<ul style="list-style-type: none"> <li>Inclut absentéisme, présentéisme et travail non rémunéré</li> <li>Simple et court</li> </ul>	<ul style="list-style-type: none"> <li>Conversion du travail non rémunéré n'est pas compatible avec toutes les méthodes</li> <li>Définition du travail non rémunéré trop large</li> <li>Évidence psychométrique limitée</li> </ul>
PRODISQ (26)	3 mois (jour le plus récent pour présentéisme)	<ul style="list-style-type: none"> <li>Adapté à la perspective de l'employeur</li> <li>Inclut des modules additionnels mesurant l'impact sur l'ensemble du lieu de travail</li> </ul>	<ul style="list-style-type: none"> <li>N'inclut pas le travail non rémunéré</li> <li>Certains modules additionnels requièrent la participation de plusieurs parties</li> <li>Évidence psychométrique limitée</li> </ul>
QQ (27)	Jour de travail le plus récent	<ul style="list-style-type: none"> <li>Simple et court</li> <li>Compréhension intuitive des concepts</li> </ul>	<ul style="list-style-type: none"> <li>Évalue uniquement le présentéisme</li> <li>Difficulté à combiner les items qualitatifs et quantitatifs</li> </ul>
SPS-13 (SPS-6) (28)	4 semaines	<ul style="list-style-type: none"> <li>Simple et pratique</li> <li>Inclut plusieurs perspectives du présentéisme</li> </ul>	<ul style="list-style-type: none"> <li>N'inclut pas le travail non rémunéré</li> <li>SPS-6 : inclut uniquement le présentéisme</li> <li>Évidence psychométrique limitée</li> </ul>
VOLP (29)	3 mois (7 jours pour présentéisme)	<ul style="list-style-type: none"> <li>Inclut absentéisme, présentéisme et travail non rémunéré</li> <li>Inclut des modules additionnels mesurant l'impact sur l'ensemble du lieu de travail</li> <li>Compatible avec 2 méthodes de conversion du temps de travail rémunéré</li> </ul>	<ul style="list-style-type: none"> <li>Le travail non rémunéré inclut uniquement la notion de remplacement</li> <li>Long et complexe</li> </ul>

<b>Instrument</b>	<b>Période de rappel</b>	<b>Forces</b>	<b>Limites</b>
WHI (30)	2 semaines	<ul style="list-style-type: none"> <li>• Facile de compréhension</li> <li>• Validité démontrée par comparaison à indicateurs directs (p.ex. données administratives du lieu de travail)</li> </ul>	<ul style="list-style-type: none"> <li>• Nécessite une entrevue conduite via un appel téléphonique automatisé</li> <li>• N'inclut pas le travail non rémunéré</li> </ul>
WLQ (31)	2 semaines	<ul style="list-style-type: none"> <li>• Validité largement démontrée</li> <li>• Formule de conversion développée à partir d'indicateurs directs (p.ex. données administratives du lieu de travail)</li> </ul>	<ul style="list-style-type: none"> <li>• Inclut uniquement le présentéisme</li> <li>• L'applicabilité de la formulation de conversion à certains lieux de travail est incertaine</li> </ul>
WPAI (32)	1 semaine	<ul style="list-style-type: none"> <li>• Inclut absentéisme, présentéisme et travail non rémunéré</li> <li>• Simple et court</li> <li>• Questionnaire le plus utilisé</li> </ul>	<ul style="list-style-type: none"> <li>• Le travail non rémunéré inclut les loisirs</li> <li>• Ne permet pas d'évaluer l'absentéisme de longue durée</li> </ul>
WPSI (33)	Versions : 2 semaines; 3 mois; 12 mois	<ul style="list-style-type: none"> <li>• Permet d'évaluer l'impact de plusieurs conditions de santé</li> <li>• Peut être court selon l'applicabilité des questions</li> </ul>	<ul style="list-style-type: none"> <li>• N'inclut pas le travail non rémunéré</li> <li>• Difficulté à séparer l'impact des différentes conditions</li> <li>• Évidence psychométrique limitée</li> </ul>

HLQ : Health and Labor Questionnaire; HPQ : Health and Work Productivity Questionnaire; HRPQ-D : Health-Related Productivity Questionnaire Diary; PRODISQ : Productivity and Disease Questionnaire; QQ : Quantity and Quality method; SPS : Stanford Presenteeism Scale; VOLP : Valuation of Lost Productivity Questionnaire; WHI : Work and Health Interview; WLQ : Work Limitations Questionnaire; WPAI : Work Productivity and Activity Impairment; WPSI : Work Productivity Short Inventory.

Adapté de Tang(21)

Dans la revue publiée par Tang(21), les auteurs reconnaissent qu'il n'y a pas d'instrument universel et que le choix de l'instrument par les chercheurs doit être adapté selon le but de l'étude, la perspective, la population, la psychométrie et l'applicabilité/fonctionnalité. Les auteurs discutent également d'un nouvel instrument de mesure de perte de productivité développé par l'*institute for Medical Technology Assessment (iMTA)* et reconnu sous le nom *iMTA Productivity Cost Questionnaire (iPCQ)*.(34) Lors de la revue, le *iPCQ* était encore à l'étape de validation, mais

les auteurs le décrivent comme un outil prometteur dont le développement a été basé sur trois instruments validés de mesure de perte de productivité. Le questionnaire couvre les trois types de productivité (absentéisme, présentéisme et travail non rémunéré) et a une période de rappel de quatre semaines. Le questionnaire inclut également les absences prolongées qui précèdent la période de rappel.

En 2021, une mise à jour des questionnaires de mesure de productivité a été publiée par Hubens et al.(35) Cette revue systématique a identifié les instruments utilisés dans le cadre d'évaluations économiques ou d'études de validation entre 2018 et 2019. Cette courte période a été choisie par les auteurs afin de présenter un portrait actuel des instruments de mesure de la productivité.(35) Les questionnaires génériques identifiés par Hubens(35) étaient très similaires à ceux identifiés par Tang(21), avec comme différence majeure l'ajout du iPCQ qui fait désormais partie des trois instruments génériques les plus utilisés, avec les questionnaires *Work Productivity and Activity Impairment (WPAI)* et *Productivity and Disease Questionnaire (PRODISQ)*.(35) Dans la revue, la pertinence des instruments de mesure de la productivité aux fins d'évaluations économiques a été jugée selon trois critères, soient les dimensions mesurées, la quantification des résultats et la conversion en unités monétaires. Pour chaque critère, la mention « atteint », « partiellement atteint » ou « non atteint » a été attribuée. Les dimensions mesurées étaient jugées complètes si elles incluent l'absentéisme, le présentéisme et le travail non rémunéré. La quantification des résultats était jugée adéquate si elle permet l'obtention de données quantifiables, telles qu'un nombre d'heures, un nombre de jours ou un pourcentage. La conversion en unités monétaires était jugée adéquate si elle est compatible avec les deux méthodes les plus couramment utilisées soient l'approche du capital humain et l'approche du coût de friction décrites plus haut. Le Tableau 4 présente l'évaluation des instruments les plus communs selon la revue par Hubens.(35)

**Tableau 4. Évaluation des instruments génériques de mesure de la productivité aux fins d'évaluations économiques**

<b>Instrument</b>	<b>Dimensions mesurées</b>	<b>Quantification</b>	<b>Conversion monétaire</b>
HLQ(23)	Atteint	Atteint	Atteint
HPQ(24)	Partiellement atteint	Partiellement atteint	Non atteint
iPCQ(34)	Atteint	Atteint	Atteint
PRODISQ(26)	Partiellement atteint	Partiellement atteint	Partiellement atteint
QQ(27)	Non atteint	Non atteint	Non atteint
SPS-6(28)	Non atteint	Non atteint	Non atteint
VOLP(29)	Atteint	Atteint	Non atteint
WLQ(31)	Partiellement atteint	Partiellement atteint	Non atteint
WPAI(32)	Atteint	Atteint	Non atteint

HLQ : Health and Labor Questionnaire; HPQ : Health and Work Productivity Questionnaire; iPCQ : Institute for Medical Technology Assessment Productivity Cost Questionnaire; PRODISQ : Productivity and Disease Questionnaire; QQ : Quantity and Quality method; SPS : Stanford Presenteeism Scale; VOLP : Valuation of Lost Productivity Questionnaire; WLQ : Work Limitations Questionnaire; WPAI : Work Productivity and Activity Impairment.

Adapté de Hubens(35)

Le questionnaire WPAI, qui est l'instrument le plus utilisé, mesure les trois dimensions de la productivité et produit des résultats quantifiables.(35) Toutefois, les auteurs ont jugé que la capacité de conversion monétaire était incomplète, car le questionnaire ne mesure pas l'absentéisme de longue durée (qui est limité à la période de rappel de sept jours) et ne reflète donc pas nécessairement la réalité du répondant.(35) Quant au questionnaire PRODISQ, il n'inclut pas la dimension du travail non rémunéré. À l'exception du questionnaire iPCQ, seul l'instrument *Health and Labor Questionnaire* (HLQ) rencontrait les trois critères. Puisque le questionnaire iPCQ a été construit à partir des questionnaires HLQ, PRODISQ et Quantity and Quality (QQ),

les auteurs prévoient une diminution significative de l'utilisation de ces trois questionnaires qui seront remplacée par la version améliorée qu'est l'iPCQ.(35)

Le questionnaire *iPCQ* a été sélectionné comme instrument de mesure de la productivité pour l'article 3 de cette thèse. Il s'agit d'un instrument récemment développé et validé qui permet d'estimer les pertes de productivité.(34) Le *iPCQ* est un questionnaire générique conçu pour déterminer la valeur de la perte de productivité pour les évaluations économiques adoptant la perspective sociétale. Le questionnaire comprend trois modules : l'absentéisme (travail rémunéré perdu), le présentéisme (productivité réduite) et la perte de productivité lors du travail non rémunéré (activités courantes). Une période de rappel de quatre semaines est utilisée pour le présentéisme et le travail non rémunéré. L'absentéisme est divisé en deux catégories : l'absence court terme (quatre semaines et moins) et l'absence prolongée (congé de maladie débutant avant la période de rappel). La période de rappel de quatre semaines pour la mesure de la productivité correspond exactement à la période de mesure des événements hypoglycémiques dans l'article 3 de cette thèse. De plus, selon la revue publiée par Tang, le choix de l'instrument de mesure de la productivité devrait être basé sur le but de l'étude, la perspective, la population, la psychométrie et l'applicabilité/fonctionnalité.(21) Le questionnaire *iPCQ* rencontre ces cinq critères en étant un instrument générique et validé, mesurant les trois dimensions de la productivité, permettant la conversion en unités monétaires et limitant le fardeau imposé aux patients. En effet, le questionnaire *iPCQ* est simple et prend entre cinq à dix minutes à compléter.(34)

### Analyse

Depuis la publication des nouvelles lignes directrices en 2017, l'ACMTS exige désormais que l'analyse de référence soit une analyse probabiliste.(3) Dans l'analyse probabiliste, la valeur des paramètres est représentée par une distribution statistique plutôt que par une estimation ponctuelle afin de caractériser l'incertitude sous-jacente.(3) La simulation de Monte Carlo est communément utilisée afin de caractériser l'incertitude entourant les estimations.(3) L'analyse probabiliste est préférable à l'analyse déterministe qui prend en compte seulement les éléments

individuels. L'analyse probabiliste nécessite d'attribuer une forme de distribution à chaque paramètre. Pour l'utilité, le choix est habituellement la distribution bêta qui s'échelonne de 0 à 1.(3) Pour les coûts et les décrets d'utilité (qui ont un déport à droite plutôt qu'à gauche), les distributions gamma ou log-normale sont souvent utilisées.(3)

Les résultats sont présentés sous forme de rapport coût-efficacité différentiel (*Incremental Cost-Effectiveness Ratio* [ICER]). Le rapport coût-efficacité différentiel consiste en la différence entre les coûts escomptés de deux interventions, divisée par la différence entre les résultats/effets escomptés de ces deux interventions.(3) Lorsque que l'intervention étudiée est moins coûteuse et plus efficace que le comparateur, l'intervention est jugée comme une stratégie dominante. À l'inverse, lorsque l'intervention est plus coûteuse et moins efficace, elle est qualifiée de stratégie dominée. Lorsque l'intervention est moins efficace et moins coûteuse, des coûts peuvent être sauvés en renonçant à des bénéfices en santé. Lorsque l'intervention est plus efficace et plus coûteuse, des bénéfices additionnels peuvent être obtenus avec des coûts supplémentaires.(3) Le seuil de propension à payer peut être fixé par les agences d'évaluation des technologies de la santé afin d'aider les payeurs à prendre une décision en termes de remboursement.

### Incertitude

L'incertitude d'une évaluation économique englobe différents aspects tels que l'incertitude méthodologique, l'incertitude structurelle et l'incertitude paramétrique.(3) L'incertitude méthodologique relève des différents choix méthodologiques, par exemple le choix du taux d'actualisation.(3) L'incertitude méthodologique peut être testée en analyse complémentaire en démontrant les différences engendrées par un choix méthodologique différent de l'analyse de référence. L'incertitude structurelle fait référence à l'aptitude du modèle économique à représenter avec exactitude de tous les aspects pertinents de la maladie et des interventions comparées.(3) À titre d'exemple, si la forme fonctionnelle choisie pour le temps écoulé avant la survenue d'un événement est incertaine, une analyse avec différentes courbes de survie est requise. L'incertitude paramétrique désigne l'incertitude associée à la justesse des paramètres, où la « vraie » valeur est

inconnue. L'incertitude entourant les paramètres du modèle peut être testée dans une analyse probabiliste dans laquelle la simulation de la distribution des probabilités permet de quantifier l'incertitude des résultats.(3)

Tout type d'incertitude affecte la confiance en les résultats d'une évaluation économique. Le présent projet de recherche se concentre sur l'incertitude paramétrique, plus précisément la justesse des données pharmacoéconomiques. Ceci nous amène au chapitre 3 dans lequel les différents biais pouvant affecter la justesse des résultats sont présentés.



## CHAPITRE 3 : BIAIS DANS LES ÉTUDES EN PRATIQUE RÉELLE

---

Au moment de la conception d'une étude, le chercheur doit prendre en considération deux types d'erreurs possibles : l'erreur aléatoire (erreur statistique affectant la précision des résultats) et l'erreur systématique (biais affectant la justesse des résultats). En effet, le biais est une erreur systématique qui fausse l'estimation de la relation entre l'exposition et l'issue, affectant ainsi la justesse (validité interne) des résultats. Il y a trois grandes catégories de biais : biais de sélection, biais d'information et biais de confusion.(36, 37) Le risque de biais est présent dans tout type d'études, mais le risque est plus élevé dans les études en pratique réelle, notamment dû à l'absence de randomisation et de suivi à double insu.(36, 37) Également, le risque de biais n'est pas exclusif à la recherche d'efficacité comparative ou à l'épidémiologie en général, mais peut certainement se produire lors de la collecte de données pharmacoéconomiques aux fins d'évaluations économiques.(38-48) Le présent chapitre résume les principaux biais auxquels les études en pratique réelle sont sujettes et plusieurs exemples généraux sont présentés afin de clarifier les concepts théoriques. L'applicabilité de ces différents biais dans la conduite d'études collectant des données pharmacoéconomiques fait partie intégrante de l'objectif 2 de cette thèse et sera donc approfondie à la section 6.2. Le recensement des écrits afin d'éviter les biais lors de la collecte de données pharmacoéconomiques en pratique réelle fait également partie de l'objectif 2 et sera présenté en détail à la section 6.2.

### 3.1 Biais de sélection

Le biais de sélection se produit lorsque certains patients éligibles, certaines périodes de temps ou certains évènements sont exclus de façon à ce que l'association entre l'exposition et l'issue du groupe à l'étude diffère de celle de la population cible. Un biais se produit à l'entrée des patients dans l'étude lorsque la sélection des patients est influencée par l'exposition et l'issue.(36, 37) Lors de la conduite d'études rétrospectives, l'issue et l'exposition se sont déjà produites, ouvrant la porte à plusieurs types de biais de sélection. Par exemple, le biais de prévalence, un sous-type de biais de sélection, peut se produire lorsque certaines périodes de temps sont exclues

de l'analyse, par exemple, lorsque des patients prévalents sont sélectionnés dans l'étude. En effet, l'issue peut être manquée si elle se produit peu de temps après le début de l'exposition, alors que le patient n'est pas encore inclus dans l'étude. Pour éviter ce type de biais, il est recommandé que le début du suivi des patients coïncide avec l'incidence de l'exposition.(37) Le biais de prévalence peut se produire par exemple lors de l'estimation de l'occurrence d'un événement indésirable. Lorsque les patients sont intolérants à un certain médicament, cela mène habituellement à l'arrêt prématuré du traitement, où l'analyse de patients prévalents, c'est-à-dire encore sous traitement, sous-estimerait la fréquence de l'événement indésirable et mènerait à une fausse représentation de la population recevant le traitement d'intérêt.

Un biais très similaire, mais spécifique aux conditions associées à un risque de mortalité élevé, est le de biais de survie. Ce biais se produit lorsque des périodes de temps durant lesquelles le patient ne pas mourir sont exclues de façon différentielle entre les groupes d'exposition.(37) Par exemple, lors de la comparaison du taux de mortalité entre les patients atteints de la maladie pulmonaire obstructive chronique recevant les télésoins à domicile ou les soins standards, l'exclusion de la période de suivi entre par exemple la sortie de l'hôpital et le début des télésoins à domicile sous-estimerait le taux de mortalité dans ce groupe. En effet, avant d'avoir accès aux télésoins à domicile, les patients doivent généralement survivre plusieurs semaines post-hospitalisation, créant ainsi une période de temps immortel. Dans cette situation, plutôt que de débiter le suivi des patients à l'initiation des soins, les patients devraient tous être suivis à partir de la même date index, par exemple la sortie de l'hôpital.

Un autre sous-type de biais de sélection est le biais de référence (biais d'hospitalisation, biais de Berkson).(36) Ce biais se produit lorsque le statut d'exposition influence l'investigation du diagnostic, donc que la raison de la référence/hospitalisation est liée au statut d'exposition. Ce biais est improbable si l'effet de l'exposition n'est pas un phénomène connu et si l'issue étudiée a peu de chance d'être manquée, c'est-à-dire que l'effet est majeur et apparent. Ce biais se produit lorsqu'un médicament contribue au processus diagnostique, par exemple en étant associé à des effets indésirables bien connus qui vont influencer la probabilité de recevoir un diagnostic de

l'issue. Par exemple, lorsqu'un patient consommant des anti-inflammatoires non-stéroïdiens présente une douleur abdominale, son médecin pourrait être plus susceptible de penser que le patient a un ulcère gastrique compte tenu des effets indésirables connus de ce médicament. Donc, ce patient a plus de chance d'être référé à l'hôpital pour investigation comparativement à un autre patient présentant avec une douleur abdominale, mais ne recevant pas d'anti-inflammatoires non-stéroïdiens. L'association entre les anti-inflammatoires non-stéroïdiens et les ulcères gastriques pourrait donc être surestimée. Une piste de solution suggérée est de restreindre l'éligibilité aux cas les plus sévères et qui seront donc tous diagnostiqués indépendamment du statut d'exposition.(36) Dans l'exemple précédent, restreindre les ulcères gastriques aux cas hémorragiques (avec saignements gastro-intestinaux) limiterait le risque de biais considérant que le diagnostic de l'issue est plus évident. Cette forme du biais de Berkson peut être généralisée plus largement, c'est-à-dire qu'il y a un biais lorsque la sélection des participants est basée sur un collisionneur (*collider*), par exemple l'hospitalisation, un type de variable qui sera décrit à la section 3.3. Une représentation graphique est également présentée à la Figure 1.(36) Un exemple de cette forme de biais a été démontré par des chercheurs qui voulaient étudier l'association entre les maladies de l'appareil locomoteur et les maladies respiratoires.(49) Suite à la sélection de patients hospitalisés, les chercheurs ont démontré une forte association entre les deux maladies. Toutefois, les maladies de l'appareil locomoteur et les maladies respiratoires sont deux causes indépendantes de l'hospitalisation, faisant de l'hospitalisation un collisionneur. Les chercheurs ont par la suite conduit la même analyse, mais dans la population générale (incluant un mélange de patients hospitalisés et non-hospitalisés), dans laquelle aucune association entre les deux maladies n'a été démontrée. Ainsi, la sélection d'un échantillon de patients basée sur un collisionneur tel que l'hospitalisation peut biaiser les résultats.(49)

Le biais protopathique est un autre sous-type de biais de sélection qui se produit lorsque l'exposition est modifiée par des manifestations initiales précédant l'issue, notamment l'apparition de symptômes. Ce problème de causalité inverse affecte principalement les maladies avec un délai de détection. Ce biais se produit lorsqu'un médicament est initié afin de traiter les symptômes apparus ou lorsqu'un médicament est arrêté dû à l'apparition des symptômes, menant à une fausse association avec l'issue. Par exemple, un patient pourrait se voir prescrire des analgésiques en

réponse à une douleur causée par un cancer non diagnostiqué, où l'association entre l'utilisation d'analgésiques et le cancer pourrait être faussement surestimée. Une solution proposée est l'ajout d'un laps de temps avant l'issue où l'exposition n'est pas mesurée, c'est-à-dire d'exclure la période de pré-symptômes.(50) Une compréhension des mécanismes pathophysiologiques du développement de l'issue est également importante.

Un biais de sélection important qui affecte les études avec collecte de données primaires est le biais d'auto-sélection. Ce biais se présente lorsque les participants à l'étude diffèrent de la population cible, représentant habituellement en plus grande proportion les patients plus en forme. Par exemple, les patients avec une forme plus sévère de la maladie peuvent être moins enclins à participer. Pour diminuer ce risque de biais, les patients exposés et non-exposés doivent être sélectionnés de la même population source. Une façon de documenter le risque de biais d'auto-sélection est de compiler les taux de refus de participation.(51) Plus les informations sur les non-participants sont complètes, plus il est possible de voir s'il y a des différences avec ceux qui ont accepté de participer à l'étude. L'autre biais de sélection significatif dans les études de collecte de données primaires est la perte au suivi, où les patients qui abandonnent l'étude sont différents. Par exemple, certains patients pourraient se retirer de l'étude lorsque leur maladie progresse à un stade plus avancé. Ce biais se produit lorsque les taux de données manquantes (données incomplètes ou attrition) sont différents entre les groupes d'expositions et que les raisons sont associées à l'issue. Il existe différentes techniques analytiques pour le traitement des données manquantes, mais la meilleure solution reste la prévention via des incitatifs, rappels, etc.

Une des techniques pour la gestion des données manquantes consiste en la restriction de l'analyse aux patients sans données manquantes. Ce type d'analyse peut toutefois réduire la taille de l'échantillon et peut entraîner un biais si les données ne sont pas manquantes complètement au hasard.(52) Parmi l'ensemble des techniques analytiques existantes, des exemples communs sont la pondération par la probabilité inverse (*inverse probability weighting*), l'approche bayésienne et l'imputation.(53) Lors de la pondération par la probabilité inverse, la probabilité de ne pas présenter de données manquantes va être calculée et inversée.(52) La probabilité inverse sera par

la suite utilisée afin de pondérer les patients de façon inversement proportionnelle, où une pondération (*weight*) plus élevée sera attribuée à un patient ayant une faible probabilité d'avoir des données complètes. Cela permet une meilleure représentativité des patients qui ont complété l'étude, mais qui selon leurs caractéristiques, avaient une forte probabilité d'abandon et donc d'avoir des données manquantes.(52) L'approche bayésienne quant à elle utilise l'information connue sur la distribution a priori des données partiellement observées afin de dériver la distribution a posteriori.(54) Les informations quant à la distribution a priori peuvent provenir des données internes (p.ex. sous-échantillon de la population à l'étude) ou de données externes (p.ex. données publiées dans la littérature). La distribution a posteriori peut ensuite être déterminée à l'aide d'un modèle d'échantillonnage, des données partiellement observées et de la distribution a priori.(54) Pour ce qui est de la méthode d'imputation, elle remplace une donnée manquante par une valeur prédite.(52) Il existe plusieurs techniques d'imputation, telles que l'imputation par la moyenne où la moyenne des valeurs d'une variable va remplacer les données manquantes de cette même variable. La dernière observation reportée est une autre technique d'imputation dans laquelle une donnée manquante est remplacée par la dernière valeur observée pour cette variable chez un même patient. Afin de conserver l'incertitude entourant une donnée manquante, l'imputation multiple peut être utilisée, laquelle remplace les variables manquantes par un ensemble de variables possibles. Les valeurs possibles sont définies à partir des distributions des données observées à l'aide d'un modèle de prédiction.(52)

### **3.2 Biais d'information**

Le biais d'information se produit lorsqu'il y a une erreur dans l'attribution du statut d'exposition ou de l'issue.(55) La classification erronée non différentielle est présente lorsque l'erreur est aléatoire et non reliée à l'exposition ou à l'issue. Par exemple, dans une étude utilisant une banque de données administratives, il serait possible de manquer le diagnostic d'un événement par manque de précision de la codification. Cette erreur est toutefois non différentielle, car elle n'est pas reliée au statut d'exposition de l'étude. Ce type de biais amoindrit l'effet observé.(55) Lors de classification erronée différentielle, où l'erreur d'attribution est liée à l'exposition ou à l'issue, la direction du biais est inconnue (surestimation ou sous-estimation de l'effet).(55) Un

sous-type de biais d'information est le biais de temps immortel, où le temps immortel représente une période de temps durant laquelle un événement ne peut pas survenir. Ce biais se produit lorsque l'exposition est attribuée de manière erronée sur une certaine période de temps immortel.(56) Par exemple, une étude de banque de données administratives chez les patients diabétiques détermine la date d'entrée dans la cohorte en fonction de la première prescription de metformine. Dans cette cohorte de patients diabétiques, les auteurs cherchent à savoir si l'utilisation chronique de statines prévient un certain événement indésirable, où l'utilisation chronique est définie comme une période d'au moins un an. Cette période d'un an devient une période de temps immortel et l'assignation de cette période au groupe exposé entraîne un biais de classification erronée différentielle. Une solution serait d'utiliser un statut d'exposition variable, plutôt que fixe, c'est-à-dire que la première année serait considérée comme non exposée et que le début de l'exposition chez ce patient commencerait après cette période de temps immortel. L'occurrence de l'événement indésirable pourrait ensuite être évaluée à l'aide d'une technique permettant de varier le statut d'exposition chez un même patient, tel qu'un modèle de Cox dépendant du temps.(56)

Un autre sous-type de biais d'information est le biais de détection. Ce biais se produit lorsque les procédures pour mesurer l'exposition ou l'issue sont réalisées de façon différentielle.(36) Par exemple, les femmes qui prennent des suppléments hormonaux après la ménopause voient habituellement leur médecin plus fréquemment afin d'assurer un suivi. Ce suivi différentiel des patientes pourrait mener à une surestimation des cas de cancers de l'endomètre qui ont plus de chance d'être détectés lors d'un suivi médical strict. Enfin, pour les études où l'exposition ou l'issue est rapportée par le patient, le biais de rappel est un enjeu important. Le biais de rappel est présent lorsqu'un groupe de patients a plus de raisons de se souvenir de l'exposition ou de l'issue étudiée. Par exemple, les femmes ayant eu un enfant avec une malformation ont plus de chance de se rappeler de l'ensemble des médicaments consommés durant la grossesse par rapport aux femmes avec un enfant sans malformation. Les études en aveugle (*blinding*) demeurent la meilleure solution pour prévenir le biais d'information. L'aveugle étant souvent impraticable en pratique réelle, l'utilisation d'outils de mesure validés et de procédures standardisées demeurent l'option la plus commune.(36)

### 3.3 Biais de confusion

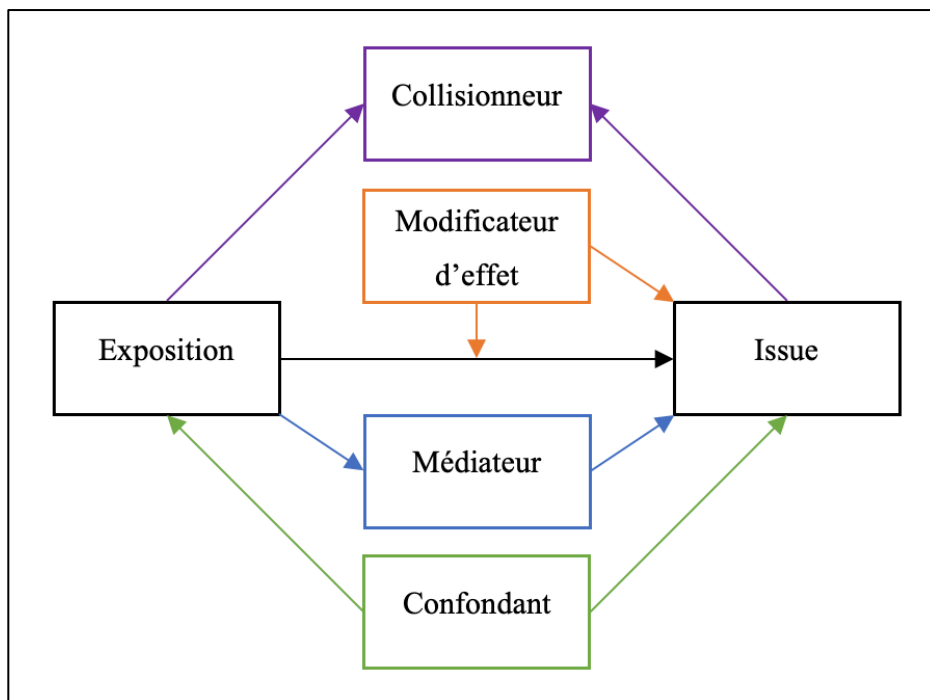
Le biais de confusion se produit lorsque la relation entre l'exposition et l'issue est modifiée par une tierce variable (facteur confondant).(36) Il y a trois conditions pour qu'une variable soit considérée comme un confondant : 1) être un facteur de risque de l'issue, 2) être associée à l'exposition et 3) ne doit pas être affectée par l'exposition ni pas l'issue, particulièrement, ne doit pas être une variable intermédiaire dans le chemin causal entre l'exposition et l'issue. L'ampleur du biais de confusion dépend de la magnitude l'effet, de la force l'association et de la prévalence du facteur confondant.(36)

Le sous-type de biais de confusion le plus commun en pharmacoépidémiologie est le biais d'indication.(36) Le biais de confusion provient du fait que la raison pour laquelle est un médicament est indiqué pour un certain patient est associé à l'issue d'intérêt. Le facteur de confusion est donc l'indication, où les patients indiqués pour différents traitements diffèrent entre eux. Ce biais englobe la variante biais d'allocation (*channelling*) et biais par sévérité.(36) Par exemple, une étude pourrait faussement trouver un taux de mortalité plus élevé chez les patients diabétiques qui reçoivent un inhibiteur de la dipeptidyl peptidase-4 (DPP-4) plutôt que la metformine. Ceci pourrait être dû au fait qu'en général les patients qui reçoivent un inhibiteur de la DPP-4 souffrent en plus grande proportion d'insuffisance rénale.(57) Un autre sous-type de biais de confusion fréquent est le biais lié à l'observance thérapeutique. En effet, les raisons de l'observance d'un patient sont souvent représentatives de ses caractéristiques personnelles où il présente de bonnes habitudes de vie et meilleure santé générale.

Les facteurs confondants sont généralement classés en quatre types : 1) connus et mesurés, 2) connus et non mesurés, 3) inconnus et mesurés et 4) inconnus et non mesurés.(36) À noter, les confondants connus mais non mesurés sont un problème typique des analyses de banques de données, mais sont moins présents dans les études collectant des données primaires. Avant de choisir une méthode de contrôle des confondants, il est essentiel de correctement identifier ces facteurs de confusion. En effet, il existe plusieurs types de variables avec différentes relations

structurales avec l'exposition et l'issue, pouvant causer un biais de confusion si incorrectement contrôlés. Les principales variables incluent le confondant, le médiateur, le modificateur d'effet et le collisionneur (*collider*).<sup>(36)</sup> Le confondant affecte l'exposition et l'issue, mais n'est pas un médiateur, c'est-à-dire un intermédiaire sur le chemin causal entre l'exposition et l'issue. Un médiateur est donc influencé par l'exposition et à son tour affecte l'issue. Le modificateur d'effet affecte la direction ou la force de la relation entre l'exposition et l'issue, en modifiant l'effet et créant ainsi de l'interaction. Le collisionneur quant à lui est un effet commun produit par l'exposition et l'issue. Alors que l'on doit contrôler pour un facteur confondant, il est incorrect de le faire pour un médiateur ou un collisionneur.<sup>(36, 37)</sup> Pour ce qui est du modificateur d'effet, la stratification est recommandée, c'est pourquoi les analyses de sous-groupes sont très fréquentes pour l'étude de l'hétérogénéité de l'effet de traitement (*heterogeneity of treatment effect / treatment effect modification*).<sup>(36, 37)</sup> La Figure 1 représente la relation structurale de chaque variable à l'aide d'un graphique acyclique dirigé (*directed acyclic graph*).

**Figure 1. Représentation de la relation structure des covariables**

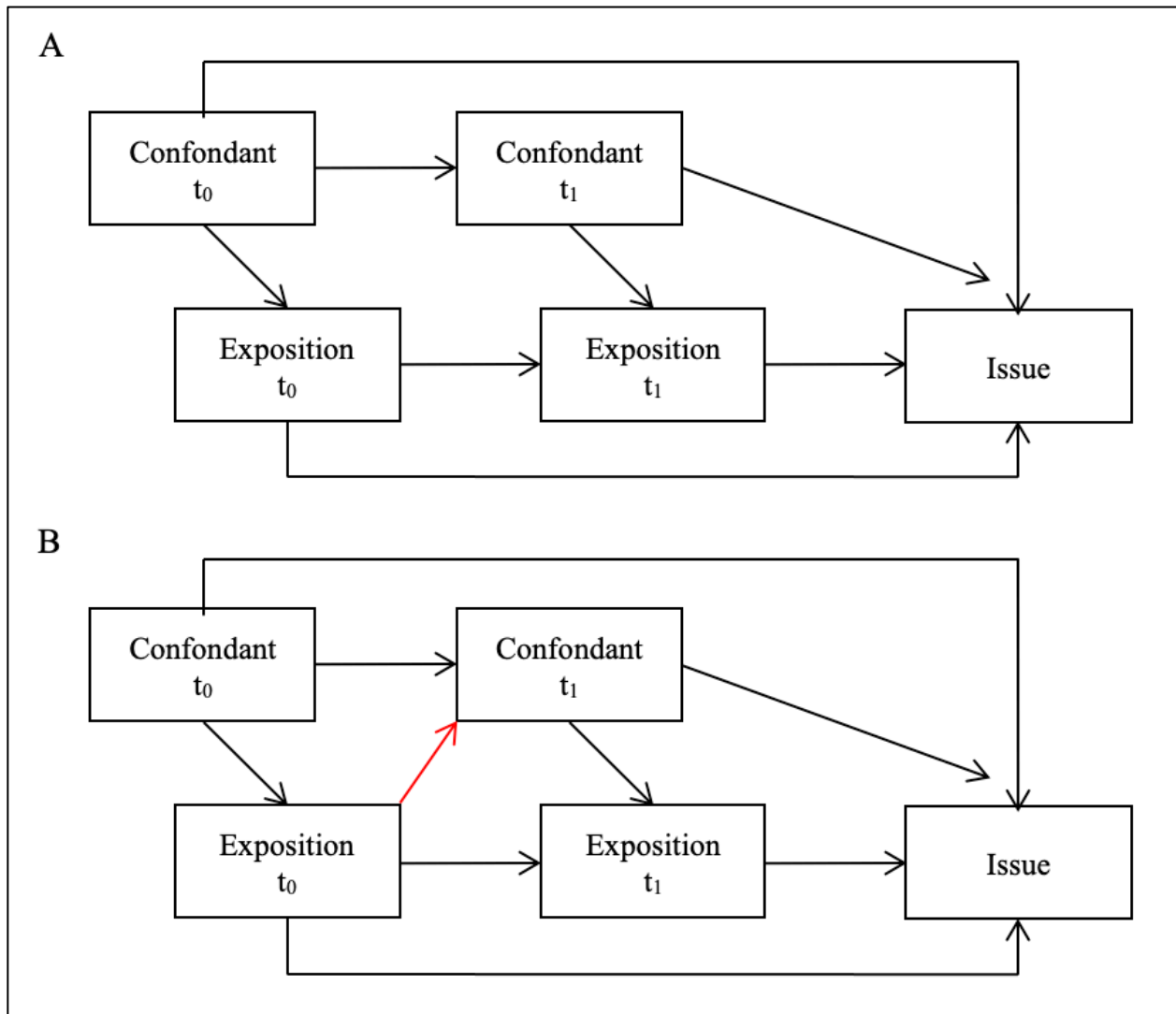


Adapté de Rothman et al.<sup>(36)</sup>



La mesure des confondants à l'entrée dans l'étude (*baseline*) est adéquate lorsque les confondants ne changent pas dans le temps (p.ex. sexe). Cependant, lorsqu'un confondant varie dans le temps (*time-varying confounding*), des mesures subséquentes sont nécessaires. Prudence est alors de mise dans le choix des méthodes de contrôle, car un biais pourrait s'en suivre.(36, 37) Plusieurs méthodes analytiques adaptées aux issues à mesures répétées sont également appropriées pour les expositions et les confondants qui varient dans le temps.(36, 37) Par contre, si le confondant mesuré après le début de l'exposition est influencé par cette exposition antérieure, ajuster pour ce facteur à l'aide des méthodes traditionnelles causera un biais. En effet, lorsque l'exposition influence le confondant dans le temps, ce dit confondant devient en fait un médiateur sur le chemin et des techniques d'ajustement plus avancées doivent être utilisées telles que la pondération en fonction de la probabilité inverse des modèles structuraux marginaux (*inverse probability weighting of marginal structural models*) ou l'estimation g (*g-estimation*).<sup>(58)</sup> La pondération de la probabilité inverse permet de générer une pseudo-population dans laquelle les expositions sont indépendantes des confondants, permettant ainsi l'estimation des paramètres des modèles structuraux marginaux.<sup>(58)</sup> Quant à la formule g, elle modélise la densité conjointe des données observées afin de générer des issues potentielles sous différents scénarios d'exposition.<sup>(58)</sup> Ces deux méthodes sont dites moins restrictives que les méthodes de régression traditionnelles qui sont soumises à de nombreuses conditions et ne permettent donc pas de correctement tenir compte des confondants qui sont affectés par une exposition antérieure (*feedback relationship*).<sup>(58)</sup> Une schématisation d'un confondant influencé par l'exposition antérieure est présentée à la Figure 2.

Figure 2. Représentation du confondant dépendant du temps



A : Confondant dépendant du temps non affecté par l'exposition

B : Confondant dépendant du temps affecté par l'exposition antérieure

Adapté de Burcu et al.(59)

Il y a de nombreuses méthodes de contrôle des confondants, soit au devis soit à l'analyse. À noter qu'il y aura confusion résiduelle si certains confondants sont incorrectement ou non mesurés, sauf dans le cas d'une randomisation qui protège contre le biais de confusion. La

randomisation, c'est-à-dire l'assignation aléatoire de l'exposition, est la meilleure méthode de contrôle de confondants. En effet, lorsque la taille d'échantillon est suffisante, la randomisation assure une distribution similaire des confondants connus et inconnus entre les groupes d'exposition. Toutefois, la randomisation n'est pas souvent réalisable en pratique réelle. Les approches quasi-expérimentales minimisent également la confusion, telles que l'expérimentation naturelle (répartition aléatoire de cause naturelle) et la régression avec discontinuité (création d'un contrefactuel autour d'un seuil prédéterminé d'attribution de l'exposition).(60) Un exemple typique d'expérimentation naturelle serait l'étude de deux régions similaires qui adoptent des politiques de santé différentes (p.ex. deux états distincts à l'intérieur des États-Unis). Quant à la régression avec discontinuité, elle nécessite que l'exposition soit assignée en fonction d'une variable continue, par exemple, les patients avec une pression artérielle au-delà d'un seuil prédéterminé seraient traités avec des antihypertenseurs, alors que les patients sous le seuil ne seraient pas traités. Dans un devis de régression avec discontinuité, seuls les patients se trouvant près du seuil d'attribution seraient inclus, c'est-à-dire les patients légèrement sous et au-delà du seuil critique. Il est attendu que ce groupe de patients présentent des caractéristiques similaires répliquant ainsi l'effet de randomisation.(60) Lorsque l'exposition varie chez un même patient, l'utilisation de modèles pour données longitudinales permet de contrôler pour les facteurs confondants individuels qui ne varient pas dans le temps.

Les principales méthodes de contrôle des confondants connus et mesurés sont l'appariement, la restriction, la stratification, le score de propension et l'analyse de régression.(36) L'appariement consiste en l'identification d'une paire d'individus qui sont similaires par rapport au facteur confondant à contrôler, mais qui diffèrent au niveau du statut d'exposition. La restriction consiste à restreindre les critères d'éligibilité aux patients présentant un certain facteur confondant. La stratification vise à séparer les patients en strates présentant différent statut au niveau de la variable confondante. La régression permet d'estimer l'association entre l'exposition et l'issue en considérant simultanément l'effet des facteurs confondants lors de l'ajustement statistique.(36) Le score de propension est une méthode répandue pour la recherche d'efficacité comparative à l'aide de banques de données administratives. Le score de propension estime la probabilité d'un individu à recevoir un traitement A plutôt que B conditionnelle à un ensemble de caractéristiques observées

liées au choix de traitement.(61) Cette technique requiert tout d'abord qu'un modèle de régression estimant le score de propension soit créé, dans lequel le choix du traitement à l'étude est la variable dépendante et les facteurs confondants sont des variables indépendantes. Ce modèle permet par la suite de calculer le score de propension propre à chaque individu. L'ajustement des facteurs confondants à l'aide des scores de propension peut finalement être réalisé via diverses méthodes, incluant l'appariement, la stratification, la régression et la pondération de la probabilité inverse (*inverse probability weighting*).(61)

Lorsque les confondants sont connus mais non mesurés, l'obtention d'information additionnelle peut être réalisée à l'aide d'étude de validation interne (sous-échantillon de la population) ou de validation externe (population distincte).(62) L'étude de validation interne (*two-stage sampling*) vise à collecter de l'information sur les confondants manquants (non mesurés), par exemple en recontactant la population à l'étude. Étant donné les ressources nécessaires (temps et argent), les variables manquantes sont souvent colligées uniquement pour un sous-échantillon de la population incluse.(62) À noter que la qualité des données doit être prise en considération, par exemple la possibilité de biais de rappel.(62) L'étude de validation externe collecte de l'information sur les confondants non mesurés dans une population distincte de la population à l'étude, par exemple, en utilisant des données d'études publiées dans la littérature. La validité et la représentativité des données doivent être considérées.(62) L'utilisation d'une variable instrumentale est un moyen reconnu de contrôle des confondants connus et inconnus en approximant le processus de randomisation. Par contre, la variable instrumentale doit être associée à l'exposition sans toutefois avoir d'impact sur l'issue, ce qui limite donc le type de variable rencontrant ces critères. Le score de propension à hautes dimensions est une méthode statistique qui permettrait d'ajuster pour des proxys des facteurs confondants inconnus.(63) Tel que pour le score de propension standard, un modèle de régression est créé, mais incluant cette fois-ci des variables identifiées à l'aide d'un algorithme automatisé. Cette approche empirique consiste en sept étapes : sélection des hautes dimensions, identification des variables dans les dimensions, évaluation de la récurrence des variables, détermination du potentiel d'ajustement des variables, sélection des variables à inclure dans le modèle, estimation du score de propension à hautes dimensions de chacun des individus et finalement, estimer l'association étudiée en ajustant à l'aide

des scores de propension à hautes dimensions.(63) L'avantage significatif de cette méthode est l'ajustement des confondants inconnus, contrairement au score de propension standard.(63)

La comparaison de groupes multiples peut être utile pour détecter la présence d'un confondant non mesuré.(64) Par exemple, pour une comparaison entre un traitement A et B à l'aide des réclamations d'une compagnie d'assurance privée, l'ajout d'un groupe de patients recevant également le traitement B, mais provenant d'un autre type d'assurance (p.ex. employés gouvernementaux) pourrait détecter la présence de confusion si l'effet entre A et B est différent à travers les deux groupes recevant B. La présence d'un confondant non mesuré peut également être testée à l'aide de l'ajout d'issues de contrôle (*negative controls*).(64) L'issue de contrôle est une issue non affectée par l'exposition, où la présence d'un effet observé serait associée à la présence de confusion. Plutôt que de simplement mentionner la possibilité de confusion résiduelle, les analyses de sensibilités permettent de discuter de la présence d'un facteur confondant inconnu de façon quantitative.(64) Il convient de noter que les analyses de sensibilités discutées ici font référence à l'étude d'une association entre l'exposition et l'issue, telle qu'en pharmacoépidémiologie, et sont un concept distinct des analyses de sensibilité des évaluations économiques présentées au chapitre 2.

Il existe plusieurs types d'analyses de sensibilité, variant de simples estimations à de complexes simulations.(65) L'approche d'éventail consiste à explorer la variation dans l'association observée en changeant la force de l'association d'un confondant et sa distribution dans les groupes d'exposition (*array approach*). L'approche par exclusion (*rule-out approach*) détermine la force d'un confondant inconnu nécessaire pour changer l'interprétation de l'association observée. Les approches bayésiennes et Monte Carlo sont des techniques plus récentes et sophistiquées qui simulent un ensemble de scénarios testant la présence de confusion résiduelle. Parmi les nouvelles analyses automatisées se trouve également la *E-value*, une programmation standardisée basée sur l'approche par exclusion (*rule-out approach*).(66, 67)

La programmation permet d'obtenir deux *E-values* qui testent le risque associé à la présence d'un confondant inconnu sur les effets estimés (*E-value for effect estimates*) et sur l'intervalle de confiance (*E-value for confidence interval*).<sup>(67)</sup> La *E-value* est une variable continue sur une échelle de ratio de risque (de 1 à l'infini) qui représente l'association d'un confondant inconnu avec l'exposition et l'issue, où une valeur plus élevée réduit le doute sur la présence de confusion résiduelle. À titre d'exemple théorique, une étude évaluant l'impact de l'exposition  $X$  sur l'issue  $Y$ , en ajustant pour  $C$  (confondants connus), obtient un effet  $\beta_x$  avec un intervalle de confiance à 95% de  $\beta_{inf} - \beta_{sup}$ . La *E-value* simule un facteur inconnu  $U$  et détermine la force minimale d'association que devrait avoir  $U$  avec à la fois  $X$  et  $Y$  pour enlever l'effet observé  $\beta_x$  et pour que l'intervalle de confiance inclut la valeur nulle, tout en ajustant pour les covariables  $C$ .<sup>(67)</sup> À titre d'exemple pratique, une étude utilisant une banque de données administratives du Royaume-Uni a démontré que le méloxicam, un anti-inflammatoire non stéroïdien, augmentait le risque d'infarctus du myocarde de 38% (rapport de cotes ajusté 1.38).<sup>(68)</sup> La *e-value* pour les effets estimés était de 2, signifiant que la force de l'association d'un confondant inconnu avec l'exposition et l'issue devrait être d'au moins 2 afin de changer la conclusion de l'étude. Les auteurs ont conclu que l'existence d'un tel confondant était peu probable étant donné que cette valeur est bien au-delà des rapports de cotes des confondants connus, et ce, dans leur étude ainsi que dans la littérature.<sup>(68)</sup>

## CHAPITRE 4 : EXEMPLE DE CAS – DIABÈTE DE TYPE 2

---

### 4.1 Prise en charge du diabète de type 2

Le DMT2 implique généralement une diminution de la production d'insuline et/ou une résistance du corps à l'action de l'insuline.(57) Ces phénomènes mènent à l'hyperglycémie, c'est-à-dire une augmentation du taux de sucre dans le sang, qui à long terme augmente significativement le risque de complications vasculaires. Le DMT2 est incurable et le but des thérapies actuelles est donc de contrôler la glycémie afin de prévenir/retarder de graves complications potentiellement mortelles.(57) Afin d'assurer un contrôle glycémique adéquat, plusieurs thérapies sont disponibles, lorsque la diète et l'exercice physique sont insuffisants. La metformine est la première ligne de traitement antidiabétique et il existe de nombreuses classes thérapeutiques pour les lignes de traitement suivantes : les inhibiteurs de la dipeptidyl peptidase-4 (DPP-4), les agonistes des récepteurs du peptide-1 apparenté au glucagon (GLP-1), les inhibiteurs du cotransporteur sodium-glucose de type 2 (SGLT-2), les inhibiteurs de l'alpha-glucosidase, les thiazolidinediones (TZD) et les sécrétagogues (sulfonylurées et méglitinides). Lorsque les antihyperglycémiantes ne permettent pas d'atteindre les cibles thérapeutiques, l'initiation d'une thérapie à l'insuline est recommandée. L'insuline est très efficace pour réduire la glycémie, à un point tel qu'un risque accru d'hypoglycémie s'en suit.(57) L'hypoglycémie se définit comme une baisse du taux de sucre dans le sang (au-dessous de 4 mmol/L) et peut être accompagnée de symptômes de différentes intensités.(69) Les diabétiques qui sont traités à l'insuline et/ou aux sécrétagogues de l'insuline sont à risque d'effets indésirables tels que l'hypoglycémie. Les symptômes neurogènes sont les premiers à apparaître et incluent les tremblements, picotements, palpitations, transpiration, anxiété, faim et nausée. Lorsque le cerveau manque de glucose, les symptômes neuroglycopéniques peuvent se manifester, incluant les troubles de concentration, maux de tête, étourdissements, confusion, faiblesse, somnolence, vision embrouillée et troubles de la parole. Les comportements et habitudes de vie du patient peuvent influencer l'occurrence d'évènements hypoglycémiques, notamment l'adhérence au traitement, la prises de mesures glycémiques, l'alimentation, la consommation d'alcool et l'exercice physique. L'hypoglycémie non sévère peut être résolue par le patient lui-même en suivant la règle 15/15 : consommation d'un

aliment qui fournit 15 grammes de glucides suivie d'une mesure de la glycémie après 15 minutes, à répéter jusqu'à rétablissement du niveau de glucose dans le sang.(69) Lorsque l'hypoglycémie est sévère, mais que la personne est toujours consciente, la même règle s'applique mais avec une ingestion de 20 grammes de glucose.(69) Lorsque la personne est inconsciente, l'administration de glucagon est recommandée. Un épisode hypoglycémique est catégorisé comme sévère lorsque le patient a besoin de l'assistance d'une autre personne.(69) Bien que les diabétiques traités à l'insuline et/ou aux sécrétagogues de l'insuline soient à risque d'hypoglycémie, l'article 3 de cette thèse se concentre seulement sur les patients traités à l'insuline. Cette décision a été prise puisque le but ultime est de servir à la modélisation économique. Au cours des cinq dernières années, l'ACMTS a évalué neuf évaluations économiques en DMT2, dont trois étaient des ACU pour des thérapies à l'insuline.(70-72) Considérant l'intérêt croissant pour des données pharmacoéconomiques chez les patients avec DMT2 traités à l'insuline, il y a un besoin de données robustes dans cette population, particulièrement avec le coût élevé de l'insuline par rapport à d'autres traitements antidiabétiques.(70-72)

La glycémie est au cœur du contrôle du diabète et l'atteinte des objectifs thérapeutiques se mesurent à l'aide de l'hémoglobine glyquée (HbA1c) qui représente la moyenne glycémique des 8 à 12 dernières semaines.(57) Pour la majorité des patients, la cible thérapeutique est de maintenir les valeurs de l'HbA1c sous 7.0%. Chez les patients souffrant d'hypoglycémie, les bénéfices du contrôle glycémique intensif sont diminués face aux risques de l'hypoglycémie, où les valeurs cibles de l'HbA1c peuvent être augmentées jusqu'à 8.5%.(57) L'HbA1c est un biomarqueur intermédiaire du contrôle du diabète et de l'efficacité du traitement qui permet de suivre l'évolution du DMT2 au quotidien. En effet, une diminution de l'HbA1c est associée à un risque réduit de complications cliniquement significatives, telles que les complications microvasculaires et cardiovasculaires.(73) Les recommandations de l'association canadienne du diabète sur les cibles thérapeutiques du contrôle glycémique ont été basées entre autres sur les évidences de cinq essais cliniques majeurs qui ont évalué l'effet d'un contrôle glycémique intensif sur les complications diabétiques.(73) Les cinq essais cliniques sont décrits au Tableau 5, où les résultats démontrent que la plupart des patients bénéficient d'une réduction de l'HbA1c. Toutefois, les cibles thérapeutiques doivent être personnalisées selon les caractéristiques des patients.(73)



L'association canadienne du diabète recommande que les cibles de l'HbA1c soient augmentées jusqu'à 8.5% chez les patients souffrant d'hypoglycémie sévère, les patients ayant une espérance de vie limitée, les patients âgés et fragiles, ainsi que les patients avec dépendance fonctionnelle ou souffrant de démence.(73) En effet, certains essais cliniques ont démontré qu'une thérapie intensive du contrôle de l'HbA1c mène à un risque accru d'hypoglycémie sévère augmentant le taux de mortalité chez ces patients.(74-76)

**Tableau 5. Essais cliniques évaluant l'effet d'un contrôle glycémique intensif**

	ACCORD (74)	ADVANCE (76)	DCCT (77)	UKPDS (78)	VADT (75)
Nombre de patients	10 251	11 140	1 394	3 277	1 791
Temps de suivi	5 ans (extension) + 3,5 ans (initial)	5 ans (extension) + (5 ans (initial))	10 ans (extension) + 6,5 ans (initial)	10 ans (extension) + 10 ans (initial)	4 ans (extension) + 5,6 ans (initial)
Type de diabète	Type 2	Type 2	Type 1	Type 2	Type 2
Population	Patients avec facteurs de risque ou antécédent d'évènements CV (durée DMT2 10 ans)	Patients avec facteurs de risque ou antécédent d'évènements CV (durée DMT2 8 ans)	Patients nouvellement diagnostiqués ou avec antécédents micro	Patients nouvellement diagnostiqués	Patients avec facteurs de risque ou antécédent d'évènements CV (durée DMT2 12 ans)
Age moyen	62 ans	66 ans	45 ans	63 ans	60 ans
Intervention	Thérapie intensive avec cible HbA1c <6% (combinaison insuline et agents oraux)	Thérapie intensive avec cible HbA1c <6,5% (SU +/- metformine)	Thérapie intensive avec cible HbA1c <6% (insuline 3-4 injections/jour)	Thérapie intensive (metformine, patients obèses : SU ou insuline)	Thérapie intensive avec cible HbA1c <6% (insuline et/ou agents oraux)
Comparateur	Thérapie standard avec cible HbA1c 7%-7,9%	Thérapie standard avec cible HbA1c ~7%	Thérapie standard (insuline 1-2 injections/jour)	Thérapie standard (diète)	Thérapie standard (objectif intensive vs standard réduction de l'HbA1c de 1,5%)
Réduction de l'HbA1c	8,1% → 6,4% vs 7,5%	7,5% → 6,5% vs 7,3%	8,8% → 7,4% 9,1% → 7,8%	7% → 7-7,4% 7% → 7,9-7,8%	9,4% → 6,9% vs 8,4%
Effet sur les complications diabétiques	<ul style="list-style-type: none"> <li>• Évènements CV (initial) : NS</li> <li>• Décès toute cause : HR=1,22 (arrêt prématuré)</li> <li>• Évènements CV et décès toute cause (extension) : NS</li> </ul>	<ul style="list-style-type: none"> <li>• Néphropathie : RRR=21%</li> <li>• Évènements CV et décès toute cause (initial et extension) : NS</li> </ul>	<ul style="list-style-type: none"> <li>• Évènements CV : RRR=42%</li> <li>• IM, AVC, décès cardiovasculaire : RRR=57%</li> <li>• Décès toute cause : RRR=33%</li> </ul>	<ul style="list-style-type: none"> <li>• IM : RRR=15% (SU-insuline) RRR=33% (metformine)</li> <li>• Décès toute cause : RRR=13% (SU-insuline) RRR=27% (metformine)</li> </ul>	<ul style="list-style-type: none"> <li>• Évènements CV (initial) : NS</li> <li>• IM, AVC, insuffisance cardiaque, amputation ou décès CV (extension) : RRR=17%</li> </ul>

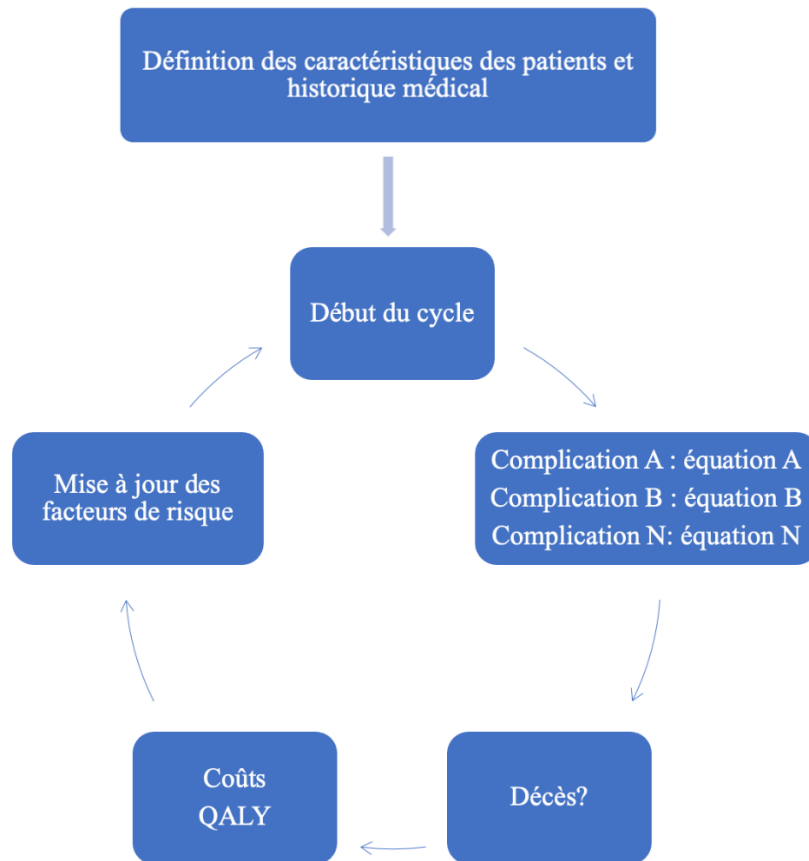
ACCORD: Action to Control Cardiovascular Risk in Diabetes; ADVANCE: Action in Diabetes and Vascular Disease: Preterax and Diamicon MR Controlled Evaluation; AVC: Accident vasculaire cerebral; CV: Cardiovasculaires; DCCT: Diabetes Control and Complications Trial; HR: Hazard ratio; RRR: Relative risk reduction; IM: Infarctus du myocarde; NS: Non statistiquement significatif; SU: Sulfonylurée; UKPDS: United Kingdom Prospective Diabetes Study; VADT: Veterans Affairs Diabetes Trial

La prise en charge du diabète représente un fardeau économique important. En effet, de par le nombre de patients à traiter, la continuité des traitements et le coût de chacun des traitements, le diabète constitue l'une des conditions les plus coûteuses pour les systèmes de santé. Ainsi, dans le choix des médicaments à rembourser pour le diabète, il est important de tenir compte de leur impact économique. Les évaluations économiques servent d'outil afin de déterminer la valeur économique des nouvelles thérapies. Tel que décrit au Chapitre 2, les évaluations économiques, notamment les ACU, ont souvent recours à la modélisation afin de refléter le cheminement clinique des patients dans la maladie d'intérêt.

## **4.2 Modélisation économique en diabète de type 2**

Les ACU sont souvent réalisées au moyen d'un modèle décisionnel qui consiste en une simplification d'une maladie étudiée nécessitant plusieurs hypothèses afin de refléter le plus justement la progression des patients dans le temps.(3) Généralement, la modélisation en DMT2 utilise une technique de microsimulation, où la simulation des patients diabétiques débute par la définition des caractéristiques individuelles des patients, leurs caractéristiques cliniques et leur historique médical.(18) À chaque cycle du modèle décisionnel, l'occurrence de complications diabétiques sera déterminée en fonction d'équations qui tiennent compte de la variation individuelle des probabilités d'événements. À chaque cycle, les coûts et les *QALY* sont calculés, certains patients vont également décéder à la suite de leurs complications. Les facteurs de risque individuels sont ensuite mis à jour et les patients débutent un nouveau cycle et ainsi de suite jusqu'à la fin de l'horizon temporel.(18) Une représentation d'un modèle économique est illustrée à la Figure 3.

**Figure 3. Exemple de modèle économique en diabète de type 2**



Les équations servant à l'estimation des probabilités de complications diabétiques incluent un ensemble de données démographiques et de facteurs de risque.(18) Les principaux facteurs sont l'âge, le sexe, l'ethnicité, la durée du diabète, le tabagisme, l'indice de masse corporelle, les antécédents de complications diabétiques, le taux de cellules blanches et d'hémoglobine, l'albuminurie, la présence de maladie vasculaire périphérique ou fibrillation auriculaire, le rythme cardiaque, le taux de filtration glomérulaire, le taux de cholestérol, la pression artérielle systolique et enfin l'HbA1c.(18) Le nombre importants de facteurs influençant l'occurrence de complications diabétiques illustrent bien que la prise en charge du DMT2 ne doit pas uniquement tenir compte de l'HbA1c. C'est pourquoi l'association canadienne du diabète recommande une approche multifactorielle de réduction du risque, c'est-à-dire qu'en plus de la réduction de l'HbA1c, des stratégies doivent également être mises en place enfin de contrôler la pression sanguine et le taux de lipides, et d'encourager l'adoption de saines habitudes de vie (cessation du tabagisme, activité physique, etc.).(73)

Ainsi, la modélisation économique en DMT2 tente de refléter l'impact des thérapies sur les complications diabétiques, ce qui peut s'avérer complexe. La modélisation doit tenir compte des nombreux facteurs de risques impliqués ainsi que de l'occurrence de diverses complications inter-reliées sur un long horizon temporel. En plus de l'efficacité, la modélisation doit aussi refléter les effets indésirables des antihyperglycémiantes, où les plus communément inclus dans les modèles sont la prise de poids et l'hypoglycémie.(18)

Il existe différents modèles économiques en DMT2 et une description des modèles largement utilisés, notamment ceux qui font partie du Mount Hood Diabetes Challenge Network, est présentée au Tableau 6.(79-84) Le Mount Hood Diabetes Challenge Network est un réseau fondé en 1999 dédié à promouvoir l'échange d'idées et d'informations entre les chercheurs en pharmacoéconomie développant et utilisant la modélisation économique en diabète.(85) Parmi les modèles présentés au Tableau 6, la source principale de données des équations estimant les probabilités des complications diabétiques est l'étude United Kingdom Prospective Diabetes Study (UKPDS).(78) De nombreux modèles incluent également des données complémentaires de cohortes épidémiologiques et de registres tels que Framingham(86), Fremantle(87), Swedish National Diabetes registry(88), etc.

La microsimulation est communément utilisée en DT2M, c'est-à-dire qu'au lieu de faire progresser une cohorte entière à travers des états de santé les patients progressent individuellement selon leurs facteurs de risque et historique.(79, 81-84) Les principales complications vasculaires incluses dans les modèles sont l'infarctus du myocarde, les maladies coronariennes, l'insuffisance cardiaque, l'accident vasculaire cérébral, l'insuffisance rénale, la cécité et l'amputation. Plusieurs modèles incluent également des états intermédiaires tels que la néphropathie, la rétinopathie et la neuropathie. Parmi les autres complications potentiellement incluses dans les modèles se trouvent la maladie périphérique vasculaire, l'œdème maculaire, la cataracte, l'ulcère, etc.(79-84)

**Tableau 6. Modèles économiques fréquemment utilisés en DMT2**

	Cardiff (82)	CDC-RTI (80)	ECHO-T2DM (83)	IQVIA-CDM (81)	MMD (84)	UKPDS-OM (79)
Pays d'origine	Royaume-Uni	États-Unis	Suède	Suisse	États-Unis	Royaume-Uni
Source principale de données cliniques	UKPDS (78)	UKPDS (78)	UKPDS (78) ADVANCE (76)	UKPDS (78) ADVANCE (76)	UKPDS (78)	UKPDS (78)
Technique de modélisation	Microsimulation	Markov	Hybride : Markov- Microsimulation	Hybride : Markov- Microsimulation	Hybride : Markov- Microsimulation	Microsimulation
Complications macrovasculaires						
Infarctus du myocarde	√	√	√	√	√	√
Maladies coronariennes	√	√	√	√	√	√
Accident vasculaire cérébral	√	√	√	√	√	√
Insuffisance cardiaque	√		√	√	√	√
Maladie vasculaire périphérique				√		
Complications microvasculaires						
Insuffisance rénale	√	√	√	√	√	√
Protéinurie (néphropathie)		√	√	√	√	

	Cardiff (82)	CDC-RTI (80)	ECHO-T2DM (83)	IQVIA-CDM (81)	MMD (84)	UKPDS-OM (79)
Cécité (perte de vision sévère)	√	√	√		√	√
Rétinopathie		√	√	√	√	
Œdème maculaire			√	√	√	
Cataracte				√		
Amputation	√	√	√	√	√	√
Neuropathie		√	√	√	√	
Ulcère diabétique			√	√		√

ADVANCE: Action in Diabetes and Vascular Disease: Preterax and Diamicon MR Controlled Evaluation; CDC: Centers for Disease Control and Prevention; ECHO: Economic and Health Outcomes Model for T2DM; IQVIA-CDM: IQVIA Core Diabetes Model; Michigan Model for Diabetes; UKPDS-OM: United Kingdom Prospective Diabetes Study Outcomes Model

Chaque des complications microvasculaires et macrovasculaires incluses dans un modèle économique en DMT2 doit être associée avec un coût et un score d'utilité, permettant ainsi de comparer l'efficience à long terme de différents agents antihyperglycémiant. Il est également important d'attribuer un coût et un score d'utilité aux effets indésirables afin de capturer l'ensemble des effets (efficacité et innocuité) des thérapies antidiabétiques.(18) Tel que mentionné précédemment, l'hypoglycémie en un enjeu sérieux associé à la thérapie intensive du contrôle de la glycémie à un point tel que les bénéfices cliniques peuvent en être annulés.(73) Malgré l'importance de capter l'impact de l'hypoglycémie dans la modélisation économique, l'ACMTS rapporte que les données d'utilité associée à l'hypoglycémie sont limitées.(18) La section 4.3 vise donc à répertorier les intrants pharmacoéconomiques associés à l'hypoglycémie aux fins de modélisation économique dans un contexte canadien.

### **4.3 Impact de l'hypoglycémie sur l'utilité et la productivité**

Dans le cadre cette thèse, deux revues de la littérature ont été conduites afin d'identifier les valeurs d'utilité et de productivité associées à l'hypoglycémie, en ciblant plus précisément les données canadiennes. Ces revues visaient tout d'abord à répertorier les données d'utilité et de productivité actuellement disponibles aux fins d'évaluations économiques en diabète et sont présentées à la section 4.3.1 et 4.3.3 respectivement. Également, les revues ont servi de support à l'article 3 de cette thèse en identifiant les différents facteurs confondants pouvant influencer l'impact l'hypoglycémie sur l'utilité et la productivité, présentés à la section 4.3.2 et 4.3.4 respectivement.

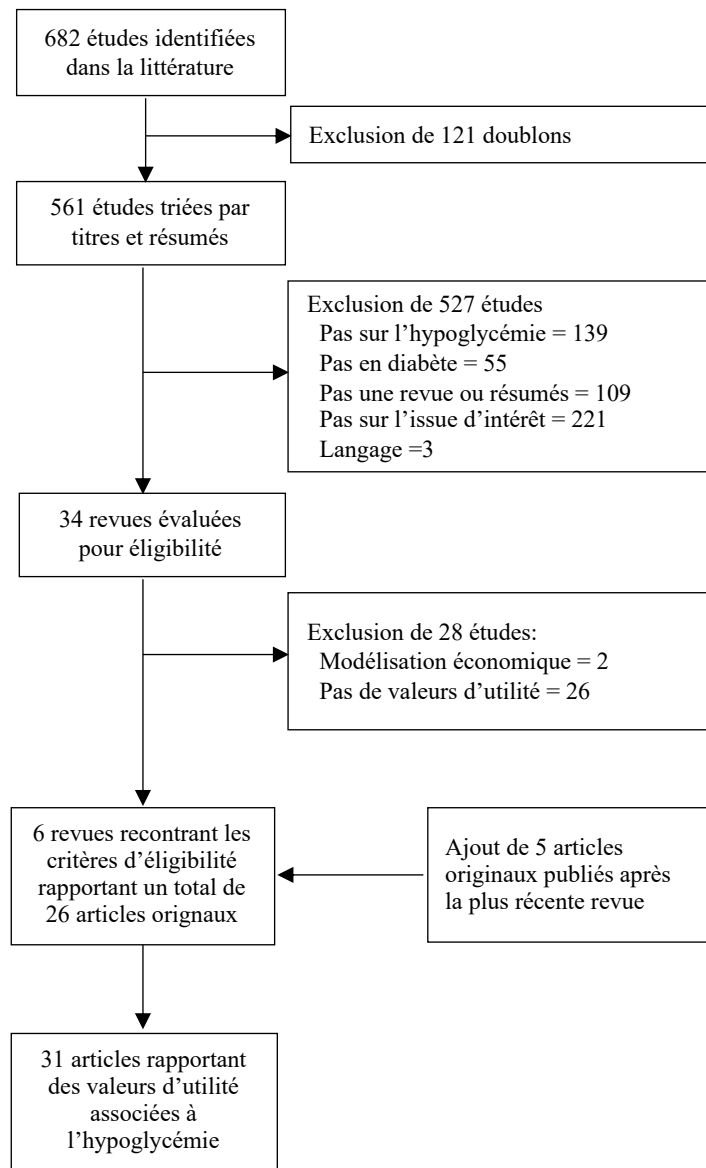
#### **4.3.1 Impact de l'hypoglycémie sur les scores d'utilité : état des connaissances**

Une revue de la littérature a été réalisée à l'aide des banques de données MEDLINE et EMBASE afin d'identifier les valeurs d'utilité associées à l'hypoglycémie. Les mots-clés incluaient le DMT2, l'hypoglycémie et l'utilité. La stratégie de recherche est présentée en Annexe 1. Stratégie de recherche en diabète de type 2 Seules les revues de la littérature ont été



sélectionnées, pour un total de 682 titres. Suite à la lecture des titres et des résumés, 34 revues ont été sélectionnées. À la lecture des publications complètes, seulement 6 revues rapportaient des valeurs d'utilité. Les 6 revues incluses présentaient 26 articles originaux estimant des valeurs d'utilité. Puisque la revue la plus récente avait été publiée en 2019, une recherche des articles originaux publiés après cette date a également été réalisée. En addition aux 26 articles préalablement inclus, 5 articles récents ont été ajoutés pour un total de 31 articles (Figure 4).

**Figure 4. Sélection des études sur les scores d'utilité associés à l'hypoglycémie**



Parmi les 31 articles identifiés, 3 étaient canadiens(89-91) et sont présentés au Tableau 7. À noter que les publications par Evans et al.(89) et Harris et al.(90) consistent en la même étude, présentant les résultats de la cohorte internationale (Canada, Allemagne, Suède, Royaume-Uni et États-Unis) et de la cohorte canadienne respectivement.

**Tableau 7. Données canadiennes d'utilité associées à l'hypoglycémie**

	Evans, 2013 (89) & Harris, 2014 (90)		Levy, 2008 (91)	
<b>Répondants et taille d'échantillon</b>	Population générale (n=1696)	Patients atteints de DTM2 (n=331)	Population générale (n=78)	Patients diabétiques traités à l'insuline depuis au moins 1 an (n=51)
<b>Sexe (hommes)</b>	874 (52%)	171 (52%)	39 (50%)	24 (47%)
<b>Age moyen</b>	46 ans	55 ans	47 ± 16 ans	55 ± 12 ans
<b>Sélection des répondants</b>	Panel en ligne		Annonces dans les journaux	Cliniques du diabète
<b>Mesure de l'hypoglycémie</b>	Vignettes (états fictifs)		Vignettes (états fictifs)	
<b>Méthode de mesure de l'utilité</b>	Arbitrage temporel		Arbitrage temporel	
<b>Résultats</b>	Par épisode par année : <ul style="list-style-type: none"> <li>• Non sévère diurne : -0,0056</li> <li>• Non sévère nocturne : -0,0076</li> <li>• Sévère diurne : -0,0592</li> <li>• Sévère nocturne : -0,0616</li> </ul>	Par épisode par année : <ul style="list-style-type: none"> <li>• Non sévère diurne : -0,0028</li> <li>• Non sévère nocturne : -0,0076</li> <li>• Sévère diurne : -0,0726</li> <li>• Sévère nocturne : -0,0826</li> </ul>	<ul style="list-style-type: none"> <li>• Épisodes rares (trimestriels) : 0,85 (0,61-1,00)</li> <li>• Épisodes intermédiaires (mensuels) : 0,77 (0,38-1,00)</li> <li>• Épisodes fréquents (hebdomadaires) : 0,66 (0,22-1,00)</li> <li>• Disutilité par épisode : -0,0032 (ajusté pour âge, sexe, éducation)</li> </ul>	<ul style="list-style-type: none"> <li>• Épisodes rares (trimestriels) : 0,91 (0,69-1,00)</li> <li>• Épisodes intermédiaires (mensuels) : 0,87 (0,63-1,00)</li> <li>• Épisodes fréquents (hebdomadaires) : 0,75 (0,42-1,00)</li> <li>• Disutilité par épisode : -0,0033 (ajusté pour âge, sexe, éducation, HbA1c, dose d'insuline et type de diabète)</li> </ul>

	<b>Evans, 2013 (89) &amp; Harris, 2014 (90)</b>	<b>Levy, 2008 (91)</b>
<b>Limites</b>	<ul style="list-style-type: none"> <li>• Membres du panel peuvent être différents de la population cible</li> <li>• Taux de participation non rapporté</li> <li>• Étude de vignettes (non privilégiée par payeurs)</li> <li>• Modèle de régression non détaillé</li> </ul>	<ul style="list-style-type: none"> <li>• Répondants (journaux) peuvent être différents de la population cible et les patients en cliniques du diabète peuvent être plus sévères</li> <li>• Groupe de patients diabétiques inclut 49% patients avec diabète de type 1</li> <li>• Taux de participation non rapporté</li> <li>• Étude de vignettes (non privilégiée par payeurs)</li> </ul>

Il n'y a que très peu d'études canadiennes rapportant des données d'utilité associée à l'hypoglycémie, et celles-ci présentent différentes limites, notamment ce sont toutes des études mesurant l'utilité à l'aide d'un arbitrage temporel utilisant des vignettes. Tel que discuté au chapitre 2, ce type d'études n'est pas privilégié par les agences d'évaluation des technologies de la santé qui recommandent plutôt d'utiliser des méthodes de mesure indirecte à l'aide de questionnaires validés (p.ex. EQ-5D).(3, 4) Il n'est donc pas surprenant que dans sa revue thérapeutique sur les différents traitements en DMT2, l'ACMTS identifie une autre source (non canadienne) de valeurs d'utilité à privilégier lors de la modélisation.(18)

La source recommandée est une publication par Currie et al., une étude financée par l'industrie pharmaceutique qui consiste en la combinaison des résultats de deux questionnaires postaux complétés par des patients diabétiques du Royaume-Uni.(92) Il convient de noter que 175 personnes ont répondu aux deux envois postaux, pour lesquels seuls les résultats du deuxième questionnaire ont été analysés. Les décrets d'utilité estimés à l'aide du questionnaire EQ-5D-3L étaient de -0,047 pour un épisode sévère et -0,014 pour un épisode non sévère sur une période de rappel de trois mois.(92) Bien que la publication de Currie et al. serve de référence, les agences d'évaluation des technologies soulignent elles-mêmes le risque de biais de sélection et de biais de rappel.(18, 93) Elles précisent également que les évidences sur l'impact de l'hypoglycémie sur l'utilité sont actuellement limités.(18, 93) L'étude par Currie et al.(92) est présentée au Tableau 8.

**Tableau 8. Impact de l'hypoglycémie sur l'utilité : source privilégiée par l'ACMTS**

	Currie, 2006(92)	
	Questionnaire postal 1	Questionnaire postal 2
<b>Nombre d'envois postaux</b>	1500 personnes	3200 personnes
<b>Année des envois postaux</b>	2000	2004
<b>Région</b>	Cardiff, Royaume-Uni	Cardiff, Royaume-Uni
<b>Devis d'étude</b>	Étude transversale	Étude transversale
<b>Population à l'étude</b>	Patients atteints du diabète de type 1 ou type 2	Patients atteints du diabète de type 1 ou type 2
<b>Sélection des patients</b>	<ul style="list-style-type: none"> <li>• Patients ayant fait un test de glycémie au cours de l'année précédente, identifiés dans une banque de données répertoriant les soins primaires et secondaires</li> </ul>	<ul style="list-style-type: none"> <li>• Patients ayant déjà été hospitalisés, identifiés par code diagnostique (ICD-10) pour le diabète dans leurs dossiers médicaux</li> <li>• Patients non-hospitalisés, auto-rapportant un diagnostic de diabète (<i>outpatient records</i>)</li> </ul>
<b>Mesure des complications diabétiques</b>	<ul style="list-style-type: none"> <li>• Complications auto-rapportées : maladies coronariennes, maladies cérébrovasculaires, pied diabétique, maladies rénales, problèmes de vision</li> </ul>	<ul style="list-style-type: none"> <li>• Complications identifiées par code diagnostique : maladies coronariennes, maladies cérébrovasculaires, pied diabétique, maladies rénales, problèmes de vision</li> </ul>
<b>Mesure de l'hypoglycémie</b>	<p>Définition de l'hypoglycémie</p> <ul style="list-style-type: none"> <li>• Épisode non-sévère : identifié par la présence de symptômes <ul style="list-style-type: none"> <li>-Léger : transpiration, étourdissements, tremblements, vision embrouillée, troubles de concentration, palpitations et/ou maux de tête.</li> <li>-Modéré : confusion, agressivité, tempérament colérique, comportement étrange (impolitesse, rire hystérique, comportement similaire à l'ivresse).</li> </ul> </li> <li>• Épisode sévère : nécessite l'assistance d'une autre personne</li> </ul> <p>Période de mesure de l'hypoglycémie</p> <ul style="list-style-type: none"> <li>• 3 mois précédant la complétion du questionnaire (période de rappel)</li> </ul>	
<b>Mesure de l'utilité</b>	EQ-5D-3L	

<b>Nombre de participants</b>	1305 répondants (taux de réponse 31,4%) 768 patients avec DMT2 (58.9% du total de répondants)
<b>Sexe (hommes)</b>	59,4% des patients DMT2
<b>Age moyen</b>	66,5 ± 12,0 ans (chez les patients DMT2)
<b>Type d'analyse</b>	Régression linéaire multivariée – Variables indépendantes : <ul style="list-style-type: none"> <li>• Épisode hypoglycémique</li> <li>• Age</li> <li>• Indice de masse corporelle</li> <li>• Maladies coronariennes</li> <li>• Maladies cérébrovasculaires</li> <li>• Pied diabétique</li> <li>• Maladies rénales</li> <li>• Problèmes de vision</li> </ul>
<b>Résultats</b>	Décrément d'utilité : <ul style="list-style-type: none"> <li>• Épisode non-sévère : -0,014</li> <li>• Épisode sévère : -0,047</li> </ul>

L'étude de Currie et al(92) présente plusieurs limites. Tout d'abord, le choix des variables indépendantes (facteurs confondants potentiels) du modèle de régression n'a pas été justifié et le risque de confusion résiduelle n'a pas été discuté. De plus, certaines données telles que le diagnostic de diabète, le type de diabète et la présence de complications microvasculaires et macrovasculaires ont été auto-rapportées pour certains patients. Le bas taux de réponse de 31,4% ouvre la porte à un risque de biais de sélection. Également, il n'est pas mentionné si les participants devaient avoir répondu à l'ensemble des questions pour être inclus dans l'étude. Enfin, le devis transversal avec une période de rappel de 3 mois précédant la complétion du questionnaire postal mène à un risque de biais de rappel au niveau de la mesure des épisodes hypoglycémiques. Bref, bien que l'étude de Currie et al(92) soit la source favorisée par les agences d'évaluation des technologies de la santé, il semblerait que ce choix soit plutôt par défaut étant donné le nombre limité d'études. En effet, en plus de présenter de nombreuses limites, l'étude de Currie et al(92) n'est pas canadienne, il est donc possible qu'il y ait des différences entre les populations du Royaume-Uni et du Canada, notamment quant à la transférabilité des ensembles de préférences (*value sets*) entre les juridictions.(22) Somme toute, l'ACMTS recommande que le choix de la

source d'utilité représente le meilleur compromis entre les critères de crédibilité, cohérence et aptitude à l'emploi.(3) À ce jour, les évidences sur l'impact de l'hypoglycémie sur l'utilité sont limitées et la génération de données robustes répondrait à un besoin non comblé. La génération de données canadiennes sur l'impact de l'hypoglycémie, adaptées à la modélisation économique en DMT2, fait donc partie des objectifs de l'article 3 de cette thèse.

#### **4.3.2 Impact de l'hypoglycémie sur les scores d'utilité : facteurs confondants**

Afin de répondre à l'objectif de l'article 3 d'évaluer l'impact de l'hypoglycémie sur l'utilité, la liste des confondants potentiels doit tout d'abord être établie. Tel que discuté à la section 4.3.1, 31 articles (canadiens et non-canadiens) ont été identifiés dans la littérature. Lors de la révision des 31 articles, 9 ont été exclus, car l'effet de l'hypoglycémie sur l'utilité n'avait pas été ajusté et ne rapportaient donc pas de facteurs confondants, pour un total de 22 articles. En plus de la liste des variables, le type de variables (p.ex. continu, dichotomique, catégoriel) a été aussi été extraits. La liste des confondants extraits est présenté au Tableau 9 **Erreur! Source du renvoi introuvable.**, dans lequel le type de variables n'a pas été précisé afin de simplifier la présentation.

Les confondants identifiés vont comme suit : âge, sexe, ethnicité, statut socioéconomique, statut marital, consommation d'alcool, tabagisme, activité physique, utilisation des services médicaux, activités d'autogestion du diabète, durée du diabète, HbA1c, indice de masse corporelle, caractéristiques des thérapies et complications du diabète. Le choix des facteurs confondants potentiels qui seront collectés dans l'article 3 est discuté à la section 6.3.3.

**Tableau 9. Facteurs confondants potentiels pour les valeurs d'utilité**

	Alvarez(94)	Bradley(95)	Currie(92)	Davis(96)	Levy(91)	Lundvist(97)	Marrett(98)	Marrett(99)	McCoy(100)	Meneghini(101)	Pagalos (102)	Praipanaw(103)	Sheu(104)	Shi(105)	Solli(106)	Tabaei(107)	Takahara(108)	Terachi(109)	Vexiau(110)	Wasem(111)	Yfantoplos(112)	Zhang(113)	
Âge	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x
Sexe	x	x			x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Ethnicité										x				x									
Éducation					x		x			x	x		x	x			x	x	x		x	x	x
Revenu										x				x			x	x			x		
Statut d'emploi		x								x	x		x		x		x	x			x		
Statut marital		x									x		x	x				x	x				
Alcool	x									x	x						x	x					
Tabagisme	x	x								x	x		x				x	x	x				
Activité physique	x									x	x		x				x	x	x				
Nombre de visites médicales <sup>1</sup>		x											x		x				x				
Autogestion du diabète <sup>2</sup>		x										x	x	x					x				

	Alvarez(94)	Bradley(95)	Currie(92)	Davis(96)	Levy(91)	Lundvist(97)	Marrett(98)	Marrett(99)	McCoy(100)	Meneghini(101)	Pagalos (102)	Pratpanaw/103	Sheu(104)	Shi(105)	Solli(106)	Tabaei(107)	Takahara(108)	Terauchi(109)	Vexiau(110)	Wasem(111)	Yfantoplos(112)	Zhang(113)
Durée du diabète	x	x				x	x	x	x	x	x		x	x			x		x		x	x
HbA1c	x	x			x	x	x	x		x	x						x		x	x	x	
Indice de masse corporelle	x	x	x	x			x			x	x		x	x	x		x	x	x	x	x	
Type de thérapie		x			x	x		x		x		x	x	x			x		x		x	
Nombre de traitements					x					x							x					
Durée de thérapie										x			x								x	
Complications macrovasculaires <sup>3</sup>	x	x	x	x		x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x
Complications microvasculaires <sup>3</sup>	x	x	x	x		x	x	x	x	x	x	x			x	x	x	x	x	x	x	x

1. Inclut visites chez un médecin de famille, médecin spécialiste, urgence et hospitalisations.

2. Inclut scores de questionnaire sur l'autogestion du diabète ou composantes individuelles telles que la diète, les mesures glycémiques, l'adhérence.

3. Inclut regroupements de complications, complications individuelles et scores de comorbidité.

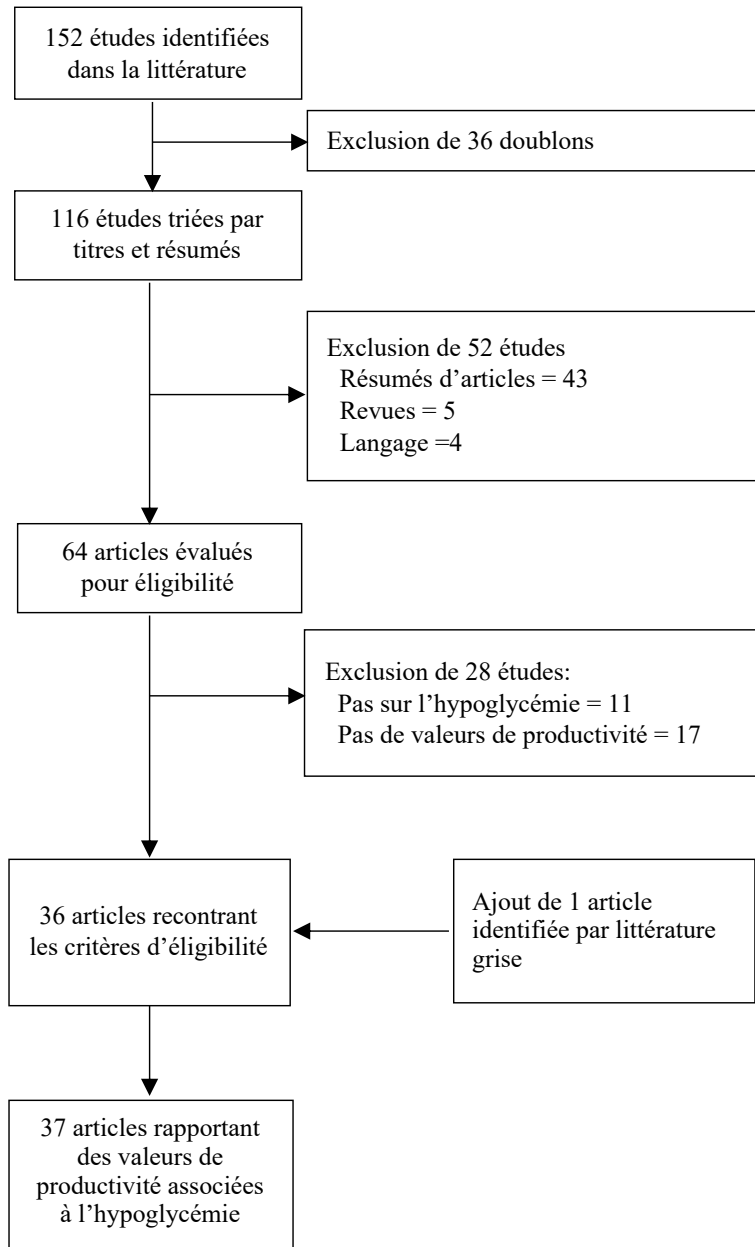


### **4.3.3 Impact de l'hypoglycémie sur la productivité : état des connaissances**

Tel mentionné plus haut, une deuxième revue de la littérature a été conduite afin d'identifier les données de productivité associées à l'hypoglycémie, plus précisément les données canadiennes. Cette revue visait à répertorier les données de productivité actuellement disponibles aux fins d'évaluations économiques en diabète et à supporter l'article 3 en identifiant les différents facteurs confondants pouvant influencer l'impact l'hypoglycémie sur la productivité. La revue a été réalisée à l'aide des banques de données MEDLINE et EMBASE afin d'identifier les pertes de productivité associées à l'hypoglycémie. Les mots-clés incluaient le DMT2, l'hypoglycémie et la productivité. La stratégie de recherche est présentée en Annexe 1. Stratégie de recherche en diabète de type 2

Puisque les publications sur la perte de productivité sont beaucoup moins nombreuses que celles sur l'utilité, les articles originaux ont été directement identifiés dans la littérature plutôt que les revues d'articles. La revue de la littérature a identifié 152 articles dont 36 étaient des doublons, 43 étaient des résumés d'articles, 5 étaient des articles de revue et 4 étaient rédigés dans une langue étrangère. Parmi les 64 articles évalués pour éligibilité, 17 ne rapportaient pas de données de perte de productivité et 11 études n'évaluaient pas l'impact de l'hypoglycémie. En addition aux 36 articles identifiés, 1 article supplémentaire a été ajouté à l'aide des références publiées dans les 5 revues, pour un total de 37 études (Figure 5).

**Figure 5. Sélection des études sur la perte de productivité associée à l'hypoglycémie**



Parmi les 37 articles inclus, 3 études étaient canadiennes et sont présentées au Tableau 10. Il convient de noter qu'étant donné que la perte de productivité est uniquement incluse dans la perspective sociétale. Comme elle ne privilégie pas cette perspective, l'ACMTS n'a pas recommandé de source à privilégier. Dans une étude par Leiter et al(114), un questionnaire a été complété par 133 patients diabétiques, et jusqu'à 9% et 26% des patients ont rapporté avoir manqué du travail ou de l'école suivant un épisode hypoglycémique non sévère et un épisode sévère respectivement. Un sondage en ligne par Brod et al(115) a inclut 150 patients diabétiques qui avaient souffert d'épisodes hypoglycémiques non sévères nocturnes lors du dernier mois. Parmi les 55 travailleurs avec DMT2, 9 (16,4%) ont rapporté en moyenne 3,5h de travail perdu.(115) Plus récemment, O'Reilly et al. ont publié des données de coûts associés à l'hypoglycémie basés sur la cohorte canadienne de l'étude internationale *Hypoglycemia Assessment Tool* (HAT).(116) L'étude HAT est une étude à grande échelle incluant des patients diabétiques traités à l'insuline. L'objectif primaire de cette étude non-interventionnelle de pratique réelle était d'estimer la prévalence des événements hypoglycémiques. L'étude avait un devis hybride, à la fois rétrospective et prospective. La perte de productivité a été estimée uniquement chez les participants ayant un emploi. Parmi les travailleurs, 6,0%, 3,8% et 7,5% ont rapporté une moyenne de 2,9 journées d'absence au travail, 3,8 journées de retard au travail et 1,7 journées raccourcies au travail.(116)

**Tableau 10. Données canadiennes de productivité associées à l'hypoglycémie**

	<b>Leiter, 2005(114)</b>	<b>Brod, 2014(115)</b>	<b>O'Reilly, 2018(116)</b>
<b>Devis d'étude</b>	Étude transversale	Étude transversale	Devis hybride (1 mois prospectif et 1 an rétrospectif)
<b>Répondants</b>	Patients avec diagnostic de diabète de type 1 ou type 2 traités à l'insuline depuis au moins 1 an	Patients avec diagnostic auto-rapporté de diabète et ayant souffert d'au moins un épisode hypoglycémique non sévère nocturne au cours du dernier mois	Patients avec diagnostic de diabète de type 1 ou type 2 traités à l'insuline depuis au moins 1 an

	<b>Leiter, 2005(114)</b>	<b>Brod, 2014(115)</b>	<b>O'Reilly, 2018(116)</b>
<b>Taille d'échantillon</b>	335 patients (total) 133 patients avec DMT2	200 patients (total) 150 patients avec DMT2	498 patients (total) 315 patients avec DMT2
<b>Taux de participation</b>	Non rapporté	83,7%	Non rapporté
<b>Sexe (hommes)</b>	51,0% (cohorte totale)	38,7% (DMT2)	59,7% (DMT2)
<b>Age moyen</b>	59,9 ans (DMT2)	57,2 ± 11,2 ans (DMT2)	60,3 ± 11,0 ans (DMT2)
<b>Sélection des répondants</b>	4 cliniques du diabète, patients recrutés à leur rendez-vous de routine	Panel en ligne (sélection stratifiée : groupe d'âge, type de diabète et statut d'emploi)	Cliniques médicales (types et nombres non détaillés)
<b>Mesure de l'hypoglycémie</b>	Définition de l'hypoglycémie <ul style="list-style-type: none"> <li>• Épisode non-sévère : ≤4,0 mmol/L</li> <li>• Épisode sévère : ≤2,8 mmol/L</li> </ul> Période de mesure de l'hypoglycémie <ul style="list-style-type: none"> <li>• Épisode non-sévère : 1 mois avant questionnaire (rappel)</li> <li>• Épisode sévère : 1 an avant questionnaire (rappel)</li> </ul>	Définition de l'hypoglycémie <ul style="list-style-type: none"> <li>• Épisode non-sévère nocturne: épisode se produisant la nuit et ne nécessitant pas l'aide d'une autre personne, identifié par mesure du glucose sanguin ou symptômes (p.ex. transpiration, confusion)</li> </ul> Période de mesure de l'hypoglycémie <ul style="list-style-type: none"> <li>• 1 mois avant questionnaire (rappel)</li> </ul>	Définition de l'hypoglycémie <ul style="list-style-type: none"> <li>• Épisode non-sévère : ≤3,1 mmol/L et/ou présence de symptômes</li> <li>• Épisode sévère : assistance d'une autre personne</li> </ul> Période de mesure de l'hypoglycémie <ul style="list-style-type: none"> <li>• Épisode non-sévère : 1 mois avant questionnaire (rappel) et 1 mois prospectif</li> <li>• Épisode sévère : 6 mois avant questionnaire (rappel)</li> </ul>
<b>Méthode de mesure de la productivité</b>	Questions développées par les chercheurs	Endicott Work Productivity Scale (EWPS)	Questions développées par les chercheurs
<b>Résultats</b>	Épisode non-sévère (n=97): <ul style="list-style-type: none"> <li>• 14,4% ont demandé à un proche de veiller sur eux</li> <li>• 10,3% ont quitté leur travail/activité plus tôt</li> <li>• 9,3% sont restés à la maison le lendemain</li> </ul> Épisode non-sévère (n=19):	Travailleurs avec DMT2 (n=55) <ul style="list-style-type: none"> <li>• 16,4% ont manqué en moyenne 3,5 heures de travail</li> <li>• 52,7% avaient de la difficulté à se concentrer le lendemain</li> <li>• Score EWPS de présentéisme : 24,1 ±</li> </ul>	Travailleurs avec DMT2 (n=134) <ul style="list-style-type: none"> <li>• 6,0% ont manqué en moyenne 2,9 jours de travail</li> <li>• 3,8% ont eu en moyenne 3,8 jours de retard au travail</li> </ul>

	<b>Leiter, 2005(114)</b>	<b>Brod, 2014(115)</b>	<b>O'Reilly, 2018(116)</b>
	<ul style="list-style-type: none"> <li>• 57,9% ont demandé à un proche de veiller sur eux</li> <li>• 32,0% ont quitté leur travail/activité plus tôt</li> <li>• 26,3% sont restés à la maison le lendemain</li> </ul>	21,6 (équivalent à l'impact de l'arthrite selon les auteurs)	<ul style="list-style-type: none"> <li>• 7,5% ont eu en moyenne 1,7 journées raccourcies au travail</li> </ul>
<b>Limites</b>	<ul style="list-style-type: none"> <li>• Patients recrutés en cliniques du diabète peuvent être plus sévères</li> <li>• Questionnaire de productivité non validé</li> <li>• Devis transversal avec période de rappel de 1 et 12 mois (rétrospectif)</li> <li>• Taux de participation non rapporté</li> <li>• Taux de données manquantes non rapporté</li> </ul>	<ul style="list-style-type: none"> <li>• Membres d'un panel peuvent être différents de la population cible</li> <li>• Questionnaire de productivité non validé (absentéisme)</li> <li>• Devis transversal avec période de rappel de 1 mois (rétrospectif)</li> <li>• Diagnostic de diabète auto-rapporté</li> </ul>	<ul style="list-style-type: none"> <li>• Type de cliniques non détaillé (indicateur de sévérité)</li> <li>• Questionnaire de productivité non validé</li> <li>• Absentéisme : devis transversal avec période de rappel de 12 mois (rétrospectif)</li> <li>• Taux de participation non rapporté</li> </ul>

Aucune des études identifiées n'a rapporté de données prospectives sur la perte de productivité associée à l'hypoglycémie, mettant les résultats à risque de biais de rappel.(114-116) Bien que les trois études aient rapporté des données sur l'absentéisme, une seule étude a rapporté des données sur le présentéisme.(115) Par contre, les données sur le présentéisme étaient limitées à l'impact des événements hypoglycémiques non sévères nocturnes.(115) Aucune des études n'a rapporté l'impact de l'hypoglycémie sur les activités courantes, limitant l'impact aux travailleurs. Les évidences sur l'impact de l'hypoglycémie sur la productivité sont actuellement limitées.(114-116) La génération de données canadiennes prospectives sur l'impact de l'hypoglycémie sur la productivité (absentéisme, présentéisme et activités courantes) fait donc partie des objectifs de l'article 3 de cette thèse. Il convient de souligner que l'étude HAT a également mesuré les données d'utilisation des ressources de la santé, où les résultats de la cohorte canadienne sont présentés dans l'article par O'Reilly et al(116). Contrairement aux données de productivité, les données d'utilisation des ressources ont été mesurées prospectivement, augmentant la crédibilité des résultats. Parmi les patients atteints de DMT2, 0,8% ont été hospitalisés en moyenne 1,8 jours durant une période de suivi de 4 semaines, représentant un coût moyen d'hospitalisation relié à l'hypoglycémie de 1881,25\$ par patient par année. Également, 2,6% des patients avec DMT2 ont nécessité en moyenne 1,5 visites médicales additionnelles dues à l'hypoglycémie durant une

période de suivi de 4 semaines, représentant un coût moyen de visites médicales reliées à l'hypoglycémie de 204,10\$ par patient par année.(116)

#### **4.3.4 Impact de l'hypoglycémie sur la productivité : facteurs confondants**

Afin de répondre à l'objectif de l'article 3 d'évaluer l'impact de l'hypoglycémie sur la productivité, la liste des confondants potentiels doit tout d'abord être établie. Tel que discuté à la section 4.3.3, 37 articles (canadiens et non-canadiens) ont été identifiés dans la littérature. Lors de la révision des 37 articles, 29 ont été exclus, car l'effet de l'hypoglycémie sur la productivité n'avait pas été ajusté et ne rapportaient donc pas de facteurs confondants, pour un total de 8 articles. En plus de la liste des variables, le type de variables (p.ex. continu, dichotomique, catégoriel) a été aussi été extraits. La liste des confondants extraits est présenté au Tableau 11 **Erreur! Source du renvoi introuvable.**, dans lequel le type de variables n'a pas été précisé afin de simplifier la présentation.

Les confondants identifiés vont comme suit : âge, sexe, ethnicité, statut socioéconomique, statut marital, consommation d'alcool, tabagisme, activité physique, activités d'autogestion du diabète, durée du diabète, HbA1c, indice de masse corporelle, caractéristiques des thérapies et complications du diabète. Le choix des facteurs confondants potentiels qui seront collectés dans l'article 3 est discuté à la section 6.3.3.

**Tableau 11. Facteurs confondants potentiels pour les valeurs de perte de productivité**

	Araki 2020 (117)	Davis 2005 (96)	Dibonaventura 2011 (118)	Meneghini 2017 (101)	Pawaskar 2017 (119)	Pawaskar 2018 (120)	Rhoads 2005 (121)	Teruchi 2019 (109)
Âge	x	x	x	x	x	x	x	x
Sexe	x		x	x	x	x	x	x
Ethnicité			x	x				
Éducation			x	x				x
Revenu			x	x				x
Statut d'emploi			x	x			x	x
Statut marital			x					x
Alcool				x				x
Tabagisme				x	x	x		x
Activité physique				x				x
Nombre de visites médicales <sup>1</sup>								
Autogestion du diabète <sup>2</sup>	x							
Durée du diabète	x		x	x	x			
HbA1c	x		x	x	x			
Indice de masse corporelle	x	x		x				x
Type de thérapie			x	x	x	x		
Nombre de traitements			x	x				

	Araki 2020 (117)	Davis 2005 (96)	Dibonaventura 2011 (118)	Meneghini 2017 (101)	Pawaskar 2017 (119)	Pawaskar 2018 (120)	Rhoads 2005 (121)	Terauchi 2019 (109)
Durée de thérapie			x	x				
Complications macrovasculaires <sup>3</sup>	x	x	x	x	x	x	x	x
Complications microvasculaires <sup>3</sup>	x	x	x	x	x	x	x	x

1. Inclut visites chez un médecin de famille, médecin spécialiste, urgence et hospitalisations.

2. Inclut scores de questionnaire sur l'autogestion du diabète ou composantes individuelles telles que la diète, les mesures glycémiques, l'adhérence.

3. Inclut regroupements de complications, complications individuelles et scores de comorbidité.

À noter, certaines études ont également inclus le type d'assurance. Par contre, ce facteur n'est pas significatif au Canada étant donné le régime médical universel et l'assurance médicaments obligatoire.



## CHAPITRE 5 : OBJECTIFS

---

Les ECR sont favorisés par les agences règlementaires pour supporter une indication thérapeutique. Au cours des dernières années, il y a une émergence des études en pratique réelle afin de recueillir de l'évidence sur l'efficacité comparative des traitements. La conduite des ECR est réalisée selon des lignes directrices strictes et requièrent une grande rigueur. Afin d'uniformiser les études en pratique réelle, plusieurs recommandations ont été publiées avec l'objectif de créer un cadre de recherche spécifique à l'efficacité comparative. Toutefois, lorsque les chercheurs désirent générer des données pharmacoéconomiques en pratique réelle, que ce soient des données d'utilité ou de coûts, aucune ligne directrice ne fournit des recommandations spécifiques à cette fin. L'absence de guide peut mener à un manque de standardisation et de rigueur dans la conduite d'études pharmacoéconomiques en pratique réelle.

L'objectif global de cette thèse est de comprendre et d'améliorer les pratiques méthodologiques pour la réalisation d'études générant des données pharmacoéconomiques en pratique réelle. Afin de répondre à cet objectif, trois projets de recherche ont été mis en place, menant à trois articles distincts. Le premier article publié est une revue de la littérature caractérisant les méthodes de collecte de données pharmacoéconomiques, en pratique réelle, chez les patients atteints de DT2M. Le constat du manque de crédibilité des résultats pharmacoéconomiques des études en pratique réelle a mené au deuxième article présentant une boîte à outils visant à éviter les biais en pratique réelle. La boîte à outils a ensuite été mise en pratique dans un troisième article qui consiste en la conception d'une étude collectant des données pharmacoéconomiques en pratique réelle en DT2M. Les objectifs spécifiques à chacun des articles sont présentés ci-dessous.

## **5.1 Article 1 – Revue des études en pratique réelle en diabète de type 2 : manque de bonnes pratiques**

### **5.1.1 Objectif général**

Les bonnes pratiques cliniques (*Good Clinical Practice*) fournissent un cadre de conduite afin de réaliser des essais cliniques de grande qualité méthodologique. En l'absence de guide de référence, le manque de rigueur est une préoccupation pour les études conduites en pratique réelle. L'objectif est donc de caractériser les pratiques méthodologiques des études collectant des données de pharmacoéconomie. Puisque les banques de données traditionnelles ne contiennent généralement pas d'importantes données pharmacoéconomiques, telles que l'utilité, la perte de productivité et les coûts non médicaux, les chercheurs en pharmacoéconomie doivent parfois générer leurs propres données. Par conséquent, la revue se concentre spécifiquement sur ce type d'études, c'est-à-dire les articles générant des données de novo (données primaires).

### **5.1.2 Objectifs spécifiques**

- Identifier les études générant des données pharmacoéconomiques en DT2M en pratique réelle;
- Caractériser la méthodologie des études identifiées;
- Évaluer la qualité des études.

## **5.2 Article 2 – Études en pratique réelle : une boîte à outils pour la collecte de données d'utilité**

### **5.2.1 Objectif général**

Un cadre de recherche spécifique à la collecte de données pharmacoéconomiques en pratique réelle peut augmenter la qualité des résultats générés. Initialement, l'objectif était d'identifier les lignes directrices spécifiques à la génération de données pharmacoéconomiques, que ce soient des données d'utilité ou de coûts. Ensuite, d'incorporer les bonnes pratiques méthodologiques dans une boîte à outils destinée spécifiquement aux investigateurs cherchant à générer ces données en pratique réelle. Lors de la soumission du deuxième article à un journal d'intérêt, l'éditeur a jugé que le contenu de l'article couvrait trop d'aspects et a demandé que l'objectif soit plus ciblé. Ainsi, la publication de la boîte à outils a été restreinte aux recommandations sur les données d'utilité publiées par les agences professionnelles de pharmacoéconomie. Les organisations professionnelles de la pharmacoéconomie incluaient European Network for Health Technology Assessment, Health Technology Assessment International, International Health Economics Association, International Network of Agencies for Health Technology Assessment et International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Cette recherche plus ciblée incluait également les lignes directrices des pays précédemment établis comme experts en pharmacoéconomie (Allemagne, Angleterre, Australie, Canada, Écosse, Espagne, France, Pays-Bas et Suède)(10).

### **5.2.2 Objectifs spécifiques**

- Identifier les lignes directrices sur les données d'utilité publiées par les agences professionnelles de pharmacoéconomie (initialement l'objectif était de répertorier plus globalement l'ensemble des lignes directrices sur la génération de données pharmacoéconomiques);
- Extraire les recommandations pratiques sur la génération de données d'utilité;

- Créer une boîte à outils à partir des recommandations existantes afin d'aider les chercheurs à éviter les biais lors de la génération de données d'utilité.

### **5.3 Article 3 – Impact de l'hypoglycémie sur la perte de productivité et l'utilité des patients atteints de diabète de type 2 et traités à l'insuline en pratique réelle canadienne : un protocole d'étude**

#### **5.3.1 Objectif général**

L'objectif est de mettre en pratique la boîte à outils en rédigeant un protocole de recherche. Pour se faire, le DMT2 est encore une fois utilisé comme champ thérapeutique, plus précisément l'hypoglycémie chez les patients recevant de l'insuline. Le protocole d'étude vise à évaluer l'impact de l'hypoglycémie sur l'utilité et la productivité. La conduite de l'étude au-delà de la conception du protocole ne faisait pas partie des objectifs de cette thèse. Dans le contexte de la pandémie COVID-19, la réalisation de ce type d'étude prospective était simplement irréaliste dans le cadre d'un projet de doctorat non financé. Cette décision a été prise en accord avec le comité aviseur.

#### **5.3.2 Objectifs spécifiques**

- Décrire l'état des connaissances actuelles et déterminer un besoin pour la génération de données pharmacoéconomiques supplémentaires;
- Utiliser les recommandations de la boîte à outils afin de mettre en place un devis d'étude et des procédures d'étude qui limitent le risque de biais;
- Rédiger un protocole de qualité, pertinent et applicable à la réalité des chercheurs en pratique réelle.

## CHAPITRE 6 : MÉTHODOLOGIE

---

En tant que première auteure de mes trois articles publiés dans le cadre de cette thèse, j'ai mené les étapes de conception, développement, réalisation et écriture des manuscrits. Les membres de mon comité aviseur ont joué un rôle de supervision et ont participé aux étapes de conception et révision des manuscrits. Plus précisément pour les articles de revues de la littérature, plusieurs étapes nécessitaient deux réviseurs indépendants et même parfois un troisième réviseur. J'ai été réviseur pour chacune des étapes des différents articles, et les deuxième et troisième réviseurs étaient des étudiantes à la maîtrise de l'Université de Montréal.

### **6.1 Article 1 – Revue des études en pratique réelle en diabète de type 2 : manque de bonnes pratiques**

#### **6.1.1 Revue de la littérature et sélection des études**

Une revue de la littérature a été conduite selon un protocole prédéfini. La question de recherche a été développée selon le format PICO :

- Population : patients atteints du DMT2;
- Intervention : toute intervention pour les soins du DMT2 ou absence d'intervention;
- Comparateur : toute intervention pour les soins du DMT2 ou absence d'intervention;
- Issue (*Outcome*) : valeurs d'utilité, utilisation des ressources et perte de productivité.

La recherche bibliographique a été effectuée à l'aide des banques électroniques MEDLINE et EMBASE. Afin de capter l'ensemble des études potentielles, les termes de recherche incluaient des mots-clés liés à la pharmacéconomie et à la pharmacoépidémiologie. Après avoir suivi les tutoriels offerts par la bibliothèque de l'Université de Montréal, la stratégie de recherche a été développée en utilisant des filtres validés, notamment ceux publiés par l'ACMTS(122) et par Scottish Intercollegiate Guidelines Network (SIGN).(123) Les mots-clés incluaient entre autres :

- Termes pharmacoépidémiologiques : *epidemiologic studies, cohort studies, cross-sectional studies, longitudinal, prospective, observational, pragmatic trial, health survey, real world, real life*;
- Termes pharmacoéconomiques :
  - Utilité : *utilities, standard gamble, analog scale, time trade off, multiattribute health index, 15 dimension, eq-5d, sf-6d, health utility index, quality adjusted life year, preference*;
  - Productivité : *productivity loss, absenteeism, presenteeism, work productivity, activity impairment, work loss, work disability, sick leave, work capacity, workplace activity limitations impairment, work ability index, work productivity and activity impairment, work productivity scale, health and work performance, employment absence, disability scale, presenteeism scale, work productivity short inventory*;
  - Ressources : *health care utilization, health resources, health care costs, hospitalizations, length of stay, health system use, medical system use, costs, cost analysis, economics, pricing*;

La recherche a par la suite été restreinte au DMT2 en ajoutant une autre série de termes spécifiques. Les études ont tout d'abord été sélectionnées à partir des titres et des résumés. Les articles complets ont ensuite été revus à l'aide d'une liste d'éligibilité prédéfinie. Les critères d'éligibilité allaient comme suit :

- Études conduites en pratique réelle définie comme études n'étant pas des ECR et représentant la pratique clinique de routine. Aucune restriction n'a été émise sur l'interférence avec l'assignation de traitement ou randomisation;
- Études générant des données de novo (pas d'études de banques de données ou de dossiers électroniques médicaux);
- Études sur le DMT2 (pas de restriction sur l'âge);
- Études collectant des données pharmacoéconomiques (utilité, utilisation des ressources, coûts ou perte de productivité);

- Études publiées en anglais ou en français;
- Études publiées entre le 1<sup>er</sup> janvier 2017 et 27 juin 2018;
- Articles complets (pas de résumés ou affiches de conférence).

Puisque l'objectif de cette revue se concentrait sur les articles générant des données de novo, les études de données secondaires (telles que les études de banques de données) ont été exclues à l'étape de l'évaluation de l'éligibilité. Également, compte tenu du nombre imposant de publications en DMT2, la recherche a été restreinte à une courte période afin d'établir un portrait plus précis de la situation actuelle. Tous les critères d'éligibilité ont été déterminée a priori et les études devaient rencontrer l'ensemble des critères afin d'être éligibles. La sélection des études a été réalisée par deux réviseurs indépendants où les désaccords ont été résolus au moyen de discussions et consultations avec un troisième réviseur lorsque nécessaire.

### **6.1.2 Extraction des données**

Pour chaque étude incluse dans la revue, les caractéristiques suivantes ont été extraites à l'aide d'une grille d'extraction prédéfinie :

- Premier auteur et pays d'étude;
- Devis;
- Population de patients;
- Intervention;
- Groupe d'exposition et mesure de l'exposition;
- Site d'étude et nombre de sites;
- Taille d'échantillon et justification;
- Méthode d'échantillonnage et source de sélection des patients;
- Méthode et durée de recrutement;
- Taux de participation;
- Approbation éthique et consentement du patient;

- Durée du suivi;
- Contrôle de la qualité;
- Publication du protocole d'étude;
- Conflits d'intérêt et source de financement;
- Issue pharmacoéconomique et nombre de mesures;
- Instrument de mesure de l'issue;
- Source de mesure de l'issue et période de rappel;
- Taux de données manquantes;
- Généralisabilité des résultats.

Il est possible que les études aient collecté des issues cliniques en plus des issues pharmacoéconomiques. Cependant, l'extraction se concentrait uniquement sur les issues pharmacoéconomiques et pourrait ne pas refléter l'ensemble des objectifs d'une étude. De plus, les résultats des études n'ont pas été extraits puisque la revue visait uniquement à caractériser les pratiques méthodologiques. L'extraction des données a été effectuée par deux réviseurs indépendants où les désaccords ont été résolus au moyen de discussions et consultations avec un troisième réviseur lorsque nécessaire. Les données extraites ont été présentées à l'aide de statistiques descriptives telles que des mesures de fréquence.

### **6.1.3 Évaluation de la qualité**

Le risque de biais des résultats des études comparatives a été évalué à l'aide de l'outil ROBINS-I (*Risk Of Bias In Non-randomized Studies - of Interventions*).<sup>(55)</sup> Pour ce qui est des études descriptives, la qualité a été évaluée à l'aide de l'outil pour l'évaluation d'études descriptives du Joanna Briggs Institute (JBI).<sup>(124)</sup> Le risque de biais a été évalué par deux réviseurs indépendants où les désaccords ont été résolus au moyen de discussions et consultations avec un troisième réviseur lorsque nécessaire. L'outil ROBINS-I couvre sept domaines : biais de confusion, sélection des participants, classification des interventions, déviations des interventions



attendues, données manquantes, mesure des issues et sélection des résultats rapportés. Le risque de biais a été évalué pour chacun des sept domaines et pour l'ensemble d'une étude. Les résultats ont été rapportés visuellement à l'aide d'histogrammes pondérés. Les critères d'évaluation de la qualité de l'outil du JBI pour études descriptives sont : sélection aléatoire des patients, critères d'éligibilité clairs, identification des facteurs confondants, validité de l'évaluation de l'issue, description suffisante des groupes, horizon temporel suffisant, inclusion des pertes au suivi, méthode fiable de mesure de l'issue et justesse des analyses statistiques. Puisque l'ensemble des critères n'est pas applicable à toutes les études, le score de qualité est présenté en pourcentage.

## **6.2 Article 2 – Études en pratique réelle : une boîte à outils pour la collecte de données d'utilité**

Tel que décrit dans les objectifs, la revue conduite afin de construire la boîte à outils avait initialement une portée beaucoup plus large. Elle visait globalement les données pharmacoéconomiques et n'était pas limitée aux données d'utilité. À cet effet, les lignes directrices publiées par de nombreuses agences professionnelles avaient été répertoriées afin d'identifier les publications fournissant des recommandations afin d'augmenter la crédibilité des études conduites en pratique réelle (peu importe l'issue étudiée). La liste d'agences professionnelles avait été établie basée une revue précédemment publiée sur la recherche d'efficacité comparative en pratique réelle.(125) Les publications avec des recommandations pour éviter les biais sont présentées au Tableau 12 accompagnées d'une brève description. Trente-quatre lignes directrices avaient été identifiées.(50, 51, 55, 64, 65, 126-154)

Les changements apportés à la boîte à outils et les différences entre l'objectif initial et la publication finale sont expliqués dans les prochains paragraphes. Une schématisation du changement dans la stratégie de recherche est également présentée à la Figure 6.

**Tableau 12. Recommandations pour éviter les biais dans la recherche en pratique réelle**

Source	Année	Titre	Description
AHRQ	2013 (152)	Developing a Protocol for Observational Comparative Effectiveness Research: A User's guide	Bonnes pratiques pour le développement d'un protocole de recherche comparative de l'efficacité
	2020 (144)	Registries for Evaluating Patient Outcomes: A User's Guide (Fourth Edition)	Bonnes pratiques sur la planification, le développement, le maintien et l'évaluation de registres
Cochrane	2016 (55)	ROBINS-I: A Tool for Assessing Risk of Bias in Non-Randomized Studies of Interventions	Outil d'évaluation du risque de biais dans les études non-randomisées couvrant 7 domaines
Duke Margolis Center for Health Policy	2017 (134)	A Framework for Regulatory Use of Real-World Evidence	Bonnes pratiques sur la génération d'évidence en pratique réelle aux fins réglementaires
	2018 (133)	Characterizing Real-World Data Quality and Relevancy for Regulatory Purposes	Bonnes pratiques sur les techniques de documentation pour le rapport de données de qualité
	2019 (135)	Determining Real-World Data's Fitness for Use and the Role of Reliability	Bonnes pratiques sur l'applicabilité des données de pratique réelle à l'aide d'une liste de vérification
	2019 (65)	Understanding the Need for Non-Interventional Studies Using Secondary Data to Generate Real-World Evidence for Regulatory Decision Making, and Demonstrating Their Credibility	Bonnes pratiques pour la recherche comparative de l'efficacité utilisant des données secondaires aux fins réglementaires
	2019 (136)	Adding Real-World Evidence to a Totality of Evidence Approach for Evaluating Marketed Product Effectiveness	Bonnes pratiques sur comment les données de pratique réelle peuvent contribuer à l'évidence pour la modification d'une indication
	2020 (137)	A Roadmap for Developing Study Endpoints in Real-World Settings	Bonnes pratiques sur le développement d'issues cliniques en pratique réelle aux fins réglementaires
ENCePP	2018 (138)	ENCePP Checklist for Study Protocols (Revision 4)	Liste de vérification pour la rédaction de protocoles pour la pharmacoépidémiologie et la pharmacovigilance

	2020 (50)	Guide on Methodological Standards in Pharmacoepidemiology (Revision 8)	Bonnes pratiques méthodologiques pour guider les chercheurs en pharmacoépidémiologie
FDA	2005 (140)	Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment	Bonnes pratiques sur l'identification et l'interprétation de signaux d'innocuité, l'évaluation pharmacoépidémiologique et la pharmacovigilance
	2017 (143)	Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices – Guidance for Industry and FDA Staff	Bonnes pratiques pour déterminer l'applicabilité des données de pratique aux fins réglementaires
	2018 (142)	Use of Electronic Health Record Data in Clinical Investigations – Guidance for Industry	Bonnes pratiques sur l'utilisation de dossiers médicaux électroniques aux fins réglementaires
	2018 (141)	Framework for FDA's Real-World Evidence Program	Bonnes pratiques sur l'évaluation de l'utilisation potentielle des données de pratique réelle aux fins réglementaires
GRACE	2010 (132)	GRACE Principles: Recognizing High-Quality Observational Studies of Comparative Effectiveness	Guide destiné aux réviseurs pour évaluer la rigueur des études en recherche comparative de l'efficacité
	2016 (131)	GRACE Principles: A Validated Checklist for Evaluating the Quality of Observational Cohort Studies for Decision-Making Support	Liste de vérification de 11 items destinée aux réviseurs pour évaluer la qualité des études de recherche comparative de l'efficacité
Green Park Collaborative	2017 (145)	RWE Decoder Framework: A Practical Tool for Assessing Relevance and Rigor of Real World Evidence	Outil d'aide à la révision pour les décideurs évaluant la qualité des études de pratique réelle
ISPE	2015 (146)	Guidelines for Good Pharmacoepidemiology Practices (Revision 3)	Bonnes pratiques sur la planification, la conduite et l'interprétation de la recherche pharmacoépidémiologique
ISPOR	2003 (148)	A checklist for retrospective database studies – Report of the ISPOR Task Force on Retrospective Databases	Contient un questionnaire de 27 items permettant d'évaluer la pertinence et validité des études rétrospectives de banques de données
	2009 (128)	Good Research Practices for Comparative Effectiveness Research: Defining, Reporting and Interpreting Nonrandomized Studies of Treatment Effects Using	Bonnes pratiques sur la formulation de question de recherche, le rapport et l'interprétation de données de recherche comparative de l'efficacité utilisant des banques de données secondaires

		Secondary Data Sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part I	
	2009 (130)	Good Research Practices for Comparative Effectiveness Research: Approaches to Mitigate Bias and Confounding in the Design of Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report—Part II	Bonnes pratiques sur les approches méthodologiques limitant le risque de biais pour la recherche comparative de l'efficacité utilisant des banques de données secondaires
	2009 (147)	Good Research Practices for Comparative Effectiveness Research: Analytic Methods to Improve Causal Inference from Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part III	Bonnes pratiques sur les techniques analytiques améliorant l'inférence causale des études de recherche comparative de l'efficacité utilisant des banques de données secondaires
	2012 (64)	Prospective Observational Studies to Assess Comparative Effectiveness: The ISPOR Good Research Practices Task Force Report	Bonnes pratiques sur la conduite d'études prospectives de recherche comparative de l'efficacité
ISPOR-AMCP-NPC	2014 (51)	A Questionnaire to Assess the Relevance and Credibility of Observational Studies to Inform Health Care Decision Making: An ISPOR-AMCP-NPC Good Practice Task Force Report	Questionnaire de 33 items destiné aux réviseurs pour évaluer la pertinence et la crédibilité des études de recherche comparative de l'efficacité
ISPOR-ISPE(129)	2017 (129)	Good Practices for Real-World Data Studies of Treatment and/or Comparative Effectiveness: Recommendations	Bonnes pratiques procédurales couvrant 7 aspects pour améliorer la transparence des études de recherche comparative de l'efficacité

		from the Joint ISPOR-ISPE Special Task Force on Real-World Evidence in Health Care Decision Making	
	2017 (154)	Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0	Bonnes pratiques pour accroître la transparence et la reproductibilité des études de données secondaires
NICE	2015 (139)	The use of observational data to inform estimates of treatment effectiveness for Technology Appraisal: Methods for comparative individual patient data	Bonnes pratiques sur l'analyse de données individuelles comparatives pour obtenir des estimations d'efficacité de traitement de qualité
	2016 (155)	The Use of Real World Data for the Estimation of Treatment Effects in NICE Decision Making	Discussion à l'aide d'exemples de cas sur les enjeux prioritaires associés au processus d'évaluation des études de recherche comparative de l'efficacité en pratique réelle
PCORI	2012 (149)	Methodological Standards and Patient-Centeredness in Comparative Effectiveness Research: The PCORI perspective	Bonnes pratiques méthodologiques sur la formulation de question de recherche, le devis, l'implication du patient et la diffusion des résultats de recherche centrée sur le patient
	2014 (150)	A RoadMap for State Policymakers to Use Comparative Effectiveness and Patient-Centered Outcomes Research to Inform Decision Making	Recommendations for the use of CER and patient-centered outcomes research by policy makers
	2019 (151)	PCORI Methodology Standards	Bonnes pratiques sur la recherche centrée sur le patient, présentant 65 standards dans 16 thèmes de recherche
RECORD	2015 (127)	The Reporting of Studies Conducted using Observational Routinely-collected Health Data (RECORD) Statement	Extension du STROBE - Liste de vérification de 22 items spécifique aux données collectées de routine
STROBE(153)	2007 (153)	The STROBE Statement: Guidelines for Reporting Observational Studies	Liste de vérification de 22 items sur les bonnes pratiques de rapport de résultats

AHRQ: Agency for Healthcare Research and Quality; AMCP: Academy of Managed Care Pharmacy; CER: Comparative Effectiveness Research; ENCePP: European Network of Centers for Pharmacoepidemiology and Pharmacovigilance; FDA: Food and Drug Administration; GRACE: Good Research for Comparative Effectiveness Initiative; ISPE: International Society for Pharmacoepidemiology; ISPOR: International Society for Pharmacoeconomics and Outcomes Research; NICE: National Institute for Health and Care Excellence; NPC: National Pharmaceutical Council; PCORI: Patient-Centered Outcomes Research Institute; RECORD: REporting of studies Conducted using Observational Routinely-collected health data; RWE: Real-World Evidence; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology.

Toutefois, suite aux commentaires de l'éditeur d'un journal d'intérêt demandant de restreindre la boîte à outils, l'angle de la stratégie de recherche a dû être modifié. Les lignes directrices des pays experts en pharmacoeconomie (Allemagne, Angleterre, Australie, Canada, Écosse, Espagne, France, Pays-Bas et Suède) ont été répertoriées afin d'identifier les recommandations existantes sur la collecte de valeurs d'utilité. La liste des pays experts répliquait celle d'une revue publiée dans une édition spéciale du journal *Pharmacoeconomics*.<sup>(10)</sup> Les lignes directrices publiées par les organisations professionnelles de la pharmacoeconomie ont également été répertoriées, incluant European Network for Health Technology Assessment, Health Technology Assessment International, International Health Economics Association, International Network of Agencies for Health Technology Assessment et International Society for Pharmacoeconomics and Outcomes Research (ISPOR). La recherche a été conduite jusqu'en avril 2021 et 58 publications potentielles ont été identifiées par titres et résumés. Après avoir lu les publications complètes et résolu les discordances par discussion, les deux réviseurs ont exclu 49 publications ne discutaient pas spécifiquement de la collecte de valeurs d'utilité :

- Publications sur la modélisation économique et techniques pour la réalisation d'ACU (n=29);
- Publications sur les techniques de cartographie (*mapping*) (n=13);
- Publications spécifiques à une certaine population (n=4);
- Publications spécifiques à un certain champ thérapeutique (n=3).

Parmi les neuf publications restantes, deux étaient des doublons et sept étaient des publications originales : cinq publiés par National Institute for Health and Care Excellence (NICE)<sup>(156-160)</sup> et deux publiés par ISPOR.<sup>(161, 162)</sup> L'extraction de données incluait les éléments méthodologiques clés affectant la validité interne :

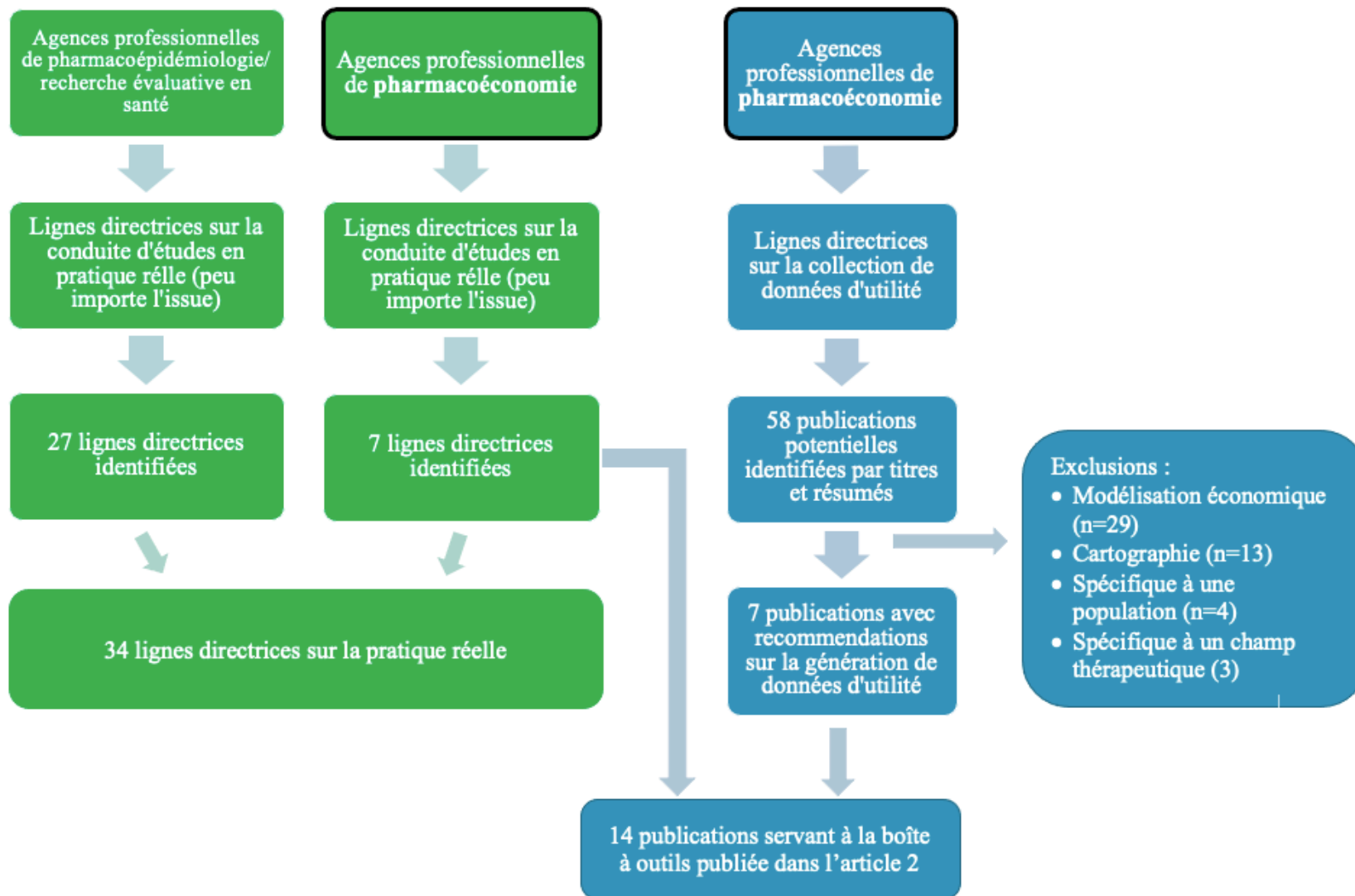
- Définition de l'état de santé;
- Choix de l'instrument de mesure et mode d'administration;
- Fréquence et horaire des mesures;
- Choix des répondants et échantillonnage;

- Méthode de recrutement et taux de participation;
- Données manquantes;
- Devis de l'étude;
- Techniques analytiques.

Aucune des sept publications incluses ne fournissait un cadre de recherche pour les investigateurs désirant estimer des données d'utilité en pratique réelle. Trois documents discutent des préférences de NICE et des alternatives de mesures de l'utilité, précisant leurs choix d'instruments et de répondants. Deux rapports proposent une méthodologie systématique pour incorporer les valeurs d'utilité obtenues dans la littérature dans un modèle économique, soulignant que la qualité et la pertinence des sources originales doivent être revues. Les taux de réponse et de données manquantes font partie de la liste de vérification minimale de qualité. Une revue par NICE démontre que le rapport d'information basique sur les modèles de régressions utilisés dans les modèles analytiques de décision est limité et fournit des recommandations pour améliorer les standards. La seule publication qui procure un cadre de recherche exhaustif est le rapport d'ISPOR sur la collecte de valeurs d'utilité dans le cadre des ECR. Toutefois, puisque le rapport se concentre sur les essais cliniques, certains aspects tels que le devis et les considérations analytiques pour les études en pratique réelle sont seulement brièvement discutés. Dans l'ensemble, seuls certains aspects sont discutés en détail : la définition de l'état de santé pour lequel sont estimées les valeurs d'utilité doit être représentative de l'état de santé du modèle économique, les méthodes de mesures indirectes avec système de classification générique sont généralement privilégiées et les répondants doivent généralement être les patients eux-mêmes et être représentatifs de la population du modèle économique. Les bonnes pratiques pour éviter les biais lors de la collecte de données d'utilité, particulièrement le biais de sélection et de confusion, sont limitées. Par conséquent, les lignes directrices sur la conduite d'études en pratique réelle publiées par les agences professionnelles de pharmacoéconomie ont été ajoutées afin de compléter les recommandations sur la collecte de données d'utilité. Cette étape était nécessaire afin de répertorier des recommandations sur la conduite d'études en pratique réelle. Sept publications ont été identifiées : une publiée par NICE(139) et six publiés par ISPOR.(51, 64, 128, 130, 147, 148) Une schématisation est présentée à la Figure 6.



Figure 6. Stratégie de recherche initiale (en vert) et stratégie finale (en bleu)



Puisqu'il n'existe aucune publication unique fournissant un cadre de recherche détaillé pour l'estimation de données d'utilité en pratique réelle, les recommandations pour la collecte de valeurs d'utilité et les recommandations pour la recherche en pratique réelle ont été combinées ensemble afin de développer une boîte à outils concise. Les recommandations de la boîte à outils visent à aider les chercheurs dans la conception et la conduite d'études tout en limitant les biais. La boîte à outils comprend quatre sections : la section 1 discute du biais de confusion; la section 2 et 3 discutent du biais de sélection (échantillon et données manquantes respectivement) et la section 4 discute du biais d'information. Les recommandations existantes sont résumées dans chaque section de la boîte à outils et synthétisées en étapes pratiques.

Les recommandations identifiées dans la revue initiale, mais qui n'ont pas été publiées dans la boîte à outils sont présentées au Tableau 13. Les recommandations non publiées incluent celles sur les autres données pharmacoéconomiques. Elles incluent également celles qui avaient été publiées par d'autres agences (pharmacoépidémiologie et recherche évaluative) que celles de pharmacoéconomie. Il convient aussi de noter que certaines lignes directrices ne faisaient qu'effleurer certains aspects et réfèrent plutôt une publication clé à lire pour entrer dans le vif du sujet. Le cas échéant, la recherche initiale avait également répertorié les publications clés et celles-ci sont citées directement afin de soutenir les différentes recommandations du Tableau 13. Il est aussi important de mentionner que même si l'ensemble des recommandations n'ont pas été incluses dans la boîte à outils de l'article 2, elles ont été prises en compte dans l'élaboration de l'article 3.

**Tableau 13. Recommandations non publiées par la boîte à outils**

Control for confounding by design or through analytical methods

- Ordinary least squares (OLS) should be used carefully as it may generate biased estimates. For HSU, proposed techniques include Tobit models, censored least absolute deviations models, two-part models and latent class models.(163) OLS coupled with robust standard errors or nonparametric bootstrap is also a valid approach.(163, 164)
- For costs, several options are available, commonly comprising two-part models and generalized linear models.(41)
- For longitudinal data (either costs or HSU), generalized estimating equations and multilevel models (e.g. mixed models) are adequate options.(41, 165, 166)
- One major advantage of the GEE is the avoidance of distributional assumptions, where several distributions can be tested. For utility, common alternatives to the Gaussian distribution with identity link function include Gaussian distribution with log link, negative binomial distribution with log link, gamma distribution with identity link, and gamma distribution with log link. The gamma distribution requires non-zero positive continuous data, thus utility has to be transposed into disutility (1- utility value). A beta binomial distribution can also be used with a transformation (linear transformation or rescaling) to fit the restrictive open interval (0,1), which excludes the endpoints 0 and 1.(167)
- For healthcare resource utilization, non-linear options mostly include negative binomial distribution (for resource use), as well as gamma and inverse-Gaussian distribution with log link function (for costs).(41)

Explore the risk of unmeasured confounding

- There is a wide range of techniques to explore the risk of unmeasured confounding, from simple spreadsheet-based analyses (e.g. array approach, rule-out approach) to complex simulation-based analyses (e.g. Bayesian, Monte Carlo analysis).(36, 65, 168-173) The

E-value is a new standardized approach for sensitivity analyses on unmeasured confounding.(66, 67)

Select participants systematically, implement standardized procedures, and document participation rate to discuss risk of selection bias

- To select the study population, there are several sampling techniques available. Probability sampling is the gold standard to ensure sample representativeness. Most often, tradeoffs have to be made, balancing validity, generalizability, feasibility and practicality. Indeed, probability sampling is resource-consuming, and might not be feasible depending on the availability of a sampling frame. In addition to simple random sampling, probability sampling techniques include other approaches adapted to different sampling frames and research questions. For large sample frames, a convenient method is systematic sampling, if frame organization permits (no underlying pattern). Another interesting technique is stratified sampling, which creates subgroups based on a specific characteristic (e.g. disease severity). The main limitation of stratified sampling is that stratification factors are often lacking in sample frames. One may also be interested in cluster sampling depending on the research question. Non-probability sampling methods include snowball sampling, purposive sampling, quota sampling, and the most widely used convenience sampling.(174, 175)

Implement a method to handle missing data

- To choose an adequate method for handling missing data, one should make assumptions on the missing data mechanism, which can be classified as MCAR, missing at random (MAR) or missing not at random (MNAR). To distinguish between MCAR and MAR, association between missingness and baseline variables (and previously observed outcomes) can be explored through regression analysis.(36, 50, 52) Another option is to test for differences in the rate of missingness for fully observed baseline variables.(176)

- One should also refer to outcome-specific guidelines, if any, to avoid bad practice (e.g. substituting for EQ-5D missing profile items an average derived from their non-missing items).(163)
- Choice of specific variable for missing data analysis should be clearly described (e.g. individual domain or score level for EQ-5D values, aggregate or disaggregate level for costs).(52)
- Irrespective of the chosen method to handle missing data, findings should be compared with CCA to ensure robustness.(37, 52, 177)
- Additional sensitivity analyses may also compare the selected method to a different analytic technique.(52, 177)

Justify the choice of instrument

- Recall bias is not uncommon for costs that are self-reported by patients.(178)
- The use of a validated instrument is recommended. More specifically, investigators selecting a work productivity instrument for cost-effectiveness analyses (CEA) should consider purpose, perspective, practicality, population and psychometrics.(21)

### **6.3 Article 3 – Impact de l’hypoglycémie sur la perte de productivité et l’utilité des patients atteints de diabète de type 2 et traités à l’insuline en pratique réelle canadienne : un protocole d’étude**

Le protocole a été développé en suivant les recommandations publiées dans la boîte à outils. Également, tel que décrit plus haut, le protocole prend également en considération les recommandations identifiées plus largement même si elles n’ont pas été incluses dans la boîte à outils.

### 6.3.1 Devis et population à l'étude

Une étude prospective sera conduite en pratique réelle afin d'évaluer l'impact de l'hypoglycémie sur la productivité et l'utilité des patients atteints de DMT2 et traités à l'insuline en Ontario et au Québec. Tel qu'expliqué dans la section 2.1, les questionnaires iPCQ(34) et EQ-5D-5L(14) ont été sélectionnés afin de répondre aux objectifs de ce projet de recherche. Cette étude non-interventionnelle et multicentrique est spécifiquement conçue pour la collecte de données pharmacoéconomiques aux fins de modélisation économique future. L'étude produira des intrants descriptifs qui pourront être incorporés dans les modèles économiques en DMT2 en plus des paramètres de précision (intervalles de confiance et erreurs standards) qui pourront servir aux analyses de sensibilité des ACU. Considérant la nature de la condition étudiée, un devis longitudinal avec mesures répétées a été jugé approprié pour limiter le biais de confusion et augmenter la qualité de l'évidence. Répliquant deux études canadiennes à grande échelle, *Hypoglycemia Assessment Tool (HAT) study*(179) et *InHypo-DM study*(180), qui avaient pour objectif de mesurer la fréquence d'évènements hypoglycémiques chez les patients traités à l'insuline, les questionnaires seront complétés à une fréquence de quatre semaines. Considérant qu'un devis longitudinal requiert généralement un minimum de trois mesures(181) et que la fréquence des mesures est de quatre semaines, l'étude aura donc une durée de trois mois.

Dans la pratique clinique canadienne, les patients atteints de DMT2 et traités à l'insuline peuvent être suivis soit par un médecin de famille soit par un spécialiste. Donc, les patients seront recrutés dans quatre cliniques médicales et dans deux cliniques du diabète / endocrinologie. Cette proportion a été choisie afin de représenter la pratique courante au Canada où un plus grand nombre de patients éligibles sont suivis en cliniques médicales qu'en cliniques spécialisées.(180) Afin d'augmenter la généralisabilité des résultats, les patients seront recrutés des deux plus grandes provinces canadiennes, l'Ontario et le Québec. La taille d'échantillon visée est de 500 patients. Les patients éligibles seront identifiés à l'aide des listes de rendez-vous et seront recrutés lors de leurs visites de routine par recrutement consécutif systématique. Pour être éligible, les patients doivent être âgés de 18 ans et plus, diagnostiqués avec le DMT2 et traités à l'insuline. Les patients doivent aussi être capable de comprendre et de lire le français ou l'anglais et de signer le formulaire

de consentement. Les patients ne sont pas éligibles s'ils participent déjà à un essai clinique. Une liste d'éligibilité (*screening log*) sera utilisée pour documenter les taux de participation.

### **6.3.2 Mesure de l'exposition : hypoglycémie**

Selon les associations canadienne et américaine du diabète, l'hypoglycémie est définie comme une concentration plasmatique de glucose de  $\leq 3.9$  mmol/L ( $\leq 70$  mg/dL).(69, 182) Afin de refléter la pratique réelle, la définition de l'hypoglycémie dans le cadre de cette étude ne sera pas restreinte aux épisodes documentés, mais inclura tout événement hypoglycémique jugé par le patient. Les symptômes qui définissent l'hypoglycémie incluent les tremblements, picotements, palpitations, transpiration, anxiété, faim, nausée, troubles de concentration, maux de tête, étourdissements, confusion, faiblesse, somnolence, vision embrouillée et troubles de la parole. Les épisodes hypoglycémiques seront auto-rapportés par les patients à l'aide de la présence de symptômes et/ou confirmés à l'aide de tests de glycémie. Cette approche a également été utilisée dans plusieurs études à grande échelle et représente mieux la pratique réelle où les patients ne testent pas toujours leur glycémie pour différentes raisons (p.ex. oubli, négligence, manque de connaissances, manque de tests). Les épisodes hypoglycémiques seront classés comme sévères ou non-sévères. Un épisode sévère requiert l'assistance d'une autre personne (pour administrer des carbohydrates, glucagon ou prendre d'autres actions correctives), alors qu'un épisode non-sévère peut être autogéré par le patient.(69, 182) Tel que mentionné plus haut, les patients compléteront un questionnaire à 4, 8 et 12 semaines afin d'auto-rapporter le nombre d'évènements hypoglycémiques sévères et non-sévères s'ayant produit au cours des 28 derniers jours.

La sévérité des épisodes hypoglycémiques, ainsi que la fréquence des épisodes, accentuent la peur d'hypoglycémie future.(69, 182) La peur elle-même est reconnue pour affecter la qualité de vie des patients.(69, 182) En plus des symptômes physiques dû aux épisodes hypoglycémiques, l'impact émotionnel négatif de la peur de l'hypoglycémie est également une préoccupation. En ce qui a trait à la modélisation économique, les agences d'évaluation des technologies de la santé, notamment l'ACMTS et National Institute for Health and Care Excellence (NICE), recommandent

que l'impact de l'hypoglycémie sur l'utilité considère à la fois la sévérité et la fréquence des épisodes.(18, 93) Ils recommandent également de ne pas appliquer de décrétement d'utilité séparément pour la peur de l'hypoglycémie.(18, 93) Par conséquent, la présente étude est conçue pour capter l'impact global de l'hypoglycémie sur l'utilité, incluant les effets temporaires des épisodes hypoglycémiques et la peur qui en résulte.

### **6.3.3 Mesure des issues pharmacoéconomiques : utilité et perte de productivité**

Les différents facteurs confondants pouvant influencer l'impact de l'hypoglycémie sur l'utilité et la productivité ont été présentés à la section 4.3.2 et 4.3.4 respectivement. Pour les deux issues, les mêmes facteurs confondants ont été identifiés à l'exception de l'utilisation des services médicaux. Puisque la revue sur la productivité incluait uniquement huit études et que l'utilisation de ressources médicales sert de proxy pour l'état de santé général des patients, ce confondant potentiel sera tout de même considéré pour les deux issues. La redondance des facteurs confondants à travers l'ensemble des études permet d'être confiant sur la liste des covariables à inclure pour limiter le biais de confusion. Les confondants identifiés vont comme suit : âge, sexe, ethnicité, statut socioéconomique, statut marital, consommation d'alcool, tabagisme, activité physique, utilisation des services médicaux, activités d'autogestion du diabète, durée du diabète, HbA1c, indice de masse corporelle, caractéristiques des thérapies et complications du diabète. Afin de valider la liste de confondants identifiés, les travaux publiés par l'*Alliance for Canadian Health Outcomes Research in Diabetes* ont été examinés, notamment la cohorte *Alberta's Caring for Diabetes* (ABCD).(183) La cohorte ABCD a été créée pour étudier les patients atteints de DMT2 à différents niveaux et collecte donc de nombreuses issues, dont l'utilité. Bien que la cohorte ABCD n'ait publié aucune étude spécifique à l'impact de l'hypoglycémie sur l'utilité, l'impact de diverses caractéristiques/complications diabétiques sur l'utilité a été évalué et les facteurs confondants en DMT2 sont toujours sensiblement les mêmes : âge, sexe, ethnicité, statut socioéconomique, statut marital, consommation d'alcool, tabagisme, activité physique, utilisation des services médicaux, activités d'autogestion du diabète et comorbidités, en incluant parfois la durée du diabète, l'indice de masse corporelle et les caractéristiques des thérapies. La liste de confondants identifiés dans la revue est donc très en ligne avec ceux sélectionnés par les experts



de l'*Alliance for Canadian Health Outcomes Research in Diabetes*. Deux facteurs confondants additionnels ont été identifiés dans les publications sur la cohorte ABCD, c'est-à-dire la vaccination annuelle contre la grippe et la vaccination contre le pneumocoque. Donc, ces deux vaccins seront ajoutés à la liste potentielle de confondants à titre de marqueur de bonnes habitudes de vie. Dans le contexte actuel, la vaccination contre la COVID-19 sera également considérée. Un modèle GEE (*generalized estimating equations*) sera utilisé afin d'évaluer les pertes de productivité et les décrets d'utilité associés aux événements hypoglycémiques incidents tout en contrôlant pour les caractéristiques individuelles des patients.

## CHAPITRE 7 : ARTICLE 1 – REVUE DES ÉTUDES DE PRATIQUE RÉELLE EN DIABÈTE DE TYPE 2 : MANQUE DE BONNES PRATIQUES

---

### **Review of real-world evidence studies in type 2 diabetes mellitus: Lack of good practices**

Lambert-Obry V, Lafrance J- P, Savoie M, Henri S, Lachaine J

International Journal of Technology Assessment in Health Care 36, 372–379 (2020)

#### **7.1 Abstract**

**Objectives:** Unlike randomized controlled trials, lack of methodological rigor is a concern about real-world evidence (RWE) studies. The objective of this study was to characterize methodological practices of studies collecting pharmaco-economic data in a real-world setting for the management of type 2 diabetes mellitus (T2DM).

**Methods:** A systematic literature review was performed using the PICO framework: population consisted of T2DM patients, interventions and comparators were any intervention for T2DM care or absence of intervention, and outcomes were resource utilization, productivity loss or utility. Only RWE studies were included, defined as studies that were not clinical trials and that collected de novo data (no retrospective analysis).

**Results:** The literature search identified 1,158 potentially relevant studies, among which sixty were included in the literature review. Many studies showed a lack of transparency by not mentioning the source for outcome and exposure measurement, source for patient selection, number of study sites, recruitment duration, sample size calculation, sampling method, missing data, approbation by an ethics committee, obtaining patient's consent, conflicts of interest, and funding. A significant proportion of studies had poor quality scores and was at high risk of bias. **Conclusions.** RWE from T2DM studies lacks transparency and credibility. There is a need for good procedural practices that can increase confidence in RWE studies. Standardized

methodologies specifically adapted for RWE studies collecting pharmaco-economic data for the management of T2DM could help future reimbursement decision making in this major public health problem.

## 7.2 Introduction

Clinical evidence required for market authorization is mostly obtained from randomized controlled trials (RCTs), which are the gold standard for assessing the efficacy of new drugs (1). However, for many countries, drug approval is not sufficient to allow the integration of innovations in the healthcare system. Indeed, the widespread use of a drug is often closely linked to the reimbursement decisions of public payers. Although therapeutic value is a key component in decision making, health technology assessment agencies also consider the cost-effectiveness of innovations, which is estimated through economic evaluations. Data required to perform a cost-utility analysis comprise utility values and costs, which allow the estimation of the incremental cost-effectiveness ratio. Indirect costs, such as productivity loss, are also considered in economic evaluations performed from a societal perspective. Although they have gained prominence, patient-reported outcomes (PROs) and economic outcomes are not systematically collected in RCTs. Even if they were, RCTs are well-known to have inherent limitations, namely low external validity, highly controlled patient follow-up, and staggering costs. It has led to the use of real-world data (RWD) as complementary evidence in reimbursement decision making (1).

Despite the increasing use of RWD, no formal consensus on its definition has yet been reached. A recent review identified thirty-eight definitions of RWD based on fifty-three published documents and twenty stakeholder interviews (2). The definition developed by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) (i.e., data that are not collected in conventional RCTs) (1) was the most commonly used (53 percent). In 24 percent of the definitions, RWD was restricted to data collected in a setting without interference with treatment assignment, patient follow-up and patient selection (2). Five definitions (13 percent) were limited to data collected in routine clinical practice (no de novo data collection, i.e., use of existing data

only) (2). The most cited data sources were registries, electronic health records, claims databases, and administrative databases, followed by PROs and health surveys (2).

Notwithstanding the lack of universal definition of RWD, a commonly voiced purpose is to complement RCT findings for health decision making (1-5). Good Clinical Practice guidance provides a quality standard for performing RCTs. As for the conduct of real-world evidence (RWE) studies, a unified standard has yet to be made. Indeed, lack of methodological rigor is a concern about RWE studies (3). The objective of this study was to characterize methodological practices of studies collecting pharmaco-economic data, namely utility values, resource utilization, and productivity loss, in a real-world setting for the management of type 2 diabetes mellitus (T2DM). T2DM has been chosen to serve as a case study because of its important burden on the healthcare system. Indeed, according to the World Health Organization (WHO), the global prevalence of diabetes (types 1 and 2) has risen from 4.7 percent in 1980 to 8.5 percent in 2014 (6). In 2014, the prevalence of diabetes was 7.3 percent in Europe, 8.3 percent in America, 8.6 percent in South-East Asia, and 13.7 percent in the Eastern Mediterranean region (6). In addition, the worldwide direct cost of diabetes is estimated to be over US\$827 billion per year (6).

## **7.3 Methods**

### **7.3.1 Literature Search and Study Selection**

A systematic literature review was performed according to a predefined protocol. The study question was established using the PICO framework: population consisted of T2DM patients, intervention and comparator were any intervention for T2DM care or absence of intervention, and outcomes were utility values, resource utilization and productivity loss. A structured literature search was performed using the electronic databases MEDLINE and EMBASE. To capture all potentially relevant studies, the search used pharmaco-economic terms combined with pharmaco-epidemiologic keywords, and was thereafter restricted to diabetes. Full details of the literature search are presented in the Supplementary Material (Supplementary Table 1). Studies

were initially selected based on titles and abstracts. Full-text articles were then reviewed using a predefined eligibility form. Inclusion criteria were:

- Studies performed in a real-world setting, herein defined as studies that were not RCTs, and that were performed during routine clinical practice. There was no restriction on interference with treatment assignment or randomization;
- Studies collecting de novo data (no studies based on electronic health records, claims databases, or administrative databases);
- Studies focusing on T2DM (no restriction on age);
- Studies collecting pharmacoeconomic data (utility values, resource utilization, and productivity loss);
- Studies published in English or French;
- Studies published between 1 January 2017 and 27 June 2018;
- Full-text articles (no conference abstracts).

The criterion of de novo data collection was used to exclude retrospective database studies, in which pharmacoeconomic outcomes are usually lacking. Although highly relevant for the assessment of clinical outcomes and direct medical costs, most clinical databases, such as electronic health records, do not provide information on health state utility, productivity loss, and non-medical costs. Even if some databases may provide pharmacoeconomic data, this review did not attempt to analyze pharmacoeconomic results in T2DM but rather focused on research practices. Analyzing the existing databases does not involve the same study conduct and research methodology as generating new evidence in studies using primary data collection. Therefore, this review focused only on studies with de novo data collection. All eligibility criteria were defined a priori and studies needed to meet all of the inclusion criteria to be eligible. Study selection was performed by two independent reviewers, resolving any disagreements through discussion and consulting a third reviewer if necessary.

### **7.3.2 Data Extraction**

For each included study, the following characteristics were extracted using a data collection form: first author, country, design, patient population, intervention, exposure group, measure of exposure, study site, number of study sites, sample size, sample size calculation, sampling method, source for patient selection, recruitment method, recruitment duration, participation rate, ethics approval, patient's consent, duration of follow-up, data monitoring, study public registration, disclosure of conflicts of interest, source of funding, pharmacoeconomic outcome, number of outcome measures, research instrument, source for measure of outcome, recall period, extent of missing data, and external validity. Studies may have collected clinical outcomes in addition to the pharmacoeconomic outcomes. However, the data extraction focused solely on data pertaining to pharmacoeconomic outcomes, and may not reflect the entire purpose of a study. In addition, results of the studies were not presented as this review aimed at characterizing methodological practices only. Data extraction was performed by two independent reviewers, resolving any disagreements through discussion and consulting a third reviewer if necessary. Extracted data were presented as descriptive statistics using frequencies.

### **7.3.3 Quality Assessment**

The risk of bias in the results of comparative studies was assessed using the ROBINS-I tool (Risk Of Bias In Non-randomized Studies of Interventions) (7). As for descriptive studies, the quality was assessed using the Joanna Briggs Institute (JBI) critical appraisal tool for descriptive studies (8). Critical appraisal was performed for each pharmacoeconomic outcome by two independent reviewers, resolving any disagreements through discussion and consulting a third reviewer if necessary. The ROBINS-I tool covers seven domains: confounding, selection of participants, classification of interventions, deviations from intended interventions, missing data, measurement of outcomes, and selection of the reported result. Domain-level and overall risk of bias judgments were made according to the ROBINS-I guidance. Results were summarized in a weighted summary plot (9). The JBI critical appraisal criteria for descriptive studies are: random sample, clear inclusion criteria, identification of confounding factors, outcome assessment

validity, sufficient description of groups, sufficient time period, inclusion of patients who withdrew, reliable method for outcome measurement, and appropriate statistical analysis. The quality score was reported in percentage as not all criteria applied to all studies.

#### **7.3.4 Results**

The literature search identified 1,158 potentially relevant studies, among which 48 were duplicates. The 1,110 remaining articles were screened by titles and abstracts. Of these, 171 full-text articles were assessed for eligibility. After excluding 111 studies not meeting eligibility criteria, 60 studies were included in the literature review (Supplementary Figure 1).

Table 1 summarizes study characteristics of included articles. Thirty-six studies (60 percent) assessed resource utilization, twenty-two studies (37 percent) collected utility values, and eight studies (13 percent) reported productivity loss. A typical diabetic population was included in 48 percent of studies, whereas 35 percent restricted the study sample to patients receiving a specific intervention, 8 percent to patients with a specific complication/comorbidity, and 8 percent to middle age and/or elderly. Twenty-five studies (42 percent) evaluated a single group whereas thirty-five studies (58 percent) included at least two exposure groups, among which seven were based on medical interventions. The main study design was cross-sectional (68 percent) followed by prospective cohort studies (18 percent). Most studies were performed in hospital outpatient clinics (57 percent), followed by primary care (25 percent) and hospital inpatient departments (15 percent). Sample sizes varied between <50 patients (8 percent), 50–100 patients (8 percent), 101–250 patients (27 percent), 251–500 patients (28 percent), 501–1,000 patients (17 percent), and >1,000 patients (12 percent). Sample size calculation was precision-based for 18 percent of studies and power-based for 15 percent of studies. A non-probability sampling method was used in 68 percent of studies whereas 15 percent used a probability sampling. Ten studies (17 percent) reported participation rates below 80 percent whereas eleven studies (18 percent) reported participation rates of 80 percent and over. Complete data extraction was not possible for many studies. Indeed, exposure measurement was unclear for 8 percent of studies. Also, 20 percent of

studies did not describe the source for patient selection, 13 percent did not report the number of study sites, and 38 percent did not report recruitment duration. The sample size calculation and the sampling method were not explained in 67 and 17 percent of studies, respectively. Twenty-two percent of studies did not mention being approved by an ethics committee and 8 percent did not mention obtaining patient's consent. Several studies did not disclose conflicts of interest (13 percent) or funding (20 percent). Moreover, 22 percent of studies did not demonstrate any evidence of a formal protocol (i.e., no study public registration and no institutional approval). In addition, external validity was not discussed in 25 percent of studies.



**Table 1. Characteristics of studies (n=60)**

<b>Study characteristics</b>	<b>%</b>
<b>Region*</b>	
Asia	62
Europe	18
Africa	12
North America	10
Australia	2
<b>Evidence of a formal protocol</b>	
Institutional review board approval	68
Study public registration	10
No evidence	22
<b>Pharmacoeconomic outcome*</b>	
Resource utilization	60
Utility	37
Productivity loss	13
<b>Study design</b>	
Cross-sectional study	68
Prospective cohort study	18
Uncontrolled before-after study	7
Cross-over study	2
Pragmatic randomized controlled trial	2
Pragmatic cluster-randomized controlled trial	2
Quasi-randomized controlled trial	2
<b>Diabetic population</b>	
Diabetic patients (no additional restriction)	48
Receiving a specific intervention	35
With a specific complication/comorbidity	8
Middle age and/or elderly	8
<b>Exposure group</b>	
Single group	42
Based on health states	17
Based on medical interventions	12
With vs. without diabetes	10
Based on patient characteristics	7
Pre- vs. post-intervention	7
Based on research instruments	3
With vs. without health care program	3

**Table 1 (continued)**

<b>Study characteristics</b>	<b>%</b>
<b>Classification of exposure*</b>	
Not applicable (single group)	42
Extracted from medical file	27
Measured by investigators	12
Self-reported	10
Assigned by investigators	7
Not mentioned	8
<b>Study site*</b>	
Hospital outpatient clinic	57
Primary care	25
Hospital inpatient department	15
Online	10
Supermarket	2
Worksite	2
<b>Number of sites</b>	
One site only	43
< 5	12
5-10	8
≥ 20	13
Online	10
Not mentioned	13
<b>Sample size</b>	
< 50	8
50-100	8
101-250	27
251-500	28
501-1000	17
> 1000	12
<b>Sample size calculation</b>	
Precision-based calculation	18
Power-based calculation	15
Not mentioned	67
<b>Sampling method</b>	
Non-probability sampling	68
Probability sampling	15
Not mentioned	17

**Table 1 (continued)**

<b>Study characteristics</b>	<b>%</b>
<b>Source for patient selection*</b>	
Database/medical file	38
Patients attending the study site	30
Web-based panel	8
Inpatient admissions	5
Friends/family suggested by patients	3
Online ads and social media	3
Not mentioned	20
<b>Recruitment method*</b>	
Face-to-face	40
Over the phone	15
By email	10
By mail	5
Online ads and social media	3
Not mentioned	38
<b>Recruitment duration</b>	
< 6 months	28
6-12 months	17
> 12 months	17
Not mentioned	38
<b>Participation rate</b>	
< 80%	17
≥ 80%	18
Not mentioned	65
<b>Ethics approval</b>	
Yes	78
Not mentioned	22
<b>Consent</b>	
Written informed consent	52
Informed consent not further specified	22
Verbal consent	7
Electronic consent	7
Completion of questionnaires	5
Not mentioned	8

**Table 1 (continued)**

<b>Study characteristics</b>	<b>%</b>
<b>Duration of follow-up</b>	
Not applicable (cross-sectional)	68
≤ 6 months	13
12-18 months	8
> 2 years	3
Treatment duration	3
Hospital stay	3
<b>Data monitoring</b>	
Yes	13
Not mentioned	87
<b>Conflict of interest</b>	
Disclosed	87
Not disclosed	13
<b>Funding</b>	
No industry funding	62
Industry funding	18
Not mentioned	20
<b>External validity discussed</b>	
Yes	75
No	25

\*Studies may be classified in more than one category, that is, exceed 100%.

Characteristics of pharmacoeconomic outcomes are presented in Table 2. Among the thirty-six studies assessing resource utilization, 39 percent were comparative and 61 percent were descriptive only. Eleven percent did not mention the source for outcome collection and 3 percent did not report the mode of administration. Only 11 percent of the research instruments were validated forms, namely the International Diabetic Federation Diabetes Impact questionnaire (3 percent), the Rapid Assessment of Avoidable Blindness - Diabetic Retinopathy survey (3 percent), the Behavioral Risk Factor Surveillance System survey (3 percent) and the Health and Social Services Utilization Inventory (3 percent). Twenty-two studies reported utility values, among which 68 percent were comparative. Only validated questionnaires were used to collect utility, namely EQ-5D index (43 percent), EQ-5D index and visual analog scale (VAS)(26 percent), VAS only (13 percent), SF-6D (9 percent), standard gamble (4 percent) and time trade-off (4 percent). The mode of administration was unclear in 23 percent of the studies. Six (75 percent) of the eight studies evaluating productivity loss were descriptive only. Research instruments were researcher developed forms for 75 percent of studies collecting productivity loss, whereas 13 percent used the Endicott Work Productivity scale and 13 percent used the International Diabetic Federation Diabetes Impact questionnaire. The extent of missing data was not reported for 22 percent of the resource utilization outcomes, 23 percent of the utility outcomes and 38 percent of the productivity loss outcomes.

**Table 2. Characteristics of pharmacoeconomic outcomes**

<b>Characteristics</b>	<b>% of included studies</b>		
	<b>Resource utilization (n=36)</b>	<b>Utility (n=22)</b>	<b>Productivity loss (n=8)</b>
<b>Research question</b>			
Comparative	39	68	25
Descriptive	61	32	75
<b>Number of measures</b>			
One measure	94	73	88
Repeated measures	3	5	13
One measure pre-test, one measure post-test	3	14	0
One pre-test, repeated measures post-test	0	9	0
<b>Mode of administration*</b>			
Face-to-face interview	47	32	38
Data extraction by researcher	39	0	0
Self-completion (electronic form)	14	5	63
Self-completion (paper form)	8	32	0
Telephonic interview	3	9	13
Not mentioned	3	23	0
<b>Research instrument</b>			
Researcher developed form	89	0	75
Validated questionnaire	11	100	25
<b>Source for outcome collection*</b>			
Patient	69	100	100
Medical file	25	0	0
Caregiver	14	0	25
Health professional	6	0	0
Not mentioned	11	0	0
<b>Recall period</b>			
Same day	11	95	13

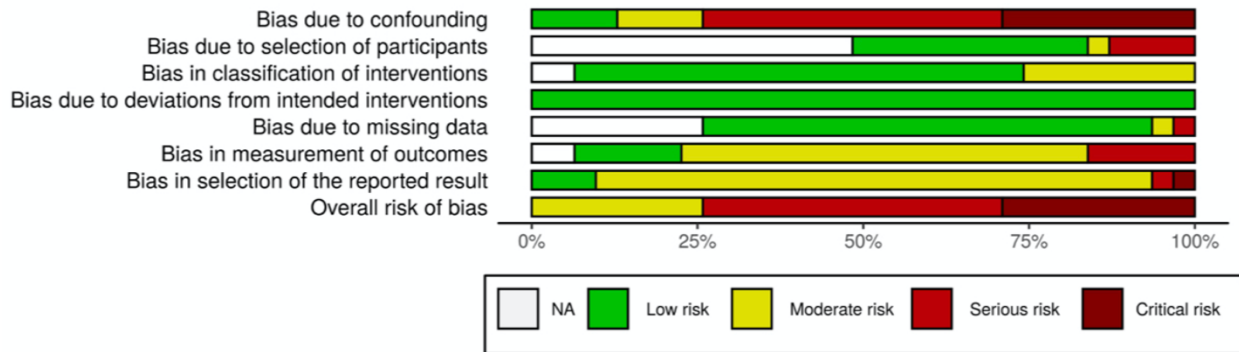
≤ 1 month	6	5	25
3 months	14	0	25
6 months	8	0	0
One year	31	0	38
Not mentioned	6	0	0
Objective measure of outcome	25	0	0
<b>Extent of missing data</b>			
< 80% completeness	6	9	13
≥ 80% completeness	72	68	50
Not mentioned	22	23	38

---

\*Studies may be classified in more than one category, i.e. exceed 100%

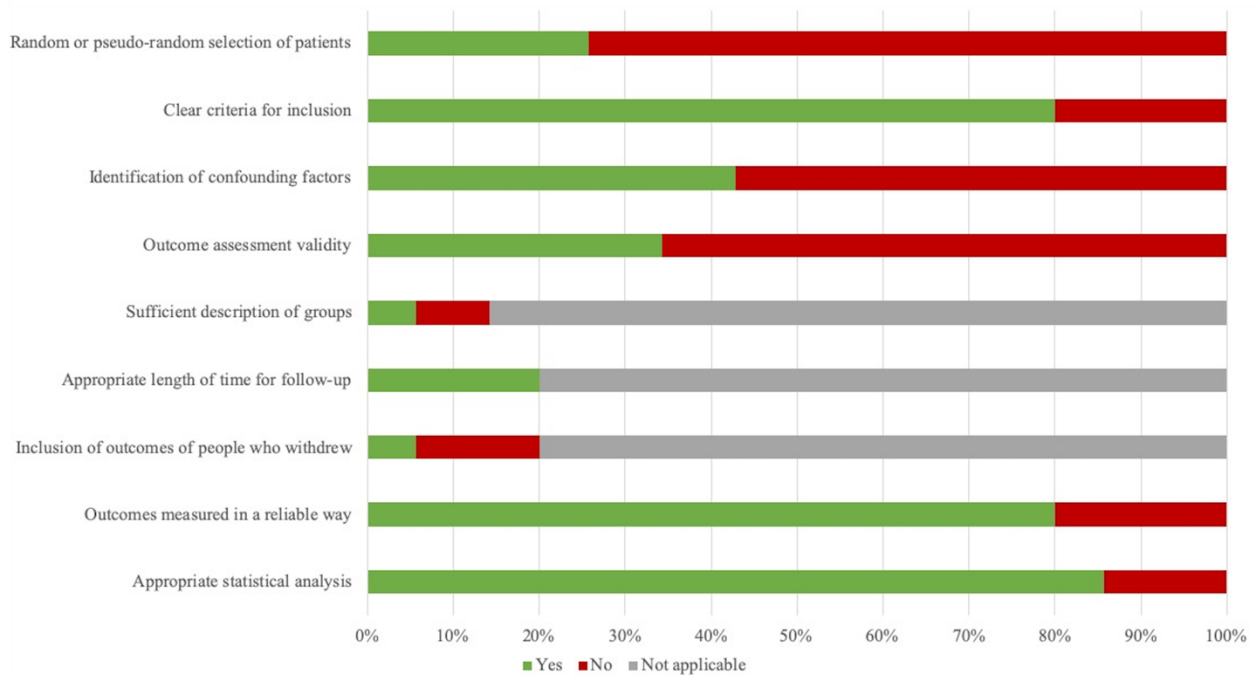
None of the comparative studies were classified at low risk of bias. Among the thirty-one analytical outcomes, 26 percent were judged at moderate risk of bias, 45 percent at serious risk and 29 percent at critical risk. The domains the most at risk of bias were confounding, outcome measurement and reporting of results. Half of the studies did not report sufficient information on which to base a judgement about the risk of selection bias. No information was reported about exposure measurement, outcome assessment and missing data in 6 percent, 6 percent and 26 percent of the studies, respectively (Figure 1). Risk of bias for each comparative outcome is presented in the electronic supplementary material (Supplementary Table 2).

**Figure 1. Risk of bias in comparative studies**



None of the descriptive studies met all the JBI critical appraisal criteria. Among the thirty-five descriptive outcomes, 6 percent had a quality score below 25 percent, 17 percent a score between 25 percent-49 percent, 37 percent a score between 50 percent-74 percent and 40 percent a score between 75 percent-90 percent. The criteria that were less likely to be met were random/pseudo-random selection of patients and outcome assessment validity (Figure 2). Quality assessment for each descriptive outcome is presented in the electronic supplementary material (Supplementary Table 3).

**Figure 2. Quality assessment of descriptive studies**



### 7.3.5 Discussion

The aim of this study was to characterize methodological practices of studies collecting pharmaco-economic outcomes in a real-world setting for the management of T2DM. This objective broadened our knowledge on how this type of study is performed and documented whether concerns about RWE procedural practices are true or not. Results of this review showed that most studies were cross-sectional, performed in hospital outpatient clinics, and measured self-reported pharmaco-economic outcomes. Another noticeable observation was that complete data extraction was not possible for many studies, showing a lack of transparency. Indeed, several studies did not clearly explain the methods of outcome and exposure assessment. A high proportion of studies did not describe sample selection, recruitment duration, number of sites, sample size calculation, and sampling method. Several studies did not demonstrate evidence of a formal protocol, and did not disclose conflict of interest and funding. Moreover, not all studies reported being approved by an

ethics committee and obtaining patient' s consent. Participation rates and missing data were also often lacking.

Quality assessment demonstrated that a significant proportion of the included studies do not provide robust evidence. Caution should therefore be exercised when interpreting results. Poor quality scores and high risk of bias question the usefulness of the results to complement RCTs and to provide credible economic evaluations. Unfortunately, this review corroborates concerns about RWE practices. Previously voiced concerns include uncertainty about internal validity, inaccurate recording of health events, missing data, opaque reporting of conduct and results, unmeasured confounding, measurement error, model misspecification, selection bias, and even fraud (3;10;11). As demonstrated by the current results, many of those threats were found to be true. It is nonetheless unfortunate that many studies lack credibility because RWE is a promising complement to RCT findings. Therefore, it is worth pursuing efforts to increase confidence in RWE studies.

The International Society for Pharmacoepidemiology (ISPE) and ISPOR have paved the way to good practices for RWE studies, yet their recommendations are not widely adopted (3;5). In addition, the Patient-Centered Outcomes Research Institute (PCORI) (12;13) and the Agency for Healthcare Research and Quality (AHRQ) (14) published important guidelines for prospective observational studies. Nevertheless, they focus on clinical outcomes. The ISPOR Task Force (15) and the AHRQ (16) also provided guidance on the conduct of prospective observational studies, but specifically for comparative treatment effectiveness. The National Institute for Health and Care Excellence (NICE) also made recommendations on the use of RWD for comparative treatment effectiveness (17). The other guidelines for prospective observational research comprise those developed for pharmacovigilance and pharmacoepidemiology methods (18). As for pharmaco-economic outcomes, the ISPOR Task Force published six reports on good practices for measuring drug costs in cost-effectiveness analyses (18). However, their recommendations mainly address the use and calculation of costs, and not primary data collection. There are other ISPOR Task Force reports adapted to pharmaco-economic outcomes, but they focus on budget impact analyses, conjoint analyses, and transferability of economic information (18). The other guidelines



that have been identified are used in synthesizing evidence rather than in generating evidence. Mostly, they provide guidance on evidence synthesis research methods, analysis methods, and the use of evidence in healthcare decisions. There are also several guidelines on reporting of evidence (18). An important publication from the ISPOR Task Force consists of a questionnaire designed to assess the relevance and credibility of observational studies (19). The questionnaire helps identify strengths, weaknesses and fatal flaws, and was developed to assist decision makers in evaluating studies. Although highly relevant to assess the quality of published studies, it was not exactly designed for investigators.

To our knowledge, there are no guidelines focusing on methodologies to collect pharmaco-economic outcomes in a real-world setting specifically. The results of this review clearly highlighted that RWE is commonly generated using poor quality methodology, at least in T2DM. Guidelines specifically adapted to pharmaco-economic outcomes may help future investigators in performing credible RWE studies for reimbursement decision making. Adapting procedural practices to T2DM, or prevalent chronic diseases, would increase relevance and usefulness for researchers.

This systematic review was conducted according to a predefined protocol and was performed by two independent reviewers. Because a transparent methodology was used, this review could be replicated, allowing the assessment of RWE practices in other clinical settings. As part of the limitations, gray literature was not screened. Moreover, a short timeframe was used for study selection. A longer observation period might have retrieved different methodologies than those of this review. Nevertheless, extending the study period would only have allowed for the retrieval of even more heterogeneous methodologies, thus worsening the lack of credibility and uniformity already demonstrated in the current review. Because this research field has greatly evolved through the years, limiting the study selection to the most recent years has permitted to provide a current overview. In addition, only data and methodologies related to pharmaco-economic outcomes were extracted, even if clinical outcomes were reported in the study. As per eligibility criteria, findings of this review only reflect practices in primary data collection.

This review was limited to T2DM for pointing out the shortcomings in collecting RWD. Any other common chronic diseases might have also been chosen and the same conclusions would have been expected. Indeed, methodological practices used for collecting pharmaco-economic outcomes were not specific to T2DM, and would probably not differ in studies assessing similar conditions. Still, it would be interesting to analyze RWE studies in other therapeutic fields to confirm this hypothesis.

### **7.3.6 Conclusion**

RWD are complementary evidence in reimbursement decision making, yet the conduct of RWE studies has not been standardized. Findings of this literature review corroborate the lack of credibility and transparency in RWE methodologies. There is a need for good procedural practices that can increase confidence in RWE studies. Guidelines adapted to RWE studies collecting pharmaco-economic data in T2DM could help future investigators in performing good quality research.

### 7.3.7 References

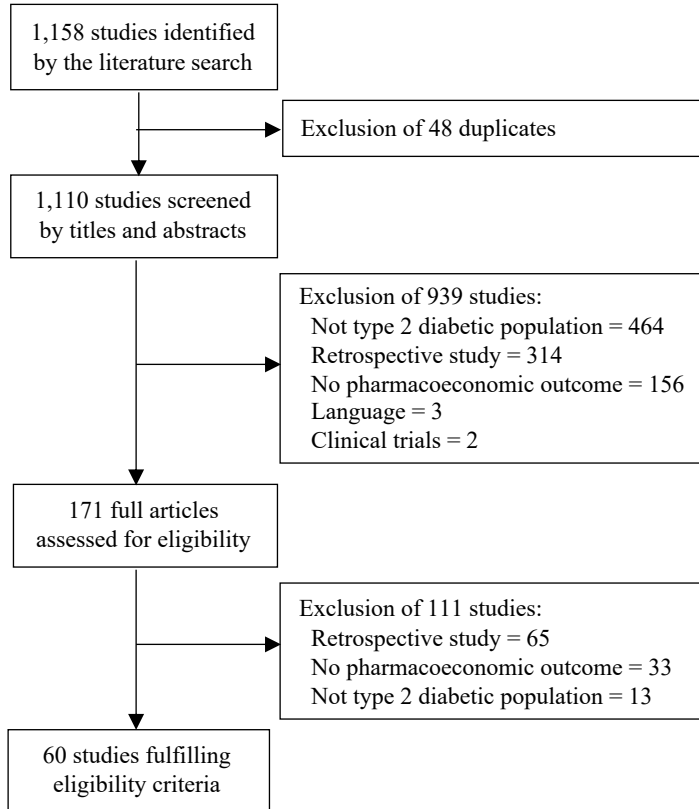
1. Garrison LP Jr, Neumann PJ, Erickson P, Marshall D, Mullins CD. Using real-world data for coverage and payment decisions: The ISPOR Real-World Data Task Force report. *Value Health*. 2007;10:326–35.
2. Makady A, de Boer A, Hillege H, Klungel O, Goettsch W. What is real-world data? A review of definitions based on literature and stakeholder interviews. *Value Health*. 2017;20:858–65.
3. Berger ML, Sox H, Willke RJ, Brixner DL, Eichler H-G, Goettsch W, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR-ISPE special task force on real-world evidence in health care decision making. *Value Health*. 2017;20:1003–08.
4. Greenfield S. Making real-world evidence more useful for decision making. *Value Health*. 2017;20:1023–24.
5. Wang SV, Schneeweiss S, Berger ML, Brown J, de Vries F, Douglas I, et al. Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies V1.0. *Value Health*. 2017;20:1009–22.
6. World Health Organization (WHO). *Global report on diabetes*. Geneva: WHO; 2016.
7. Sterne JAC, Hernan MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *Br Med J*. 2016;355:i4919.
8. Lockwood C, Sfetcu R, Oh EG. *Synthesizing quantitative evidence*. Lippincott Williams & Wilkins; 2011. Australia: Wolters Kluwer.
9. McGuinness LA. robvis: An R package and web application for visualizing risk-of-bias assessments [cited 2020 Apr 26]; 2019. Available from: <https://github.com/mcguinlu/robvis>.
10. Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005;2:e124.

11. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-world evidence—what is it and what can it tell us? *N Engl J Med*. 2016;375:2293–97.
12. Patient-Centered Outcomes Research Institute (PCORI). Methodology standards [cited 2020 Apr 26]; 2019. Available from: <https://www.pcori.org/sites/default/files/PCORI-Methodology-Standards.pdf>.
13. Patient-Centered Outcomes Research Institute (PCORI). Our questions, our decisions: standards for patient-centered outcomes research [cited 2020 Apr 26]; 2012. Available from: <https://www.pcori.org/assets/Methodology-Report-072312.pdf>.
14. Agency for Healthcare Research and Quality (AHRQ). 21st century patient registries—registries for evaluating patient outcomes: A user’s guide 3rd Edition, Addendum [cited 2020 Apr 26]; 2018. Available from: [https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/registries-guide-3rd-ed-addendum-research-2018-revised\\_0.pdf](https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/registries-guide-3rd-ed-addendum-research-2018-revised_0.pdf).
15. Berger ML, Dreyer N, Anderson F, Towse A, Sedrakyan A, Normand SL. Prospective observational studies to assess comparative effectiveness: The ISPOR good research practices task force report. *Value Health*. 2012;15: 217–30.
16. Agency for Healthcare Research and Quality (AHRQ). Developing a protocol for observational comparative effectiveness research: A user’s guide [cited 2020 Apr 26]; 2013. Available from: [https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/observational-cer-protocol\\_research.pdf](https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/observational-cer-protocol_research.pdf).
17. Bell H, Wailoo AJ, Hernandez M, Grieve R, Faria R, Gibson L, et al. The use of real world data for the estimation of treatment effects in NICE decision making [cited 2020 Apr 26]; 2016. Available from: <http://nicedsu.org.uk/wp-content/uploads/2018/05/RWD-DSU-REPORT-Updated-DECEMBER-2016.pdf>.
18. International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Guideline index for outcomes research and use in health care decision making [cited 2020 Apr 26]; 2020. Available from: <https://www.ispor.org/heor-resources/more-heor-resources/outcomes-research-guidelines-index>.

19. Berger ML, Martin BC, Husereau D, Worley K, Allen JD, Yang W, et al. A questionnaire to assess the relevance and credibility of observational studies to inform health care decision making: An ISPOR- AMCP-NPC Good Practice Task Force report. *Value Health*. 2014;17: 143–56.

### 7.3.8 Supplementary material

#### Supplementary Figure 1 – Flowchart of study selection



#### Supplementary table 1 – Literature search strategy

##### Literature search strategy from Ovid MEDLINE database

- 1 Epidemiologic Studies/
- 2 exp Cohort Studies/
- 3 Cross-Sectional Studies/
- 4 (epidemiologic adj (study or studies)).ab,ti.
- 5 (cohort adj (study or studies)).ab,ti.
- 6 cross sectional.ab,ti.
- 7 cohort analy\$.ab,ti.
- 8 (follow up adj (study or studies)).ab,ti.
- 9 longitudinal.ab,ti.
- 10 prospective\$.ab,ti.
- 11 (observ\$ adj3 (study or studies)).ab,ti.
- 12 adverse effect?.ab,ti.
- 13 multicenter study/

14 (pragmatic clinical trial or pragmatic trial).mp.  
15 exp questionnaires/  
16 exp health survey/ or exp health care survey/  
17 (survey or surveys or questionnaire or questionnaires).mp.  
18 epidemiologic methods/  
19 cohort\$.tw.  
20 (real world or real\*world).ti.  
21 (real life or real\*life).ti.  
22 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or  
23 (utilit\$ approach\$ or health gain or hui or hui2 or hui 2 or hui3 or hui 3).ti,ab.  
24 (health measurement\$ scale\$ or health measurement\$ questionnaire\$).ti,ab.  
25 (standard gamble\$ or categor\$ scal\$ or linear scal\$ or linear analog\$ or visual scal\$ or magnitude  
26 (time trade off\$ or rosser\$ classif\$ or rosser\$ matrix or rosser\$ distress\$ or hrqol).ti,ab.  
27 (index of wellbeing or quality of wellbeing or qwb).ti,ab.  
28 (multiattribute\$ health ind\$ or multi attribute\$ health ind\$).ti,ab.  
29 (health utilit\$ index or health utilit\$ indices).ti,ab.  
30 (multiattribute\$ theor\$ or multi attribute\$ theor\$ or multiattribute\$ analys\$ or multi attribute\$  
31 (health utilit\$ scale\$ or classification of illness state\$ or 15 dimension).ti,ab.  
32 health state\$ utilit\$.ti,ab.  
33 well year\$.ti,ab.  
34 (multiattribute\$ utilit\$ or multi attribute\$ utilit\$).ti,ab.  
35 health utilit\$ scale\$.ti,ab.  
36 (euro qol or euro qual or eq-5d or eq5d or eq 5d or euroqol or euroqual).ti,ab.  
37 (qaly or qaly or qualys or qalys or quality adjusted life year\$).ti,ab.  
38 (hye or hyes or health\$ year\$ equivalent\$).ti,ab.  
39 (person trade off\$ or person tradeoff\$ or time tradeoff\$ or time trade off\$).ti,ab.  
40 theory utilit\$.ti,ab.  
41 (sf 6d or short from 6d).ti,ab.  
42 (sf6 or sf 6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six or  
43 Quality-Adjusted Life Years/  
44 quality adjusted life.ti,ab,kf.  
45 (qaly\* or qald\* or qale\* or qtime\* or life year or life years).ti,ab,kf.  
46 (health adj3 utilit\*).ti,ab,kf.  
47 (utilit\* adj3 (valu\* or measur\* or health or life or estimat\* or elicit\* or disease or score\* or  
48 (preference\* adj3 (valu\* or measur\* or health or life or estimat\* or elicit\* or disease or score\* or  
49 disutilit\*.ti,ab,kf.  
50 standard gamble\*.ti,ab,kf.  
51 tto.ti,ab,kf.  
52 (hui or hui1 or hui2 or hui3).ti,ab,kf.  
53 (eq or euroqol or euro qol or eq5d or eq 5d or euroqual or euro qual).ti,ab,kf.  
54 rosser.ti,ab,kf.  
55 (health\* adj2 year\* adj2 equivalent\*).ti,ab,kf.  
56 (quality of wellbeing or quality of well being or index of wellbeing or index of well being or  
57 (time trade off or time tradeoff).ti,ab,kf.  
58 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or  
59 (productivity loss or productivity losses).mp.  
60 absenteeism.mp. or exp Absenteeism/  
61 presenteeism.mp. or exp Presenteeism/  
62 work productivity.mp.  
63 activity impairment.mp.  
64 (work loss or work\*loss).mp.  
65 work disability.mp.  
66 Workplace Activity Limitations Scale.mp.  
67 Work Limitations Questionnaire.mp.  
68 Work Ability Index.mp.  
69 Work Productivity Survey.mp.  
70 (Work Productivity and Activity Impairment).mp.

71 WPAI.mp.  
72 Endicott Work Productivity Scale.mp.  
73 (Health and Labour Questionnaire).mp.  
74 (Health and Work Performance Questionnaire).mp.  
75 (Health and Work Questionnaire).mp.  
76 (Lam Employment Absence and Productivity Scale).mp.  
77 Sheehan Disability Scale.mp.  
78 Stanford Presenteeism Scale.mp.  
79 (Work and Health Interview).mp.  
80 Worker Productivity Index.mp.  
81 Work Productivity Short Inventory.mp.  
82 (sick leave or medical leave).mp.  
83 Work Capacity Evaluation.mp. or exp work capacity/  
84 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70 or 71 or 72 or 73 or 74 or 75 or  
85 \*health care utilization/  
86 \*Health Resources/  
87 \*Health Care Costs/  
88 \*"Length of Stay"/ or \*Hospitalization/  
89 (health system utili\*ation or "health system use" or medical system utili\*ation or "medical system  
90 Economics/  
91 exp "Costs and Cost Analysis"/  
92 Economics, Nursing/  
93 Economics, Medical/  
94 Economics, Pharmaceutical/  
95 exp Economics, Hospital/  
96 (economic\* or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic\*  
97 (economic\* or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic\*  
98 (cost\* adj2 (effective\* or utilit\* or benefit\* or minimi\* or analy\* or outcome or outcomes)).ab,kf.  
99 85 or 86 or 87 or 88 or 89 or 90 or 91 or 92 or 93 or 94 or 95 or 96 or 97 or 98  
100 letter/  
101 editorial/  
102 news/  
103 exp historical article/  
104 anecdotes as topic/  
105 comment/  
106 case report/  
107 (letter or comment\*).ti.  
108 animals/ not humans/  
109 exp animals, laboratory/  
110 exp animal experimentation/  
111 exp models, animal/  
112 exp rodentia/  
113 (rat or rats or mouse or mice).ti.  
114 (reliability or reproducib\* or inter rate or interrater or valid\* or test retest).mp.  
115 exp Retrospective Studies/  
116 retrospective\$.ab,ti.  
117 (database analy\* or secondary data).ti,ab.  
118 Randomized Controlled Trials as Topic/  
119 randomized controlled trial/  
120 clinical trial/  
121 clinical trial, phase i.pt.  
122 clinical trial, phase ii.pt.  
123 clinical trial, phase iii.pt.  
124 clinical trial, phase iv.pt.  
125 controlled clinical trial.pt.  
126 randomized controlled trial.pt.  
127 clinical trial.pt.



128 exp Clinical Trials as topic/  
129 (clinical adj trial\$.tw.  
130 PLACEBOS/  
131 placebo\$.tw.  
132 meta-analysis.pt.  
133 meta-analysis/ or systematic review/ or meta-analysis as topic/ or "meta analysis (topic)"/ or  
134 ((systematic\* adj3 (review\* or overview\*)) or (methodologic\* adj3 (review\* or  
135 ((quantitative adj3 (review\* or overview\* or synthes\*)) or (research adj3 (integrati\* or  
136 ((integrative adj3 (review\* or overview\*)) or (collaborative adj3 (review\* or overview\*)) or (pool\*  
137 (data synthes\* or data extraction\* or data abstraction\*).ti,ab,kf,kw.  
138 (handsearch\* or hand search\*).ti,ab,kf,kw.  
139 (mantel haenszel or peto or der simonian or dersimonian or fixed effect\* or latin  
140 (met analy\* or metanaly\* or technology assessment\* or HTA or HTAs or technology overview\* or  
141 (meta regression\* or metaregression\*).ti,ab,kf,kw.  
142 (meta-analy\* or metaanaly\* or systematic review\* or biomedical technology assessment\* or bio-  
143 (medline or cochrane or pubmed or medllars or embase or cinahl).ti,ab,hw.  
144 (cochrane or (health adj2 technology assessment) or evidence report).jw.  
145 (comparative adj3 (efficacy or effectiveness)).ti,ab,kf,kw.  
146 (outcomes research or relative effectiveness).ti,ab,kf,kw.  
147 ((indirect or indirect treatment or mixed-treatment) adj comparison\*).ti,ab,kf,kw.  
148 (reference list\* or bibliograph\* or hand search\* or manual search\* or relevant journals).ab.  
149 (search strategy or search criteria or systematic search or study selection or data extraction).ab.  
150 (search\* adj4 literature).ab.  
151 (medline or pubmed or cochrane or embase or psychlit or psychlit or psychinfo or psycinfo or cinahl  
152 (claims database or administrative database or administrative data).ti,ab.  
153 100 or 101 or 102 or 103 or 104 or 105 or 106 or 107 or 108 or 109 or 110 or 111 or 112 or 113 or  
154 58 or 84 or 99  
155 22 and 154  
156 155 not 153  
157 (diabetic or diabetes).ti,ab,kw.  
158 156 and 157  
159 limit 158 to ((english or french) and yr="2017-Current")

### Literature search strategy from Ovid EMBASE database

1 Epidemiologic Studies/  
2 exp Cohort Studies/  
3 Cross-Sectional Studies/  
4 (epidemiologic adj (study or studies)).ab,ti.  
5 (cohort adj (study or studies)).ab,ti.  
6 cross sectional.ab,ti.  
7 cohort analy\$.ab,ti.  
8 (follow up adj (study or studies)).ab,ti.  
9 longitudinal.ab,ti.  
10 prospective\$.ab,ti.  
11 (observ\$ adj3 (study or studies)).ab,ti.  
12 adverse effect?.ab,ti.  
13 Clinical study/  
14 Family study/  
15 Longitudinal study/  
16 Prospective study/  
17 Cohort analysis/  
18 (Cohort adj (study or studies)).mp.  
19 (follow up adj (study or studies)).tw.

20 (observational adj (study or studies)).tw.  
 21 (epidemiologic\$ adj (study or studies)).tw.  
 22 (cross sectional adj (study or studies)).tw.  
 23 multicenter study/  
 24 (pragmatic clinical trial or pragmatic trial).mp.  
 25 exp questionnaires/  
 26 exp health survey/ or exp health care survey/  
 27 (survey or surveys or questionnaire or questionnaires).mp.  
 28 exp follow up/  
 29 cohort\$.tw.  
 30 exp longitudinal study/  
 31 (real world or real\*world).ti.  
 32 (real life or real\*life).ti.  
 33 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or  
 34 (utilit\$ approach\$ or health gain or hui or hui2 or hui 2 or hui3 or hui 3).ti,ab.  
 35 (health measurement\$ scale\$ or health measurement\$ questionnaire\$).ti,ab.  
 36 (standard gamble\$ or categor\$ scal\$ or linear scal\$ or linear analog\$ or visual scal\$ or magnitude  
 37 (time trade off\$ or rosser\$ classif\$ or rosser\$ matrix or rosser\$ distress\$ or hrqol).ti,ab.  
 38 (index of wellbeing or quality of wellbeing or qwb).ti,ab.  
 39 (multiattribute\$ health ind\$ or multi attribute\$ health ind\$).ti,ab.  
 40 (health utilit\$ index or health utilit\$ indices).ti,ab.  
 41 (multiattribute\$ theor\$ or multi attribute\$ theor\$ or multiattribute\$ analys\$ or multi attribute\$  
 42 (health utilit\$ scale\$ or classification of illness state\$ or 15 dimension).ti,ab.  
 43 health state\$ utilit\$.ti,ab.  
 44 well year\$.ti,ab.  
 45 (multiattribute\$ utilit\$ or multi attribute\$ utilit\$).ti,ab.  
 46 health utilit\$ scale\$.ti,ab.  
 47 (euro qol or euro qual or eq-5d or eq5d or eq 5d or euroqol or euroqual).ti,ab.  
 48 (qualy or qaly or qualys or qalys or quality adjusted life year\$).ti,ab.  
 49 (hye or hyes or health\$ year\$ equivalent\$).ti,ab.  
 50 (person trade off\$ or person tradeoff\$ or time tradeoff\$ or time trade off\$).ti,ab.  
 51 theory utilit\$.ti,ab.  
 52 (sf 6d or short form 6d).ti,ab.  
 53 Quality-Adjusted Life Year/  
 54 quality adjusted life.ti,ab,kw.  
 55 (qaly\* or qald\* or qale\* or qtime\* or life year or life years).ti,ab,kw.  
 56 (sf6 or sf 6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six or  
 57 (health\* adj2 year\* adj2 equivalent\*).ti,ab,kw.  
 58 (health adj3 utilit\*).ti,ab,kw.  
 59 (utilit\* adj3 (valu\* or measur\* or health or life or estimat\* or elicit\* or disease or score\* or  
 60 (preference\* adj3 (valu\* or measur\* or health or life or estimat\* or elicit\* or disease or score\* or  
 61 disutilit\*.ti,ab,kw.  
 62 rosser.ti,ab,kw.  
 63 standard gamble\*.ti,ab,kw.  
 64 tto.ti,ab,kw.  
 65 (time trade off or time tradeoff).ti,ab,kw.  
 66 (hui or hui1 or hui2 or hui3).ti,ab,kw.  
 67 (eq or euroqol or euro qol or eq5d or eq 5d or euroqual or euro qual).ti,ab,kw.  
 68 (quality of wellbeing or quality of well being or index of wellbeing or index of well being or  
 69 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or  
 70 (productivity loss or productivity losses).mp.  
 71 absenteeism.mp. or exp Absenteeism/  
 72 presenteeism.mp. or exp Presenteeism/  
 73 work productivity.mp.  
 74 activity impairment.mp.  
 75 (work loss or work\*loss).mp.  
 76 work disability.mp.

77 Workplace Activity Limitations Scale.mp.  
78 Work Limitations Questionnaire.mp.  
79 Work Ability Index.mp.  
80 Work Productivity Survey.mp.  
81 (Work Productivity and Activity Impairment).mp.  
82 WPAI.mp.  
83 Endicott Work Productivity Scale.mp.  
84 (Health and Labour Questionnaire).mp.  
85 (Health and Work Performance Questionnaire).mp.  
86 (Health and Work Questionnaire).mp.  
87 (Lam Employment Absence and Productivity Scale).mp.  
88 Sheehan Disability Scale.mp.  
89 Stanford Presenteeism Scale.mp.  
90 (Work and Health Interview).mp.  
91 Worker Productivity Index.mp.  
92 Work Productivity Short Inventory.mp.  
93 (sick leave or medical leave).mp.  
94 Work Capacity Evaluation.mp. or exp work capacity/  
95 70 or 71 or 72 or 73 or 74 or 75 or 76 or 77 or 78 or 79 or 80 or 81 or 82 or 83 or 84 or 85 or 86 or  
96 \*health care utilization/  
97 \*Health Resources/  
98 \*"health care cost"/  
99 \*"Length of Stay"/ or \*Hospitalization/  
100 (health system utili\*ation or "health system use" or medical system utili\*ation or "medical system  
101 Cost/  
102 exp Health Economics/  
103 (economic\* or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic\*  
104 (economic\* or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic\*  
105 (cost\* adj2 (effective\* or utilit\* or benefit\* or minimi\* or analy\* or outcome or outcomes)).ab,kw.  
106 Economics/  
107 96 or 97 or 98 or 99 or 100 or 101 or 102 or 103 or 104 or 105 or 106  
108 letter.pt. or letter/  
109 note.pt.  
110 editorial.pt.  
111 case report/ or case study/  
112 (letter or comment\*).ti.  
113 animal/ not human/  
114 nonhuman/  
115 exp animal experiment/  
116 exp experimental animal/  
117 animal model/  
118 exp rodent/  
119 (rat or rats or mouse or mice).ti.  
120 (reliability or reproducib\* or inter rate or interrater or valid\* or test retest).mp.  
121 exp Retrospective Studies/  
122 retrospective\$.ab,ti.  
123 (database analy\* or secondary data).ti,ab.  
124 Clinical trial/  
125 Randomized controlled trial/  
126 Placebo/  
127 Randomi?ed controlled trial\$.tw.  
128 Rct.tw.  
129 Placebo\$.tw.  
130 factorial\*.ti,ab.  
131 randomized controlled trial/  
132 meta-analysis.pt.  
133 meta-analysis/ or systematic review/ or meta-analysis as topic/ or "meta analysis (topic)"/ or

134 ((systematic\* adj3 (review\* or overview\*)) or (methodologic\* adj3 (review\* or  
135 ((quantitative adj3 (review\* or overview\* or synthes\*))) or (research adj3 (integrati\* or  
136 ((integrative adj3 (review\* or overview\*)) or (collaborative adj3 (review\* or overview\*))) or (pool\*  
137 (data synthes\* or data extraction\* or data abstraction\*).ti,ab,kw.  
138 (handsearch\* or hand search\*).ti,ab,kw.  
139 (mantel haenszel or peto or der simonian or dersimonian or fixed effect\* or latin square\*).ti,ab,kw.  
140 (met analy\* or metanaly\* or technology assessment\* or HTA or HTAs or technology overview\* or  
141 (meta regression\* or metaregression\*).ti,ab,kw.  
142 (meta-analy\* or metaanaly\* or systematic review\* or biomedical technology assessment\* or bio-  
143 (medline or cochrane or pubmed or medlars or embase or cinahl).ti,ab,hw.  
144 (cochrane or (health adj2 technology assessment) or evidence report).jw.  
145 (comparative adj3 (efficacy or effectiveness)).ti,ab,kw.  
146 (outcomes research or relative effectiveness).ti,ab,kw.  
147 ((indirect or indirect treatment or mixed-treatment) adj comparison\*).ti,ab,kw.  
148 ((systematic or evidence) adj2 (review\* or overview\*)).ti,ab.  
149 (reference list\* or bibliograph\* or hand search\* or manual search\* or relevant journals).ab.  
150 (search strategy or search criteria or systematic search or study selection or data extraction).ab.  
151 (search\* adj4 literature).ab.  
152 (medline or pubmed or cochrane or embase or psychlit or psychlit or psychinfo or psycinfo or cinahl  
153 ((pool\* or combined) adj2 (data or trials or studies or results)).ab.  
154 exp Randomized Controlled Trials as Topic/  
155 (claims database or administrative database or administrative data).ti,ab.  
156 108 or 109 or 110 or 111 or 112 or 113 or 114 or 115 or 116 or 117 or 118 or 119 or 120 or 121 or  
157 69 or 95 or 107  
158 33 and 157  
159 158 not 156  
160 (diabetic or diabetes).ti,ab,kw.  
161 159 and 160  
162 limit 161 to ((english or french) and yr="2017-Current")  
163 limit 162 to (embase or medline)

**Supplementary table 2 – Risk of bias in comparative studies**

Author	Country - Population	Exposure groups	Pharmaco-economic outcome	Bias due to confounding	Bias in selection of participants	Bias in classification of exposures	Bias due to departures from intended exposures	Bias due to missing data	Bias in measurement of outcomes	Bias in selection of the reported result	Overall bias
Aro, AK.(1)	Finland – Diabetic patients aged ≥ 65 years	3 groups: -Good glycemic control -Intermediate glycemic control -Poor glycemic control	EQ-5D index	Serious	No information	Low	Low	Low	Moderate	Moderate	Serious
Azmi, S.(2)	Malaysia – T2DM patients with chronic kidney disease	2 groups: -With anemia -Without anemia	Direct medical cost	Serious	No information	Moderate	Low	Low	Serious	Moderate	Serious
Bar-Oz, D.(3)	Israel – T2DM patients scheduled for cataract surgery	Pre and post cataract surgery	EQ-5D index	Serious	Low	Low	Low	No information	Moderate	Serious	Serious
Budiningsari, D.(4)	Indonesia – Diabetic patients	2 groups: -Pictorial dietary assessment tool -Comstock tool	Direct medical cost	Low	Low	Low	Low	Low	Low	Moderate	Moderate
Elrayah-Eliadarous, HA.(5)	Sudan – Diabetic patients and participants without diabetes	2 groups: -With diabetes -Without diabetes	Direct medical and non-medical cost (International Diabetes Federation questionnaire)	Serious	Serious	Moderate	Low	No information	Serious	Moderate	Serious
Elrayah-Eliadarous, HA.(5)	Sudan – Diabetic patients and participants without diabetes	2 groups: -With diabetes -Without diabetes	Being prevented from paid work (International Diabetes Federation questionnaire)	Serious	Serious	Moderate	Low	No information	Serious	Moderate	Serious

Author	Country - Population	Exposure groups	Pharmaco-economic outcome	Bias due to confounding	Bias in selection of participants	Bias in classification of exposures	Bias due to departures from intended exposures	Bias due to missing data	Bias in measurement of outcomes	Bias in selection of the reported result	Overall bias
Emara, KM.(6)	Egypt – Diabetic patients with charcot neuroarthropathy undergoing tibio-talocalcaneal arthrodesis	Pre and post tibio-talocalcaneal arthrodesis	EQ-5D index	Serious	Low	Low	Low	No information	Moderate	Moderate	Serious
Gireesha, KS.(7)	India – T2DM patients receiving insulin	2 groups: -Insulin users -Insulin analogs users	EQ-5D index and visual analog scale	Critical	No information	Low	Low	Low	Moderate	Moderate	Critical
Icks, A.(8)	Germany – Diabetic patients	2 groups: -3-month recall survey -6-month recall survey	Number of physician visits	Low	Low	Low	Low	Low	Low	Moderate	Moderate
Janssen, EM.(9)	USA – T2DM patients	3 groups: -High school or less -Some college -College or more	Willingness-to-pay	Critical	Moderate	Moderate	Low	Low	Moderate	Moderate	Critical
Ji, L.(10)	China – Diabetic patients receiving insulin by pen injection	2 groups: -With lipohypertrophy -Without lipohypertrophy	Cost of therapy	Critical	No information	Moderate	Low	Low	Moderate	Moderate	Critical
Jiao, F.(11)	China – Diabetic patients	2 groups: -Without diabetic complication -With at least one diabetic complication	SF-6D	Moderate	No information	Low	Low	Low	Moderate	Moderate	Moderate
Klinge, MW.(12)	Denmark – Diabetic patients with severe recurrent nausea/vomiting undergoing gastric electrical stimulation	Pre and post gastric electrical stimulation	Cost-utility (SF-6D converted from SF-36)	Serious	Low	Low	Low	Serious	Moderate	Moderate	Serious

Author	Country - Population	Exposure groups	Pharmaco-economic outcome	Bias due to confounding	Bias in selection of participants	Bias in classification of exposures	Bias due to departures from intended exposures	Bias due to missing data	Bias in measurement of outcomes	Bias in selection of the reported result	Overall bias
Kundra, TS.(13)	India – Diabetic patients undergoing coronary artery bypass grafting	2 groups: -Prayer sign positive -Prayer sign negative	Hospital length of stay, length of intensive care unit stay, ventilatory hours	Critical	No information	Low	Low	Low	No information	Moderate	Critical
Laelasari, E.(14)	Indonesia – T2DM patients receiving insulin or sulfonylurea	3 groups: -Insulin -Sulfonylurea -Sulfonylurea + metformin	Direct medical cost	Critical	Low	Low	Low	Low	Low	Moderate	Critical
Lee, MH.(15)	Australia – Diabetic patients and patients without diabetes	2 groups: -With diabetes -Without diabetes	Hospital length of stay	Serious	Low	Low	Low	Low	Low	Moderate	Serious
Lee, Y.(16)	Canada – T2DM (and pre-T2DM) patients	2 groups: -Managerial position -Nonmanagerial position	Work productivity (Endicott Work Productivity Scale)	Serious	No information	Moderate	Low	Low	Moderate	Low	Serious
Mahdavi, M.(17)	Finland, Germany, Greece, Netherlands, Spain, UK – T2DM patients	6 groups: -Regional provider networks from the 6 countries	EQ-5D index	Serious	Serious	Low	Low	Low	Moderate	Moderate	Serious
Mapa-Tassou, C.(18)	Cameroun – T2DM patients and participants without diabetes	2 groups: -With diabetes -Without diabetes	Use of outpatient and inpatient services and use of medicines	Serious	Serious	Moderate	Low	Low	Serious	Moderate	Serious
Markle-Reid, M.(19)	Canada – T2DM patients aged ≥ 65 years with multiple chronic conditions	2 groups: -With program -Without program	Direct medical cost (Health and Social Services Utilization Inventory)	Low	Low	Low	Low	Moderate	Moderate	Low	Moderate
Modarresnia, L.(20)	Iran – T2DM patients	2 groups: -With restless legs syndrome -Without restless legs syndrome	EQ-5D index and visual analog scale	Serious	No information	Low	Low	No information	Moderate	Moderate	Serious

Author	Country - Population	Exposure groups	Pharmaco-economic outcome	Bias due to confounding	Bias in selection of participants	Bias in classification of exposures	Bias due to departures from intended exposures	Bias due to missing data	Bias in measurement of outcomes	Bias in selection of the reported result	Overall bias
Pan, CW.(21)	China – T2DM patients	3 groups: -No diabetic retinopathy -Unilateral retinopathy -Bilateral retinopathy	EQ-5D index	Moderate	No information	Low	Low	Low	Moderate	Moderate	Moderate
Panisch, S.(22)	Austria – T2DM patients	2 groups: -With program -Without program	EQ-5D index and visual analog scale	Low	Low	Low	Low	Low	Moderate	Low	Moderate
Riandini, T.(23)	Singapore – T2DM patients aged ≥ 40 years	2 groups: -With neuropathy -Without neuropathy	EQ-5D index	Moderate	No information	Low	Low	Low	Moderate	Moderate	Moderate
Sankar, V.(24)	India – Diabetic patients receiving amitriptyline or pregabalin for neuropathy	2 groups: -Amitriptyline -Pregabalin	Visual analog scale	Critical	No information	Low	Low	Low	Moderate	Moderate	Critical
Skordis-Worrall, J.(25)	Kyrgyzstan – Diabetic patients and tuberculosis patients	3 groups: -Diabetic patients -Tuberculosis patients -Co-affected patients	Direct medical and non-medical cost	Serious	No information	No information	Low	No information	Serious	Moderate	Serious
Soleymanian, T.(26)	Iran – Patients on maintenance dialysis (diabetic and non-diabetic)	2 groups: -With diabetes -Without diabetes	Hospitalization rate	Critical	No information	No information	Low	Low	No information	Moderate	Critical
Sothornwit, J.(27)	Thailand – Diabetic patients	3 groups: -Diabetic foot -Other diabetic complications -No complication	EQ-5D index	Critical	No information	Low	Low	No information	Moderate	Moderate	Critical



Author	Country - Population	Exposure groups	Pharmacoeconomic outcome	Bias due to confounding	Bias in selection of participants	Bias in classification of exposures	Bias due to departures from intended exposures	Bias due to missing data	Bias in measurement of outcomes	Bias in selection of the reported result	Overall bias
Ugurlar, M.(28)	Turkey – T2DM patients with foot wound undergoing vacuum-assisted closure treatment	2 groups: -Vacuum-assisted closure (Y-connector) -Vacuum-assisted closure (bridge)	Direct medical cost	Critical	Low	Low	Low	Low	Low	Critical	Critical
Xu, RH.(29)	China – Patients with diabetes, heart disease, hypertension and/or cancer	4 groups: -With diabetes -With heart disease -With hypertension -With cancer	EQ-5D index	Moderate	No information	Moderate	Low	No information	Moderate	Moderate	Moderate
Zhu, B.(30)	China – Diabetic patients with retinopathy undergoing cataract surgery	Pre and post cataract surgery	Cost-utility (Time trade off)	Serious	Low	Low	Low	Low	Moderate	Moderate	Serious

## References

1. Aro, A.K., et al., *Glycemic control and health-related quality of life among older home-dwelling primary care patients with diabetes*. Primary Care Diabetes, 2017. **11**(6): p. 577-582.
2. Azmi, S., et al., *The Cost and Quality of Life of Malaysian Type 2 Diabetes Mellitus Patients with Chronic Kidney Disease and Anemia*. Value in Health Regional Issues, 2018. **15**: p. 42-49.
3. Bar-Oz, D., et al., *Glycemic Control and Quality of Life Following Cataract Surgery in Patients with Type 2 Diabetes and without Maculopathy*. Current Eye Research, 2018. **43**(1): p. 96-101.
4. Budiningsari, D., et al., *Evaluation of Pictorial Dietary Assessment Tool for Hospitalized Patients with Diabetes: Cost, Accuracy, and User Satisfaction Analysis*. Nutrients, 2017. **10**(1): p. 28.
5. Elrayah-Eliadarous, H.A., et al., *Economic and social impact of diabetes mellitus in a low-income country: A case-control study in Sudan*. Journal of Diabetes, 2017. **9**(12): p. 1082-1090.
6. Emara, K.M., R. Ahmed Diab, and M. Amr Hemida, *Tibio-calcaneal fusion by retrograde intramedullary nailing in charcot neuroarthropathy*. Foot, 2018. **34**: p. 6-10.
7. Gireesha, K.S. and M. Sumithra, *A prospective study to compare insulin and insulin analogs in type II diabetic patients*. Asian Journal of Pharmaceutical and Clinical Research, 2017. **10**(7): p. 356-359.
8. Icks, A., et al., *Agreement found between self-reported and health insurance data on physician visits comparing different recall lengths*. Journal of Clinical Epidemiology, 2017. **82**: p. 167-172.
9. Janssen, E.M., et al., *Education and patient preferences for treating type 2 diabetes: A stratified discrete-choice experiment*. Patient Preference and Adherence, 2017. **11**: p. 1729-1736.
10. Ji, L., et al., *Lipohypertrophy in China: Prevalence, Risk Factors, Insulin Consumption, and Clinical Impact*. Diabetes Technology and Therapeutics, 2017. **19**(1): p. 61-67.
11. Jiao, F., et al., *Health-related quality of life and health preference of Chinese patients with diabetes mellitus managed in primary care and secondary care setting: decrements associated with individual complication and number of complications*. Health & Quality of Life Outcomes, 2017. **15**(1): p. 13.
12. Klinge, M.W., et al., *Early assessment of cost-effectiveness of gastric electrical stimulation for diabetic nausea and vomiting*. Journal of Neurogastroenterology and Motility, 2017. **23**(4): p. 541-549.
13. Kundra, T.S., P. Kaur, and N. Manjunatha, *Prayer sign as a marker of increased ventilatory hours, length of intensive care unit and hospital stay in patients undergoing coronary artery bypass grafting surgery*. Annals of Cardiac Anaesthesia, 2017. **20**(1): p. 90-92.
14. Laelasari, E., R. Sauriasari, and A. Banun, *Cost-effectiveness analysis of insulin, sulfonylurea, and sulfonylurea- Metformin in type 2 diabetes mellitus*. Asian Journal of Pharmaceutical and Clinical Research, 2017. **10**(Special Issue October): p. 50-53.
15. Lee, M.H., et al., *Factors associated with duration of inpatient hospital stay for patients with diabetes mellitus admitted to a medical unit in a community public hospital*. Australian Journal of Primary Health, 2017. **23**(1): p. 23-30.
16. Lee, Y., et al., *Cognitive Impairment Mediates Workplace Impairment in Persons with Type 2 Diabetes Mellitus: Results From the Motivaction Study*. Canadian Journal of Diabetes., 2017.

17. Mahdavi, M., et al., *The relationship between context, structure, and processes with outcomes of 6 regional diabetes networks in Europe*. PLoS ONE, 2018. **13 (2) (no pagination)**(e0192599).
18. Mapa-Tassou, C., et al., *Use of medical services and medicines attributable to type 2 diabetes care in Yaounde, Cameroon: a cross-sectional study*. BMC health services research, 2017. **17(1)**: p. 262.
19. Markle-Reid, M., et al., *Community Program Improves Quality of Life and Self-Management in Older Adults with Diabetes Mellitus and Comorbidity*. Journal of the American Geriatrics Society, 2018. **66(2)**: p. 263-273.
20. Modarresnia, L., et al., *Restless Legs Syndrome in Iranian People With Type 2 Diabetes Mellitus: The Role in Quality of Life and Quality of Sleep*. Journal of Clinical Sleep Medicine, 2018. **14(2)**: p. 223-228.
21. Pan, C.W., et al., *Diabetic retinopathy and health-related quality of life among Chinese with known type 2 diabetes mellitus*. Quality of Life Research, 2018: p. 1-7.
22. Panisch, S., et al., *The impact of a disease management programme for type 2 diabetes on health-related quality of life: multilevel analysis of a cluster-randomised controlled trial*. Diabetology & metabolic syndrome, 2018. **10**: p. 28.
23. Riandini, T., et al., *Functional status mediates the association between peripheral neuropathy and health-related quality of life in individuals with diabetes*. Acta Diabetologica, 2018. **55(2)**: p. 155-164.
24. Sankar, V., et al., *Efficacy, safety and cost effectiveness of amitriptyline and pregabalin in patients with diabetic peripheral neuropathy*. Indian Journal of Pharmaceutical Sciences, 2017. **79(4)**: p. 646-650.
25. Skordis-Worrall, J., et al., *Addressing the double-burden of diabetes and tuberculosis: lessons from Kyrgyzstan*. Global Health, 2017. **13(1)**: p. 15.
26. Soleymanian, T., et al., *Clinical outcomes and quality of life in hemodialysis diabetic patients versus non-diabetics*. Journal of Nephropathology, 2017. **6(2)**: p. 81-89.
27. Sothornwit, J., et al., *Decreased health-related quality of life in patients with diabetic foot problems*. Diabetes, Metabolic Syndrome and Obesity Targets and Therapy, 2018. **11**: p. 35-43.
28. Ugurlar, M., et al., *Comparison of two different vacuum-assisted closure (VAC) treatments of multiple chronic diabetic foot wounds in the same extremity*. Journal of Foot & Ankle Surgery, 2017. **23(3)**: p. 173-178.
29. Xu, R.H., A.W.L. Cheung, and E.L.Y. Wong, *Examining the health-related quality of life using EQ-5D-5L in patients with four kinds of chronic diseases from specialist outpatient clinics in Hong Kong SAR, China*. Patient Preference and Adherence, 2017. **11**: p. 1565-1572.
30. Zhu, B., Y. Ma, and H. Zou, *Cost-utility analyses of cataract surgery in vision-threatening diabetic retinopathy*. Journal of Cataract and Refractive Surgery, 2017. **43(1)**: p. 95-101.

**Supplementary table 3 – Quality assessment of descriptive studies**

Author	Country - Population	Exposure groups	Pharmacoeconomic outcome	Random or pseudo-random selection of patients	Clear criteria for inclusion	Identification of confounding factors	Outcome assessment validity	Sufficient description of groups	Appropriate length of time for follow-up	Inclusion of outcomes of people who withdrew	Outcome measured in a reliable way	Appropriate statistical analysis	Score
Baharlooei, O.(1)	Iran – T2DM patients aged ≥ 60 years	Single group	Hospital length of stay	No	Yes	Yes	Unclear	N/A	N/A	N/A	Unclear	No	33%
Bradley, C.(2)	Netherlands, Spain, UK, Belgium, France, Germany, Greece, Italy, Turkey – T2DM patients aged ≥ 40 years	Single group	EQ-5D visual analog scale	No	Yes	Yes	Yes	N/A	N/A	N/A	Yes	Yes	83%
Flores, N.(3)	USA – T2DM patients who underwent a non-medical switch for anti-hyperglycemic agents	2 groups: -Below 65 years -65 years and over	Number of medical visits and out-of-pocket costs	No	Yes	No	No	Yes	N/A	N/A	Yes	Yes	57%
Iqbal, Q.(4)	Pakistan – T2DM patients without comorbidities	Single group	EQ-5D index and visual analog scale	Unclear	Yes	Yes	Yes	N/A	N/A	N/A	Yes	Yes	83%
Jose, JV.(5)	India – Diabetic patients with nephropathy	Single group	Direct medical cost	Unclear	No	Yes	Yes	N/A	Yes	Unclear	Yes	Yes	63%
Jose, JV.(5)	India – Diabetic patients with nephropathy	Single group	Indirect cost	Unclear	No	Yes	No	N/A	Yes	Unclear	Yes	Yes	50%
Kamradt, M.(6)	Germany – T2DM patients with ≥ 2 chronic comorbidities	Single group	EQ-5D index	Yes	Yes	No	Yes	N/A	N/A	N/A	Yes	Yes	83%
Katibeth, M.(7)	Iran – Diabetic patients aged ≥ 50 years	Single group	Use of cataract surgery	Yes	Yes	No	Yes	N/A	N/A	N/A	Yes	Yes	83%
Ke, L.(8)	China – Diabetic patients	Single group	Out-of-pocket expenses	Yes	No	Yes	No	N/A	N/A	N/A	Yes	Yes	67%

Author	Country - Population	Exposure groups	Pharmacoeconomic outcome	Random or pseudo-random selection of patients	Clear criteria for inclusion	Identification of confounding factors	Outcome assessment validity	Sufficient description of groups	Appropriate length of time for follow-up	Inclusion of outcomes of people who withdrew	Outcome measured in a reliable way	Appropriate statistical analysis	Score
Khatib, ST.(9)	Palestine – T2DM patients receiving hemodialysis	Single group	EQ-5D index and visual analog scale	No	Yes	Yes	Yes	N/A	N/A	N/A	Yes	Yes	83%
Koki, G.(10)	Cameroun – Diabetic patients receiving laser treatment for retinopathy	Single group	Direct medical and non-medical cost	No	Yes	No	No	N/A	N/A	N/A	Yes	Unclear	33%
Le, NTD.(11)	Vietnam – T2DM patients receiving insulin or oral antidiabetics	Single group	Direct medical and non-medical cost	Yes	Yes	Yes	No	N/A	N/A	N/A	Yes	Yes	83%
Le, NTD.(11)	Vietnam – T2DM patients receiving insulin or oral antidiabetics	Single group	Indirect cost	Yes	Yes	Yes	No	N/A	N/A	N/A	Yes	Yes	83%
Lin, YJ.(12)	Taiwan – T2DM patients	Single group	Standard gamble and visual analog scale	No	Yes	Yes	Yes	N/A	N/A	N/A	Yes	Yes	83%
Long, Q.(13)	China – T2DM patients	Single group	Out-of-pocket expenses	No	Yes	No	No	N/A	N/A	N/A	Yes	Yes	50%
Mangi, AA.(14)	Pakistan – Diabetic patients	Single group	Out-of-pocket expenses	Unclear	No	No	No	N/A	N/A	N/A	Unclear	Unclear	0%
McEwen, MM.(15)	Mexico – T2DM patients	Single group	Number of medical visits (Behavioral Risk Factor Surveillance System questionnaire)	No	Yes	Yes	Yes	N/A	N/A	N/A	Yes	Yes	83%
Mentock, SM.(16)	India – T2DM patients	Single group	Most recent contact with a health care provider	No	Yes	No	No	N/A	N/A	N/A	Yes	Yes	50%
Mwangi, N.(17)	Kenya – Diabetic patients	Single group	Uptake of retinal examination	Yes	Yes	Yes	No	N/A	N/A	N/A	Yes	Yes	83%

Author	Country - Population	Exposure groups	Pharmacoeconomic outcome	Random or pseudo-random selection of patients	Clear criteria for inclusion	Identification of confounding factors	Outcome assessment validity	Sufficient description of groups	Appropriate length of time for follow-up	Inclusion of outcomes of people who withdrew	Outcome measured in a reliable way	Appropriate statistical analysis	Score
Mwaura, LW.(18)	Kenya – Diabetic patients	Single group	Frequency of diabetes clinic attendance	Yes	No	No	No	N/A	N/A	N/A	Yes	Yes	50%
Ohashi, Y.(19)	Japan – Diabetic patients with non-severe hypoglycemic episodes receiving insulin	Single group	Contact with a health care provider	No	Yes	No	No	N/A	N/A	N/A	Yes	Yes	50%
Ohashi, Y.(19)	Japan – Diabetic patients with non-severe hypoglycemic episodes receiving insulin	Single group	Productivity (work and daily activities)	No	Yes	No	No	N/A	N/A	N/A	Yes	Yes	50%
Olesen, K.(20)	Denmark – T2DM workers	Single group	Sickness absence	No	Yes	Yes	No	N/A	N/A	N/A	Unclear	Yes	50%
Pfeiffer, KM.(21)	Italy, UK, USA – Diabetic patients treated with basal-bolus insulin	Single group	Number of medical visits	Yes	Yes	Yes	No	N/A	N/A	N/A	Yes	Yes	83%
Pfeiffer, KM.(21)	Italy, UK, USA – Diabetic patients treated with basal-bolus insulin	Single group	Work productivity	Yes	Yes	Yes	No	N/A	N/A	N/A	Yes	Yes	83%
Rowen, D.(22)	UK – Diabetic patients and participants from the general population	2 groups: -Diabetic patients -General population	Willingness-to-pay	No	No	No	Yes	N/A	N/A	N/A	Yes	Yes	50%
Rowen, D.(22)	UK – Diabetic patients and participants from the general population	2 groups: -Diabetic patients -General population	EQ-5D index	No	No	No	Yes	N/A	N/A	N/A	Yes	Unclear	33%

Author	Country - Population	Exposure groups	Pharmacoeconomic outcome	Random or pseudo-random selection of patients	Clear criteria for inclusion	Identification of confounding factors	Outcome assessment validity	Sufficient description of groups	Appropriate length of time for follow-up	Inclusion of outcomes of people who withdrew	Outcome measured in a reliable way	Appropriate statistical analysis	Score
Sangam, K.(23)	India – T2DM patients	6 groups: -Diabetic without complications -Diabetic with hypertension -Diabetic with cardiovascular disease -Diabetic with hypertension + cardiovascular disease -Diabetic with peripheral neuropathy -Diabetic with peripheral neuropathy + hypertension + cardiovascular disease	Direct medical and non-medical cost	Unclear	Yes	No	No	No	Yes	Unclear	Unclear	No	22%
Singh, R.(24)	India – Diabetic patients receiving insulin	2 groups: -Insulin pen device -Insulin disposable syringe	Cost of therapy	No	Yes	No	No	Yes	N/A	N/A	Yes	Yes	57%
Stefanickova, J.(25)	France, Germany, Greece, Portugal, Russia, Slovakia, Spain, UK – Diabetic patients with macular edema receiving ranibizumab	Single group	Number of medical visits and treatments	No	Yes	No	Yes	N/A	Yes	Yes	Yes	Yes	75%
Tamilselvan, T.(26)	India – T2DM patients receiving oral hypoglycemic therapy	14 groups: Hypoglycemic monotherapy or combination	Cost of therapy	Unclear	Yes	No	Unclear	No	Yes	Unclear	Yes	Yes	44%

Author	Country - Population	Exposure groups	Pharmacoeconomic outcome	Random or pseudo-random selection of patients	Clear criteria for inclusion	Identification of confounding factors	Outcome assessment validity	Sufficient description of groups	Appropriate length of time for follow-up	Inclusion of outcomes of people who withdrew	Outcome measured in a reliable way	Appropriate statistical analysis	Score
Tan, ML.(27)	Singapore – T2DM patients	Single group	Presenteeism	Unclear	Yes	No	No	N/A	Yes	Yes	Yes	Yes	63%
Ugwu, ET.(28)	Nigeria – T2DM patients	Single group	Use and cost of self-monitoring of blood glucose and hospitalization rate	No	Yes	No	No	N/A	N/A	N/A	Unclear	Yes	33%
Ur Rashid Nazir, S.(29)	Pakistan – T2DM patients	Single group	EQ-5D index and visual analog scale	Unclear	Yes	No	Yes	N/A	N/A	N/A	Unclear	Yes	50%
Vijayan, M.(30)	India – T2DM patients receiving laser therapy for retinopathy	6 groups: -Ranibizumab + pan retinal photo-coagulation laser -Bevacizumab + pan retinal photo-coagulation laser -Pan retinal photo-coagulation laser -Ranibizumab + focal laser -Becavizumab + focal laser -Focal laser	Direct medical cost	No	Yes	No	Unclear	No	Yes	Unclear	Unclear	Yes	33%



## References

1. Baharlooei, O., M. Alavi, and M. Adelmehraban, *Psychosocial factors predicting length of hospitalization in elderly individuals with diabetes in selected hospitals of Isfahan University of Medical Sciences, Isfahan, Iran, in 2015*. ARYA Atherosclerosis, 2017. **13**(3): p. 103-108.
2. Bradley, C., et al., *Predictors of Quality of Life and Other Patient-Reported Outcomes in the PANORAMA Multinational Study of People With Type 2 Diabetes*. Diabetes Care, 2018. **41**(2): p. 267-276.
3. Flores, N.M., et al., *Consequences of non-medical switch among patients with type 2 diabetes*. Current Medical Research and Opinion, 2018: p. 1-7.
4. Iqbal, Q., et al., *Profile and predictors of health related quality of life among type II diabetes mellitus patients in Quetta city, Pakistan*. Health and Quality of Life Outcomes, 2017. **15** (1) (no pagination)(142).
5. Jose, J.V., et al., *Pharmacoeconomic evaluation of diabetic nephropathic patients attending nephrology department in a tertiary care hospital*. Journal of Postgraduate Medicine, 2017. **63**(1): p. 24-28.
6. Kamradt, M., et al., *Health-related quality of life in primary care: Which aspects matter in multimorbid patients with type 2 diabetes mellitus in a community setting?* PLoS ONE, 2017. **12** (1) (no pagination)(e0170883).
7. Katibeh, M., et al., *Rapid Assessment of Avoidable Blindness and Diabetic Retinopathy in Gilan Province, Iran*. Ophthalmic Epidemiology, 2017. **24**(6): p. 381-387.
8. Ke, L., et al., *Assessment of diabetes care and the healthcare system in economically and transport underdeveloped rural mountain areas of western China: A cross-sectional survey*. Journal of Diabetes, 2017. **9**(5): p. 475-481.
9. Khatib, S.T., et al., *Quality of life in hemodialysis diabetic patients: A multicenter cross-sectional study from Palestine*. BMC Nephrology, 2018. **19** (1) (no pagination)(49).
10. Koki, G., et al., *[What is the cost of treatment of diabetic retinopathy by argon laser in Yaounde?]*. Journal Francais d Ophthalmologie, 2018. **41**(4): p. 357-362.
11. Le, N.T.D., L.D. Pham, and T.Q. Vo, *Type 2 diabetes in Vietnam: A cross-sectional, prevalence-based cost-of-illness study*. Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy, 2017. **10**: p. 363-374.
12. Lin, Y.J., et al., *Patient preferences for diabetes-related complications in Taiwan*. Current Medical Research and Opinion, 2018: p. 1-7.
13. Long, Q., et al., *Treatment of Type 2 diabetes mellitus in Chongqing of China: unaffordable care for the poor*. Diabetic Medicine, 2017. **34**(1): p. 120-126.
14. Mangi, A.A., et al., *Role of pharmacist to subsidize the substantial cost of diabetic patients spent on their disease management at tertiary care hospital of Hyderabad, Sindh, Pakistan*. Rawal Medical Journal, 2017. **42**(3): p. 400-403.
15. McEwen, M.M., et al., *A Modified Behavior Risk Factor Surveillance System to Assess Diabetes Self-management Behaviors and Diabetes Care in Monterrey Mexico: A Cross-sectional Study*. Frontiers in Public Health, 2017. **5**: p. 97.

16. Mentock, S.M., et al., *Treatment-seeking behavior and obstacles to treatment compliance in diabetic patients in Mangaluru, India*. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, 2017. **11**(Supplement 2): p. S617-S622.
17. Mwangi, N., et al., *Predictors of uptake of eye examination in people living with diabetes mellitus in three counties of Kenya*. *Tropical Medicine and Health*, 2017. **45** (1) (no pagination)(41).
18. Mwaura, L.W., S. Wandibba, and C.O. Olungah, *Effect of distance on access to health services among women with type 2 diabetes in a rural community in Kenya*. *African Journal of Diabetes Medicine*, 2017. **25**(1): p. 18-20.
19. Ohashi, Y., et al., *Diabetes management and daily functioning burden of non-severe hypoglycemia in Japanese people treated with insulin*. *Journal of Diabetes Investigation*, 2017. **8**(6): p. 776-782.
20. Olesen, K., et al., *Characteristics associated with non-disclosure of Type 2 diabetes at work*. *Diabetic Medicine*, 2017. **34**(8): p. 1116-1119.
21. Pfeiffer, K.M., et al., *Postprandial glucose and healthcare resource use: a cross-sectional survey of adults with diabetes treated with basal-bolus insulin*. *Journal of Medical Economics*, 2017: p. 1-8.
22. Rowen, D., et al., *Using a Discrete-Choice Experiment Involving Cost to Value a Classification System Measuring the Quality-of-Life Impact of Self-Management for Diabetes*. *Value in Health*, 2018. **21**(1): p. 69-77.
23. Sangam, K., et al., *Evaluation of pharmacoeconomic direct cost in diabetes patients*. *Asian Journal of Pharmaceutical and Clinical Research*, 2017. **10**(4): p. 38-40.
24. Singh, R., C. Samuel, and J.J. Jacob, *A comparison of insulin pen devices and disposable plastic syringes - Simplicity, safety, convenience and cost differences*. *European Endocrinology*, 2018. **14**(1): p. 47-51.
25. Stefanickova, J., et al., *A noninterventional study to monitor patients with diabetic macular oedema starting treatment with ranibizumab (POLARIS)*. *Acta Ophthalmologica.*, 2018.
26. Tamilselvan, T., et al., *Pharmacoeconomical evaluation of oral hypoglycemic agents for type-2 diabetes mellitus in a multispeciality hospital*. *International Journal of Pharmaceutical Sciences and Research*, 2017. **8**(5): p. 2243-2248.
27. Tan, M.L., et al., *Factors associated with diabetes-related distress over time among patients with T2DM in a tertiary hospital in Singapore*. *BMC Endocrine Disorders*, 2017. **17** (1) (no pagination)(36).
28. Ugwu, E.T., C.J.G. Orjioke, and E.E. Young, *Self monitoring of blood glucose among patients with type 2 diabetes mellitus in eastern Nigeria: Need for multi-strategic interventions*. *Current Diabetes Reviews*, 2018. **14**(2): p. 175-181.
29. Ur Rashid Nazir, S., et al., *Does adherence to the therapeutic regimen associate with Health Related Quality of life: Findings from an observational study of type 2 diabetes mellitus patients in Pakistan*. *Pakistan Journal of Pharmaceutical Sciences*, 2017. **30**(6): p. 2159-2165.
30. Vijayan, M., et al., *Study on quality of life assessment in diabetic retinopathy among patients with type 2 diabetic patients*. *Asian Journal of Pharmaceutical and Clinical Research*, 2017. **10**(7): p. 116-119.

## CHAPITRE 8 : ARTICLE 2 – ÉTUDES EN PRATIQUE RÉELLE : UNE BOÎTE À OUTILS POUR LA COLLECTE DE DONNÉES D'UTILITÉ

---

### **Real-world evidence: a practical toolbox for collecting health state utilities**

Lambert-Obry V, Lafrance J-P, Savoie M, Lachaine J

Journal of Comparative Effectiveness Research, 2022 Jan; 11(1): 57-64

#### **8.1 Abstract**

Health state utilities (HSU) data collected in real-world evidence (RWE) studies are at risk of bias. Although numerous guidance documents are available, practical advice to avoid bias in HSU studies is limited. Thus, the objective of this article was to develop a concise toolbox intended for investigators seeking to collect HSU in a real-world setting. The proposed toolbox builds on existing guidance, and provides practical steps to help investigators perform good quality research. The toolbox aims at increasing the credibility of HSU data for future reimbursement decision-making.

#### **8.2 Introduction**

Health technology assessment (HTA) is a multidisciplinary process that encompasses diverse dimensions such as clinical, economic, organizational, social, cultural and ethical considerations.[1] Cost-utility analyses (CUA), which present results in terms of cost per quality-adjusted life-year (QALY), can be used to inform decision-making.[2] Utility data are required to calculate QALYs. Health state utilities (HSU) describe the value of a health state on a scale where 1 represents full health, 0 represents dead, and negative values represent states worse than death.[2] Well-designed randomized controlled trials (RCTs) can be a reliable source for generating utility

data. However, RCTs often provide estimates by treatment arm rather than HSU, which refer to estimates for given economic model health states. Moreover, RCTs have a low external validity, and may not capture all health states relevant to economic modeling. Therefore, HSU may have to be collected in real-world evidence (RWE) studies (herein defined as studies that are not conventional RCTs[3]), which are at higher risk of bias than RCTs. Regulatory and reimbursement agencies have shown growing interest in RWE to demonstrate the effectiveness of new therapies. Threats to internal validity of RWE studies have led to the publication of several guidance documents providing good procedural practices for comparative effectiveness research. In the same way as clinical outcomes, the estimation of HSU data can also be biased. Several publications showed that pharmacoeconomic data collected in RWE studies are at risk of bias, which can directly impact the incremental cost-effectiveness ratio (ICER) and lead to the wrong conclusions on cost-effectiveness.[4-15] Furthermore, sensitivity analyses of economic evaluations often show that the ICER is sensitive not only to clinical inputs, but also to utility values.[16, 17] Considering the risk of bias and the potential significant impact on the ICER, RWE studies collecting HSU should also entail good procedural practices.

According to stakeholder interviews, one of the challenges when using RWE for coverage decisions is the lack of investigator expertise (at interpreting data and adjusting for bias).[18] Recommendations adapted for the collection of HSU in a real-life practice could help to estimate high quality inputs for reimbursement decision-making. To our knowledge, none of the existing guidelines focus on methodologies for generating HSU in real-life practice specifically. Thus, the objective of this article is to review current recommendations for HSU generation, and incorporate methodological standards into a single toolbox intended for investigators seeking to collect HSU in a real-world setting.

### **8.3 Methods**

To identify existing guidance on the collection of HSU, guidelines and methodology reports of countries with a long tradition of HTA[19] were screened: Australia,[20] Canada,[21] England and Wales,[22] France,[23] Germany,[24] the Netherlands,[25] Scotland,[26] Spain

(Catalonia)[27] and Sweden.[28] Guidelines published by international HTA organizations were also screened, including the European Network for Health Technology Assessment (EUnetHTA),[29] Health Technology Assessment International (HTAi),[30] International Health Economics Association (iHEA),[31] International Network of Agencies for Health Technology Assessment (INAHTA)[32] and International Society for Pharmacoeconomics and Outcomes Research (ISPOR).[33] The search was conducted up to April 2021. Using the keywords "utility" and "utilities", 58 potentially relevant publications were identified by titles and abstracts (when available). Full-text publications were then reviewed for eligibility, resolving any disagreements through discussion. Selection was limited to publications providing guidance on how to collect HSU. Following full-text screening, 49 publications did not specifically discuss HSU collection, and were therefore excluded for the following reasons: focus on economic modeling/CUA methodologies (n=29), focus on mapping (n=13), specific to a certain population (n=4) and specific to certain disease areas (n=3). Among the 9 remaining publications, 2 were duplicates (journal articles based on previously published Technical Support Document), leading to 7 original publications: 5 from the National Institute for Health and Care Excellence (NICE)[34-38] and 2 from ISPOR.[2, 17] Data extraction included key methodological elements affecting internal validity: definition of health state, choice of instrument, mode of administration, features of assessment (e.g. timing and frequency), choice of respondents, sampling, method of recruitment, response rate, missing data, study design and analytical techniques.

None of the seven publications provide a framework for investigators seeking to estimate HSU in a real-world setting (data extraction is summarized in the Supplementary Table 1). Three documents discuss the NICE reference case and alternative methods for measuring HSU, indicating preferred choices of instruments and respondents.[34, 35, 38] Two reports focus on secondary research, providing a systematic methodology to incorporate HSU obtained from the literature into economic models. These two reports provide a framework for the identification, review and synthesis of HSU, noting that original sources should be reviewed for data quality and relevance. Response rate and missing data are presented as part of the minimum quality checks.[2, 37] A review by the NICE Decision Support Unit shows that reporting standards relating to basic information on regression models used in decision analytic models are poor, and provides

recommendations for improvement.[36] The only publication providing an exhaustive framework for primary research on HSU is the ISPOR Task Force Report on the collection of HSU for economic models in clinical trials.[17] However, as the report mainly focuses on clinical trials, some aspects, like design and analytical considerations for RWE studies, are only briefly discussed. Overall, only a few methodological elements are extensively addressed: the health state definition for utility estimation should be representative of the economic model health state, the preferred method for HSU generation is generally a generic preference-based measure (PBM), and respondents should usually be patients themselves and be representative of the economic model population. Practical advice to avoid bias in HSU generation, particularly selection bias and confounding, is limited. Therefore, a second search of the HTA publications was conducted to identify good practices for observational research (regardless of type of outcomes) in order to complement recommendations for HSU collection. This was a necessary step to gather practical guidance on study conduct in a real-world setting. Seven publications providing recommendations to avoid bias were identified: 1 from NICE[39] and 6 from ISPOR.[40-45]

As there is no single guideline for investigators seeking guidance on methodological steps to estimate HSU in a real-world setting, recommendations for HSU generation and for observational research were combined together to develop a concise toolbox. The toolbox recommendations aim to help investigators design and conduct a study as well as avoid methodological flaws. The toolbox is organized into 4 sections, as follows: Section 1 discusses confounding, Section 2 and 3 address selection bias (study sample and missing data respectively) and Section 4 discusses information bias. Existing recommendations were summarized in each section of the toolbox, and were then synthesized into practical steps.

## **8.4 Results**

The toolbox summarizing the practical steps for HSU generation in a real-world setting is presented in Table 1.

**Table 1. Toolbox to generate health state utilities in a real-world setting**

**Confounding: Study design and analytical techniques**

- Justify the choice of study design
- Explain how confounding variables are identified
- Describe how confounding variables are measured and the risk of residual confounding
- Control for confounding by design or through analytical methods
- Assess basic statistical assumptions and conduct sensitivity analyses
- Explore the risk of unmeasured confounding

**Selection bias: Study sample**

- Determine sample to ensure representativeness and relevance
- Select participants by probability sampling or in a systematic manner
- Implement similar enrollment procedures at all study sites and for each exposure group
- Document participation rates to discuss risk of selection bias

**Selection bias: Missing data**

- Take steps to minimize missing data
- Conduct a descriptive analysis of missing data
- Implement a method to handle missing data

**Information bias: Exposure and outcome measurement**

- Justify the choice of instrument
- Implement standardized procedures for exposure and outcome measurement
- Select features of assessment to ensure efficiency and suitability

### 8.4.1 Confounding: Study design and analytical techniques

#### *Justify the choice of study design*

An important step is to choose between a cross-sectional design or a longitudinal design (e.g. before-after study, prospective study).[40] Tradeoffs of different study designs (e.g. speed, cost, quality, relevance) should be discussed to ensure the study provides useful evidence and is fit for purpose.[40, 45] A well-known limitation of longitudinal studies is non-random missing data.[17] Conversely, cross-sectional data may provide biased estimates due to underlying heterogeneity across patients.[17] Indeed, there is evidence that a cross-sectional design may overestimate the impact of a health state on utility.[17] However, it is often impractical to wait for patients to progress through different health states.[17] The nature of the condition and health state (e.g. acute/episodic vs. chronic disease, severity within a single health state, speed of progression) will influence the applicability of the different designs.

#### *Explain how confounding variables are identified and measured, and describe the risk of residual confounding*

At the design stage, potential confounding domains should be identified based on existing literature and/or expert opinion.[39, 42, 44] Adjusting for intermediate and collider variables is inappropriate and causes bias.[39, 44] The use of directed acyclic graphs (causal models that contain directed edges [arrows], linking nodes [variables], and their paths) can help to clarify the structural relationship among variables (confounder, mediator, collider).[43, 44] Naturally, to limit residual confounding, measures of the confounding variables should be valid and reliable.

#### *Control for confounding by design or through analytical methods*

Without the advantage of randomization, all observational study designs are vulnerable to time-invariant and/or time-varying confounding.[39-45] Study designs using patients as their own controls (i.e. individual exposure varies) protect against time-invariant confounding.[40, 41, 44]



Any studies without a comparison group are vulnerable to confounding by secular trends or natural history (historical bias).[40, 41, 45] Methods controlling for confounding most commonly include matching, restriction, stratification, multivariate regression analysis and propensity score analysis.[39, 40, 42-45] Methods for dealing with time-varying confounding (affected by prior exposure) include inverse probability weighting of marginal structural models and g-estimation.[40, 43, 44]

### ***Assess basic statistical assumptions and conduct sensitivity analyses***

Although regression analysis is common practice in observational studies, it can lead to biased estimates if improperly performed.[39, 44, 45] Certainly, the functional form of the dependent variable drives the choice of the regression model (along with number of measures and/or clustering).[39, 44, 45] However, it may not be an easy choice considering the unique nature of HSU (left-skewed, censored, ceiling effect).[17] One option is to transform HSU to the decrement scale (right-skewed), where the back transformation to the original scale does not impact standard errors.[17] Certainly, the original HSU distribution should be displayed using a histogram as a starting point for model selection.[36] The widely used ordinary least squares may not be the best option, and different model types might need to be tested.[36, 39, 44, 45] For longitudinal studies, a common choice is generalized estimating equations.[17, 44] As for independent variables, the choice of their inclusion into the model should be justified.[36, 39, 44] When possible, controlling for baseline utility is recommended.[17] It is also recommended to present the full regression model, not only the adjusted effect.[42, 44] Coefficients should be presented along with their uncertainty parameters (e.g. standard errors, confidence intervals).[36, 42] The statistical assumptions that underlie the regression analysis should be tested (e.g. functional form, normality, heteroscedasticity, multicollinearity, overlap).[36, 39, 42, 44, 45] Model performance and goodness of fit should be reported.[36, 42, 44, 45] Key assumptions should be tested in sensitivity analyses, such as distribution of covariates (e.g. creation of categories, cutoffs), functional form (e.g. assumption of non-linearity), and outliers (e.g. inclusion or exclusion).[36, 39, 42, 44, 45]

### *Explore the risk of unmeasured confounding*

Unmeasured confounding may remain after adjustment, mainly due to unmeasured known confounders and unknown confounders. High-dimensional propensity score, additional adjustment (e.g. two-stage sampling, external adjustment) and instrumental variable analysis are proposed approaches to address unmeasured confounding.[44] The use of negative controls (outcomes not impacted by intervention) helps explore the presence of unmeasured confounding.[40, 42, 43] To substitute qualitative discussions of unmeasured confounding, sensitivity analyses should be performed.[42, 44] Simple options include the array approach (explores how the observed association changes by varying the covariates distribution and the strength of association), and the rule-out approach (determines the strength of unmeasured confounding needed to fully explain away the observed association).[44]

## **8.4.2 Selection bias: Study sample**

### *Determine sample to ensure representativeness and relevance*

Except for special patient populations (e.g. young children, cognitively impaired), patients should rate their own health.[2, 17, 34, 35, 37, 38] At design, the choice of eligibility criteria (e.g. disease severity, comorbidities, therapies) and study sites (e.g. primary, secondary, tertiary care) will directly impact generalizability and relevance for economic modeling.[2, 17, 34-38] To ensure the representativeness of an economic model, sampling should allow reflecting individual variability (i.e. patient heterogeneity) and variability over time (e.g. disease progression) within each health state.[17] The method of recruitment will also affect the representativeness, where an online sample (e.g. social media, recruitment panels, patient associations) may differ (e.g. sociodemographic and health profiles) from a sample of patients recruited through clinical sites.[17] The need to access detailed medical information certainly influences the choice of sample.[17]

***Select participants systematically, implement standardized enrollment procedures, and document participation rates to discuss risk of selection bias***

For HSU generation, a significant threat to internal validity is self-selection bias (e.g. volunteer bias, nonresponse bias). Most often, healthier patients are more likely to participate, which may bias estimates for severe health states.[17] To limit the impact of self-selection bias, patients (exposed and unexposed) should be selected from the same population. Moreover, although information on non-participants is often limited, participation rates have to be recorded to document the risk of selection bias.[2, 17, 37, 42] To select the study sample, there are several sampling techniques available. Although probability sampling is the gold standard to ensure sample representativeness, convenience samples are common practice for HSU generation in a real-world setting. To minimize selection bias, systematic participant recruitment such as consecutive sampling should be used.[40]

#### **8.4.3 Selection bias: Missing data**

***Take steps to minimize missing data, and conduct a descriptive analysis of missing data***

Minimizing missing data by taking preventive steps (e.g. reminders, incentives, multiple attempts) is the best way to limit bias.[17, 40] The extent of missing data should be reported by outcome, exposure group and timepoint when applicable.[2, 17, 36, 37, 42, 44] For longitudinal studies, missing data pattern should be described (i.e. monotonic [i.e. no return after dropout] or non-monotonic [i.e. intermittent missing data]).[17] Possible reasons for missing data should be discussed or baseline predictors should be explored.[2, 17, 37, 42]

***Implement a method to handle missing data***

One should explicitly report how missing data are handled.[2, 17, 36, 37, 42, 44] There are several techniques to deal with missing data, including complete case analysis (CCA) (i.e. restricting analysis to patients without missing data). Drawbacks of CCA are reduced sample size

and biased estimates if data are not missing completely at random (MCAR).[44] Restricting to CCA when data are not MCAR will overestimate HSU, representing a healthier subsample of the target population.[17] The appropriate choice of the analytical approach to deal with missing data (e.g. inverse probability weighting, multiple imputation, Bayesian methods) depends on missing data pattern and mechanism assumption (i.e. MCAR, missing at random or missing not at random). Standard imputation such as last observation carried forward is not recommended for HSU not MCAR.[17]

#### **8.4.4 Information bias: Exposure and outcome measurement**

##### *Justify the choice of instrument*

The use of a validated instrument is recommended.[2, 17, 34, 35, 37, 38] Generic PBM are preferred over condition-specific PBM (and over direct valuation), unless there is good evidence that generic measures are not able to measure changes in health.[2, 17, 34, 35, 37, 38] When selecting a generic PBM, one should consider the psychometric assessment in the particular health condition and requirements from HTA agencies.[2, 17, 34, 35, 37, 38]

##### *Implement standardized procedures for exposure and outcome measurement, and select features of assessment to ensure efficiency and suitability*

The health state for utility estimation should be clearly defined and representative of the economic model health state.[2, 17, 34, 35, 37, 38] It is not uncommon that the exposure is self-reported (rather than confirmed with medical records), which may impact categorization into health states depending on the complexity of the definitions (e.g. when progression through health states is defined by a laboratory or radiological marker).[17] Although blinding is the optimal way to protect against information bias, patients are usually aware of their health condition at utility measurement. When applicable, one option is to schedule the timing of assessment before the patient is informed of disease evolution (e.g. at the beginning of a medical visit before receiving

progression results), where standardization of the timing would be essential if other measures (e.g. test results) can affect HSU.[17] Frequency of assessments should also be carefully considered and adapted to disease progression, changes, events, and duration.[17] One should anticipate the needs of a future economic model and optimize measurement accordingly (e.g. capture changes within and between health states, capture transient effects).[17] The recall period of a questionnaire should also be considered when planning assessments, particularly when health states fluctuate over a short period of time (e.g. acute events).[2, 17] Moreover, one should choose the most suitable mode of administration (e.g. paper questionnaire, electronic questionnaire, interview), as they offer different advantages and disadvantages, which may best suit different populations and conditions.[17] To minimize the risk of bias, data collection, including mode of administration, should be standardized.[17]

## **8.5 Discussion**

A review of guidelines published by HTA organizations was conducted to identify standards for HSU generation. Existing recommendations mainly discuss the choice of instruments and respondents. Practical advice to design and conduct a RWE study collecting HSU is limited. Currently, only guidance documents for observational research provide methodological steps to increase internal validity, which are adapted for comparative effectiveness research. Without specific recommendations, RWE studies may generate poor quality HSU data, which can affect the accuracy of assessment of cost-effectiveness. The recent ISPOR report on the identification, review and use of HSU in CUA highlights that the possible bias introduced into CUA may occur at different stages: searching, selecting, synthesizing and use in the model.[2] Considering the numerous steps that can lead to bias, a good starting point to ensure the validity of CUA is the quality of original sources (i.e. studies generating HSU estimates). Although it is recognized that the quality of HSU inputs can ultimately impact reimbursement decision-making, none of the HTA organizations have published recommendations for investigators collecting HSU in a real-world setting. Existing guidelines focus either on HSU generation or on RWE, but a single framework is lacking. Therefore, the purpose was to develop a toolbox combining the different steps to help investigators avoid bias. The toolbox recommendations build on existing guidance, and address

important aspects such as confounding, selection bias and information bias. A limitation of the toolbox is that it may not cover all potential issues arising from HSU studies, as it is designed to be concise and practical for common investigators. Moreover, the toolbox is neither a protocol checklist nor a critical appraisal tool. The toolbox focuses on a narrow scope, and does not cover important topics, such as the conduct of economic evaluations, modeling techniques and decision-making processes. For in-depth discussion on a broad range of issues in HTA, further reading is available.[33] Although toolbox recommendations are supported by authoritative references, the next step would be that an international organization as ISPOR publishes an official framework.

## **8.6 Conclusion**

The proposed toolbox provides practical steps to help investigators perform good quality research. It is specifically designed for HSU generation in a real-world setting and is complementary to existing good practices. The toolbox aims at increasing the credibility of HSU data for future reimbursement decision-making.

## **8.7 Future perspective**

In a resource-constrained healthcare system, CUA is essential to help decision-making. Although clinical inputs for CUA are generally collected in RCTs, HTA agencies/regulatory bodies have shown growing interest in real-world comparative effectiveness research. Methodological principles for RWE studies have gained prominence in guidelines, and are highly topical in the era of “big data”. Considering the opportunities created by RWE studies, a rising uptake is to be expected. Pitfalls of RWE studies are well-known, yet guidelines for pharmacoeconomics outcomes as HSU provide limited advice. Practical recommendations for investigators would help generate high quality estimates for CUA. Standardization of RWE studies collecting HSU has yet to come.

## 8.8 Executive summary

- Health state utilities (HSU) data collected in real-world evidence (RWE) studies are at risk of bias. This can directly impact the incremental cost-effectiveness ratio (ICER), and lead to the wrong conclusions on cost-effectiveness.
- A review of the guidelines published by health technology assessment (HTA) organizations was conducted. Data extraction included key methodological elements affecting internal validity: definition of health state, choice of instrument, mode of administration, features of assessment, choice of respondents, sampling, method of recruitment, response rate, missing data, study design and analytical techniques.
- None of the existing guidelines focus on methodologies for HSU generation in real-life practice. Therefore, current recommendations from the different publications were combined together into a single toolbox.
- The toolbox was developed to help investigators design and conduct a study as well as limit confounding, selection bias and information bias.
- The toolbox aims at increasing the credibility of HSU data for future reimbursement decision-making.

## 8.9 References

1. O'rourke B, Oortwijn W, Schuller T, International Joint Task G. The new definition of health technology assessment: A milestone in international collaboration. *Int J Technol Assess Health Care* 36(3), 187-190 (2020).
2. Brazier J, Ara R, Azzabi I et al. Identification, Review, and Use of Health State Utilities in Cost-Effectiveness Models: An ISPOR Good Practices for Outcomes Research Task Force Report. *Value Health* 22(3), 267-275 (2019).  
  
\*\*Contains a checklist of minimum reporting standards of systematic review of utilities for cost-effectiveness models
3. Garrison LP, Jr., Neumann PJ, Erickson P, Marshall D, Mullins CD. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. *Value Health* 10(5), 326-335 (2007).
4. Kreif N, Grieve R, Sadique MZ. Statistical methods for cost-effectiveness analyses that use observational data: a critical appraisal tool and review of current practice. *Health Econ* 22(4), 486-500 (2013).
5. Guertin JR, Conombo B, Langevin R et al. A Systematic Review of Methods Used for Confounding Adjustment in Observational Economic Evaluations in Cardiology Conducted between 2013 and 2017. *Med Decis Making* 40(5), 582-595 (2020).
6. Handorf EA, Heitjan DF, Bekelman JE, Mitra N. Estimating cost-effectiveness from claims and registry data with measured and unmeasured confounders. *Stat Methods Med Res* 28(7), 2227-2242 (2019).
7. Lambert-Obry V, Lafrance JP, Savoie M, Henri S, Lachaine J. Review of real-world evidence studies in type 2 diabetes mellitus: Lack of good practices. *Int J Technol Assess Health Care* doi:10.1017/S0266462320000392 1-8 (2020).
8. Faries D, Peng X, Pawaskar M, Price K, Stamey JD, Seaman JW, Jr. Evaluating the impact of unmeasured confounding with internal validation data: an example cost evaluation in type 2 diabetes. *Value Health* 16(2), 259-266 (2013).
9. Frappier J, Tremblay G, Charny M, Cloutier LM. Costing bias in economic evaluations. *J Med Econ* 18(8), 596-599 (2015).
10. Fukuda H, Imanaka Y. Assessment of transparency of cost estimates in economic evaluations of patient safety programmes. *J Eval Clin Pract* 15(3), 451-459 (2009).
11. Harkanen T, Maljanen T, Lindfors O, Virtala E, Knekt P. Confounding and missing data in cost-effectiveness analysis: comparing different methods. *Health Econ Rev* 3(1), 8 (2013).



12. Guertin JR, Bowen JM, De Rose G, O'reilly DJ, Tarride JE. Illustration of the Impact of Unmeasured Confounding Within an Economic Evaluation Based on Nonrandomized Data. *MDM Policy Pract* 2(1), 2381468317697711 (2017).
  13. Clement Nee Shrive FM, Ghali WA, Donaldson C, Manns BJ. The impact of using different costing methods on the results of an economic evaluation of cardiac care: microcosting vs gross-costing approaches. *Health Econ* 18(4), 377-388 (2009).
  14. Evans CJ, Crawford B. Data collection methods in prospective economic evaluations: how accurate are the results? *Value Health* 3(4), 277-286 (2000).
  15. Franklin M, Lomas J, Walker S, Young T. An Educational Review About Using Cost Data for the Purpose of Cost-Effectiveness Analysis. *Pharmacoeconomics* 37(5), 631-643 (2019).
  16. Ara R, Hill H, Lloyd A, Woods HB, Brazier J. Are Current Reporting Standards Used to Describe Health State Utilities in Cost-Effectiveness Models Satisfactory? *Value in Health* 23(3), 397-405 (2020).
  17. Wolowacz SE, Briggs A, Belozeroff V et al. Estimating Health-State Utility for Economic Models in Clinical Studies: An ISPOR Good Research Practices Task Force Report. *Value Health* 19(6), 704-719 (2016).
- \*\*Provides guidance on how to collect health state utilities in clinical trials
18. The Institute for Clinical and Economic Review (ICER) and The Office of Health Economics (OHE). *Real World Evidence for Coverage Decisions: Opportunities and Challenges* (2018) <https://icer-review.org/wp-content/uploads/2018/03/ICER-Real-World-Evidence-White-Paper-03282018.pdf>
  19. Rowen D, Azzabi Zouraq I, Chevrou-Severac H, Van Hout B. International Regulations and Recommendations for Utility Data for Health Technology Assessment. *Pharmacoeconomics* 35(Suppl 1), 11-19 (2017).
  20. Department of Health and Ageing, Australian Government, Pharmaceutical Benefits Advisory Committee (PBAC) <https://pbac.pbs.gov.au/>
  21. Canadian Agency for Drugs and Technologies in Health (CADTH) <http://www.cadth.ca/>
  22. National Institute for Health and Care Excellence (NICE) <http://www.nice.org.uk/>
  23. Haute Autorité de Santé (HAS) <http://has-sante.fr>
  24. German national institute for quality and efficiency in health care (IQWiG) <http://www.iqwig.de/index.736.en.html>
  25. National Health Care Institute (ZIN) <http://www.zorginstituutnederland.nl>

26. Scottish Medicines Consortium (SMC) <http://www.scottishmedicines.org.uk>
27. CatSalut Catalan Health Service <https://web.gencat.cat/en/temes/salut/>
28. The Dental and Pharmaceutical Benefits Agency (TLV) [www.tlv.se](http://www.tlv.se)
29. European Network for Health Technology Assessment (EUnetHTA). Methodology Guidelines <https://eunetha.eu/methodology-guidelines/>
30. Health Technology Assessment International (HTAi). HTAi Vortal <https://vortal.htai.org>
31. International Health Economics Association (iHEA). Health Preference Research <https://www.healtheconomics.org/page/HealthPreference>
32. International Network of Agencies for Health Technology Assessment (INAHTA). HTA Tools & Resources <https://www.inahta.org/hta-tools-resources/>
33. International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Good Practices for Outcomes Research <https://www.ispor.org/heor-resources/good-practices-for-outcomes-research>
34. Brazier J. Valuing health States for use in cost-effectiveness analysis. *Pharmacoeconomics* 26(9), 769-779 (2008).
35. Brazier JE, Rowen D. NICE DSU Technical Support Document 11: Alternatives to EQ-5D for generating health state utility values. The National Institute for Health and Care Excellence (2011) <http://www.nicedsu.org.uk>
36. Kearns B, Ara R, Wailoo AG. A Review of the Use of Statistical Regression Models to Inform Cost Effectiveness Analyses within the NICE Technology Appraisals Programme: Report by the Decision Support Unit. The National Institute for Health and Care Excellence (2012) [www.nicedsu.org.uk](http://www.nicedsu.org.uk)
37. Papaioannou D, Brazier JE, Paisley S. NICE DSU Technical Support Document 9: The identification, review and synthesis of health state utility values from the literature. The National Institute for Health and Care Excellence (2011) <http://www.nicedsu.org.uk>
38. Rowen D, Brazier JE, Wong R, Wailoo AG. Measuring and valuing health-related quality of life when sufficient EQ-5D data is not available. The National Institute for Health and Care Excellence (2020) <http://nicedsu.org.uk>
39. Faria R, Alava MH, Manca A, Wailoo AG. NICE DSU Technical Support Document 17: The Use of Observational Data to Inform Estimates of Treatment Effectiveness in Technology Appraisal: Methods for Comparative Individual Patient Data. The National Institute for Health and Care Excellence (2015) <http://www.nicedsu.org.uk>

\*\*Provides guidance on how to conduct real-world evidence studies for comparative effectiveness research

40. Berger ML, Dreyer N, Anderson F, Towse A, Sedrakyan A, Normand SL. Prospective observational studies to assess comparative effectiveness: the ISPOR good research practices task force report. *Value Health* 15(2), 217-230 (2012).

\*\*Provides guidance on how to conduct prospective comparative effectiveness studies

41. Berger ML, Mamdani M, Atkins D, Johnson ML. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report--Part I. *Value Health* 12(8), 1044-1052 (2009).

42. Berger ML, Martin BC, Husereau D et al. A questionnaire to assess the relevance and credibility of observational studies to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. *Value Health* 17(2), 143-156 (2014).

\*\*Contains a 33-item questionnaire to assess the relevance and credibility of comparative effectiveness studies

43. Cox E, Martin BC, Van Staa T, Garbe E, Siebert U, Johnson ML. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report--Part II. *Value Health* 12(8), 1053-1061 (2009).

44. Johnson ML, Crown W, Martin BC, Dormuth CR, Siebert U. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report--Part III. *Value Health* 12(8), 1062-1073 (2009).

\*\*Discusses analytical methods to avoid bias in retrospective comparative effectiveness studies

45. Motheral B, Brooks J, Clark MA et al. A checklist for retrospective database studies--report of the ISPOR Task Force on Retrospective Databases. *Value Health* 6(2), 90-97 (2003).

\*\*Contains a 27-item questionnaire to assess the relevance and validity of retrospective database studies

**Supplementary Table 1.** Recommendations published by health technology assessment bodies for collecting health state utilities

Publication	Aim	Confounding	Selection bias	Information bias
Brazier, 2008 NICE DSU Article[1]	Provide guidance on reference case and alternative methods for measuring HSU	-	Selection of respondents	Choice of instrument Definition of HS
Brazier, 2011 NICE DSU TSD[2]	Provide guidance on alternative methods for measuring HSU when EQ-5D is insufficient	-	Selection of respondents	Choice of instrument Definition of HS
Papaioannou, 2011 NICE DSU TSD[3] & Article[4]	Provide guidance to identify, review and synthesize HSU	-	Selection of respondents Response rate Missing data	Choice of instrument Definition of HS
Kearns, 2012 NICE DSU TSD[5] & Article[6]	Provide guidance on the reporting of statistical regression models in CUA	Analytical techniques	Missing data	-
Wolowacz, 2016 ISPOR Report[7]	Provide guidance on the collection of HSU in clinical studies for CUA	Study design Analytical techniques	Selection of respondents Method of recruitment Response rate Missing data	Choice of instrument Mode of administration Frequency of assessment Timing of assessment Definition of HS
Brazier, 2019 ISPOR Report[8]	Provide guidance to identify, review and use HSU in CUA	-	Selection of respondents Response rate Missing data	Choice of instrument Timing of assessment Definition of HS
Rowen, 2020 NICE DSU Report[9]	Provide guidance on alternative methods for measuring HSU when EQ-5D is insufficient	-	Selection of respondents	Choice of instrument Definition of HS

CUA: Cost-utility analysis; DSU: Decision Support Unit; HS: Health state; HSU: Health state utility; ISPOR: International Society for Pharmacoeconomics and Outcomes Research; NICE: National Institute for Health and Care Excellence; TSD: Technical Support Document

1. Brazier J. Valuing health States for use in cost-effectiveness analysis. *Pharmacoeconomics* 26(9), 769-779 (2008).
2. Brazier JE, Rowen D. NICE DSU Technical Support Document 11: Alternatives to EQ-5D for generating health state utility values. The National Institute for Health and Care Excellence (2011) <http://www.nicedsu.org.uk>
3. Papaioannou D, Brazier JE, Paisley S. NICE DSU Technical Support Document 9: The identification, review and synthesis of health state utility values from the literature. The National Institute for Health and Care Excellence (2011) <http://www.nicedsu.org.uk>
4. Papaioannou D, Brazier J, Paisley S. Systematic searching and selection of health state utility values from the literature. *Value Health* 16(4), 686-695 (2013).
5. Kearns B, Ara R, Wailoo AG. A Review of the Use of Statistical Regression Models to Inform Cost Effectiveness Analyses within the NICE Technology Appraisals Programme: Report by the Decision Support Unit. The National Institute for Health and Care Excellence (2012) [www.nicedsu.org.uk](http://www.nicedsu.org.uk)
6. Kearns B, Ara R, Wailoo A *et al.* Good practice guidelines for the use of statistical regression models in economic evaluations. *Pharmacoeconomics* 31(8), 643-652 (2013).
7. Wolowacz SE, Briggs A, Belozeroff V *et al.* Estimating Health-State Utility for Economic Models in Clinical Studies: An ISPOR Good Research Practices Task Force Report. *Value Health* 19(6), 704-719 (2016).
8. Brazier J, Ara R, Azzabi I *et al.* Identification, Review, and Use of Health State Utilities in Cost-Effectiveness Models: An ISPOR Good Practices for Outcomes Research Task Force Report. *Value Health* 22(3), 267-275 (2019).
9. Rowen D, Brazier JE, Wong R, Wailoo AG. Measuring and valuing health-related quality of life when sufficient EQ-5D data is not available. The National Institute for Health and Care Excellence (2020) <http://nicedsu.org.uk>

## CHAPITRE 9 : ARTICLE 3 – IMPACT DE L’HYPOGLYCÉMIE SUR LA PERTE DE PRODUCTIVITÉ ET L’UTILITÉ DES PATIENTS ATTEINTS DU DIABÈTE DE TYPE 2 ET TRAITÉS À L’INSULINE EN PRATIQUE RÉELLE CANADIENNE : UN PROTOCOLE D’ÉTUDE PROSPECTIVE

---

### **The Impact of Hypoglycemia on Productivity Loss and Utility in Type 2 Diabetes Patients Treated with Insulin in Real-World Canadian Practice: Protocol for a Prospective Study**

Lambert-Obry V, Lafrance J-P, Savoie M, Lachaine J

JMIR Research Protocols 2022 Mar 28; 11(3): e35461

#### **9.1 Abstract**

**Background:** Type 2 diabetes mellitus (T2DM) imposes a significant burden, with its increasing prevalence and life-threatening complications. In patients not achieving glycemic targets on oral antidiabetic drugs, initiation of insulin is recommended. However, a serious concern about insulin is drug-induced hypoglycemia. Hypoglycemia is known to affect quality of life and healthcare resource utilization. However, health economics and outcomes research (HEOR) data for economic modeling are limited, particularly in terms of utility values and productivity losses.

**Objective:** The aim of this real-world prospective study is to assess the impact of hypoglycemia on productivity and utility in insulin-treated T2DM adults from Ontario and Quebec (Canada).

**Methods:** This noninterventional, multicenter, 3-month prospective study will recruit patients from four medical clinics and two endocrinology/diabetes clinics. Patients will be identified using appointment lists, and enrolled through consecutive sampling during routinely scheduled consultations. To be eligible, patients must be  $\geq 18$  years of age, diagnosed with T2DM, and treated with insulin. Utility and productivity will be collected using the EQ-5D-5L questionnaire and the iMTA Productivity Cost Questionnaire (*iPCQ*), respectively. Questionnaires will be completed at 4, 8 and 12 weeks after recruitment. Generalized estimating equations (GEE) models will be used to investigate productivity losses and utility decrements associated with incident hypoglycemic events while controlling for individual patient characteristics. A total of 500 patients will be enrolled to ensure precision of HEOR estimates.

**Results:** This study is designed to fill a gap in the Canadian evidence on the impact of hypoglycemia on HEOR outcomes. More specifically, it will generate productivity and utility inputs for economic modeling in T2DM.

**Conclusions:** Insulin therapies are expensive, and hypoglycemia is a significant component of economic evaluations. Robust HEOR data may help health technology assessment (HTA) agencies in future reimbursement decision making.

## 9.2 Introduction

Diabetes imposes a significant burden, with its increasing prevalence and life-threatening complications. According to Diabetes Canada, one in three Canadians is living with diabetes or prediabetes, among which type 2 diabetes mellitus (T2DM) accounts for 90% to 95%. [1] Glycemic control is key to the management of diabetes, and there are several types of oral therapies approved for T2DM. In patients not achieving glycemic targets on oral antidiabetic drugs, initiation of insulin is recommended. [1] However, a serious concern about insulin is drug-induced hypoglycemia. [2] Hypoglycemic events are a major obstacle to the achievement of glycemic

targets, and represent a challenge for both patients and treating physicians. The severity of hypoglycemia is defined by clinical manifestations, from mild symptoms to seizure and coma.[2]

### 9.2.1 Current state of knowledge

Hypoglycemia is known to affect quality of life and healthcare resource utilization. However, health economics and outcomes research (HEOR) data for economic modeling are limited, particularly in terms of utility values and productivity losses. O'Reilly *et al.* recently published data on healthcare costs associated with hypoglycemia based on the Canadian cohort of the international Hypoglycemia Assessment Tool (HAT) study.[3] Although providing relevant data on direct costs, productivity losses were subject to recall bias (1-year recall) and were limited to absenteeism (excluding presenteeism and unpaid work). As for utility values, a highly-cited literature review by Beaudet *et al.* identified utilities for 20 modeling complications associated with T2DM.[4] For hypoglycemia, the review found only one study with utility estimates suitable for economic modeling, i.e. taking into account the number of events (rather than presence vs. absence of hypoglycemia). Economic modeling in T2DM is particularly challenging since multiple interrelated organ systems are involved, over a long time horizon, with numerous risk factors. For an adverse event like hypoglycemia, HEOR data should depend on the number of episodes to fit T2DM economic models, as in Currie *et al.*, the current preferred source for utility values for hypoglycemia.[5] Although Currie *et al.* is the source recommended by several health technology assessment (HTA) agencies, it is criticized by the National Institute for Health and Care Excellence (NICE) for being subject to selection bias and recall bias.[6,7] Worthy of note, Beaudet *et al.* also highlighted that variability around the utility estimates (e.g. standard errors) are not always reported, thereby limiting future sensitivity analyses in cost-utility analyses (CUA).

Canadian data on productivity losses due to hypoglycemia include a 2005 survey on 133 T2DM patients of whom up to 9% and 26% missed work or study following non-severe hypoglycemic events (NSHE) and severe hypoglycemic events (SHE) respectively.[8] A web-based survey included 150 T2DM patients having a nocturnal NSHE in the last month. Among the



87 working respondents, 15 (17%) reported on average 3.5h of lost work.[9] In the Canadian HAT cohort, 6.0%, 3.8% and 7.5% of T2DM workers reported on average 2.9 days taken off, 3.8 days arriving late and 1.7 days leaving early respectively, based on a 1-year recall.[3] As for utility, time trade-off (TTO) values were elicited from 51 Canadian respondents with diabetes and 79 respondents from the general population. The mean utility ranged between 0.85-0.94, 0.77-0.90 and 0.66-0.83 for rare hypoglycemic events (quarterly), intermediate (monthly) and frequent (weekly) health states, respectively. In a multivariate linear regression, the estimated utilities associated with a single hypoglycemic event (all types combined) were -0.0033 and -0.0032 for respondents with diabetes and the general population, respectively.[10] Another TTO study estimated utility decrements of 0.006, 0.008, 0.059 and 0.062 for daytime NSHE, nocturnal NSHE, daytime SHE and nocturnal NSHE, respectively.[11] A third TTO study including diabetes patients and respondents from the general population estimated utility decrements ranging from 0.0028-0.0056, 0.0076, 0.0592-0.0726, 0.0616-0.0826 for daytime NSHE, nocturnal NSHE, daytime SHE and nocturnal SHE, respectively.[12] No Canadian studies meeting the Canadian Agency for Drugs and Technologies in Health (CADTH) recommendation to use an indirect method for utility measurement (e.g. EQ-5D) were identified.

## 9.2.2 Study rationale and relevance

The population at risk of drug-induced hypoglycemia is significant, with a proportion of T2DM patients requiring insulin around 20%-30%.[13,14] In the last five years, CADTH reviewed nine economic evaluations in T2DM, among which three were CUA for insulin therapies.[15-17] Considering the growing interest for HEOR data in insulin-treated T2DM patients, there is a need for robust estimates. Data on productivity losses are lacking, which are particularly meaningful in Quebec where the *Institut national d'excellence en santé et en services sociaux* preconizes economic evaluations adopting a societal perspective. In addition, in the three recent pharmacoeconomic reports on insulin therapies for T2DM, CADTH highlighted the uncertainty around utility values for hypoglycemia.[15-17] The committee described the current evidence on the impact of hypoglycemia on utility as limited and of low quality. In CADTH's reanalyses for those three CUA, variation in utility values for hypoglycemia led to a significant change in the

estimates of quality-adjusted life years gains. The incremental cost-effectiveness ratios were very sensitive to any changes in utility decrements for hypoglycemic events, thereby determined as an important driver of the results.[15-17] Thus, there is a need for HEOR data that are robust enough to make an informed decision with reasonable uncertainty. The aim of this real-world prospective study is to assess the impact of hypoglycemia on productivity and utility in insulin-treated T2DM adults from Ontario and Quebec.

## **9.3 Methods**

### **9.3.1 Research purpose and study design**

This noninterventional, multicenter, 3-month prospective study is designed to collect HEOR inputs for future economic modeling. The study will generate descriptive HEOR estimates that can be incorporated into T2DM models, along with precision parameters (i.e. confidence intervals and standard errors) to evaluate uncertainty in CUA sensitivity analyses. Considering the nature of the disease, a longitudinal design with repeated measures was deemed appropriate to limit confounding and increase the quality of evidence.

### **9.3.2 Study population**

In the Canadian clinical practice, T2DM patients on insulin can be followed by a family physician or a specialist.[18] Therefore, patients will be recruited from four medical clinics and two endocrinology/diabetes clinics. To increase generalizability, patients will be recruited from the two largest Canadian provinces, Ontario and Quebec. Patients will be identified using appointment lists, and enrolled through consecutive sampling during routinely scheduled consultations. To be eligible, patients must be  $\geq 18$  years of age, diagnosed with T2DM, and treated with insulin. There is no restriction on insulin regimens (e.g. short-acting, long-acting, mixed insulin). Given the lack of reimbursement, the use of insulin pumps is limited in T2DM.[19] Patients must also be able to understand and read English or French, and able to provide informed

consent. Patients will be excluded if they participate in a clinical trial. A screening log will be used to document participation rates.

### **9.3.3 Definition of hypoglycemia**

According to the Canadian Diabetes Association (CDA) and the American Diabetes Association (ADA), hypoglycemia is defined as plasma glucose concentrations of  $\leq 3.9$  mmol/L ( $\leq 70$  mg/dL).[2,20] In order to reflect a real-world setting, the definition of hypoglycemia in this study is not restricted to documented episodes, but includes any hypoglycemic event as judged by the patient. Symptoms defining hypoglycemia include trembling, palpitations, sweating, anxiety, hunger, nausea, tingling, difficulty concentrating, confusion, weakness, drowsiness, impaired vision, difficulty speaking, headache and/or dizziness. Patients will self-report hypoglycemia episodes recorded either by symptoms or blood glucose testing alone, or in a combination. This approach was also used in several large real-world studies,[18,19,21-24] and better represents real-life practice where patients do not always test their blood glucose for different reasons (e.g. forgetting, neglecting, lack of knowledge, lack of testing materials). Hypoglycemic events will be categorized as either severe or non-severe. A SHE is defined as an event requiring assistance of another person (to administer carbohydrates, glucagon, or take other corrective actions), whereas a NSHE can be managed by the patient alone, as per CDA and ADA definitions.[2,20]

Along with the severity of hypoglycemic events, frequency of events promotes fear of future hypoglycemia.[2,20] The fear itself is recognized to affect patients' quality of life.[2,20] In addition to the physical symptoms due to hypoglycemic events, the negative emotional impact of the fear of hypoglycemia is also a concern. For modeling purposes, HTA agencies recommend that the impact of hypoglycemia on utility considers both severity and frequency of episodes. They also recommend not to apply a utility decrement separately for the fear of hypoglycemia. Therefore, the current study was designed to capture the overall impact of hypoglycemia on utility, including the transient effect of events and the resulting fear.

### 9.3.4 Measurement of outcomes

#### Utility

The EQ-5D-5L questionnaire will be used to collect utility values.[25,26] This validated instrument is widely used and covers five dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression), which are themselves divided into five levels (no problems, slight problems, moderate problems, severe problems and extreme problems). The EQ-5D health states may be converted through a crosswalk value set into a single summary index. An EQ-5D summary index of 1 represents full health, 0 represents dead, and negative values represent states worse than death. The questionnaire also includes a visual analog scale (VAS) that records the respondent's self-rated health using a vertical scale ranging from 0 to 100 with endpoints labeled as 'the best health you can imagine' and 'the worst health you can imagine'. EQ-5D scores represent the patient's health status on the day of questionnaire completion.

#### Productivity loss

The institute for Medical Technology Assessment (iMTA) of Erasmus University Rotterdam recently developed and validated a questionnaire to estimate productivity losses, referred to as the iMTA Productivity Cost Questionnaire (*iPCQ*).[27] The *iPCQ* is a generic questionnaire designed to determine the value of productivity losses for economic evaluations adopting a societal perspective. The questionnaire includes three modules: lost productivity at paid work due to absenteeism, lost productivity at paid work due to presenteeism (reduced productivity) and lost productivity at unpaid work. A recall period of 4 weeks is used for presenteeism and unpaid work. Absenteeism is divided into two categories: short-term absence (no longer than 4 weeks) and long-term absence (sick leave that started prior to the recall period).

### 9.3.5 Study procedures

#### Data collection

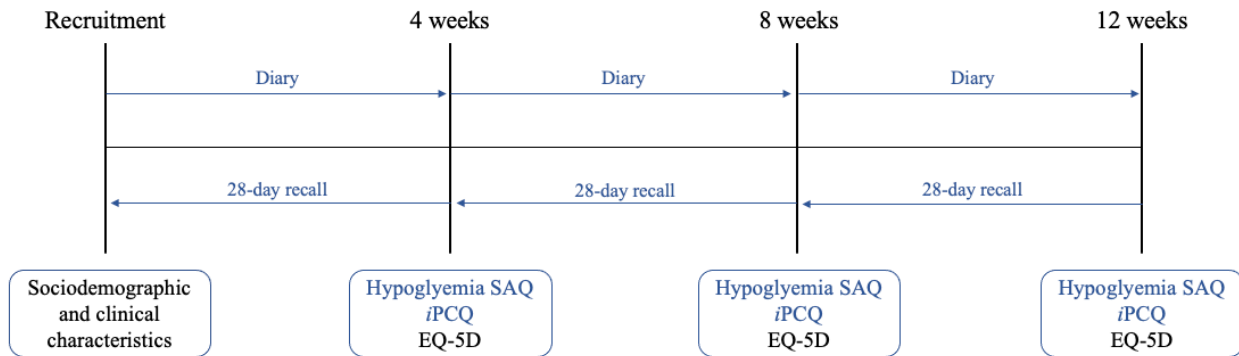
Sociodemographic and clinical characteristics of patients will be collected at recruitment. Patients will complete a baseline questionnaire to report their age, gender, ethnicity, education, income level, employment status, living status, alcohol consumption, smoking and physical activity. As a proxy for frailty, the self-reported number of visits to a healthcare provider over the last three months, and the number of emergency visits and hospitalizations over the last six months will also be collected.[28] As a proxy for lifestyle habits, immunization status (pneumococcal, influenza, COVID-19) will also be self-reported. Self-care (disease management activities) will be assessed using the Diabetes Self-Management Questionnaire (DSMQ), a validated questionnaire covering behaviors related to glycemic control (diet, blood glucose monitoring, medication adherence, physical activity and contact with health care professionals).[29] The type of device used for glucose monitoring will also be recorded (e.g. traditional finger-prick monitor, flash glucose monitoring device). Of note, continuous glucose monitoring systems are not reimbursed in T2DM, thereby limiting their use. To describe history of hypoglycemia, patients will also be asked if they have experienced NSHE and SHE over the last three months and the last year, respectively. Duration of diabetes, glycated hemoglobin (HbA<sub>1c</sub>), body mass index, therapy, duration of insulin, and vascular complications will be extracted from patients' medical records.

#### Assessments and study calendar

Patients will complete a self-assessment questionnaire (SAQ) at 4 weeks, 8 weeks and 12 weeks to report the number of NSHE and SHE experienced in the last 28 days. Patients will also be provided with a diary to record hypoglycemic events prospectively. Both the SAQ and the diary will be used to estimate the number of hypoglycemic events. The EQ-5D and *i*PCQ questionnaires will also be completed at 4, 8 and 12 weeks. Study calendar is shown in **Figure**. Questionnaires will be completed electronically via an online platform. To ensure that patients complete their questionnaires in a timely manner, reminders will be sent to patients by email. Paper-based

questionnaires can be provided to participants who do not have access to or are less familiar with using the Internet (phone call reminders).

**Figure 1. Study calendar**



### 9.3.6 Ethical considerations

The study will be conducted in accordance with the Declaration of Helsinki.[30] An independent review board (IRB) will approve the protocol before study initiation, and each patient will give written informed consent. After receiving IRB approval, the protocol will be registered on ClinicalTrials.gov. All study documents, including validated versions of questionnaires, will be made available to participants in English or French according to their preferences. Patients will not incur any costs for volunteering to take part in this research study. Patients will receive a compensation of \$20 for their participation at each completion of the set of questionnaires.

### 9.3.7 Statistical analyses

Data will be screened for accuracy, and questionnaires completed in paper format (if any) will entail double data entry to minimize transcription errors. Baseline characteristics will be presented for all patients participating in the study. Categorical variables will be summarized in frequencies (number and proportion), and descriptive statistics (mean and standard deviation) will be provided for continuous variables. Hypoglycemia will be measured in terms of hypoglycemic events and treated as a continuous variable, as recommended by HTA agencies for economic modeling in T2DM.[6,7] If hypoglycemic events are reported at a higher rate in the patient diary vs. the SAQ, diary values will be used to calculate the number of events, and SAQ values will be used in sensitivity analysis. For the EQ-5D-5L, questionnaire scores will be calculated according to the EuroQoL scoring manual[31], using the Canadian crosswalk value set.[32] For the *i*PCQ, productivity losses will be calculated for each module (absenteeism, presenteeism, unpaid work) using the iMTA scoring manual[33], and sum up to a total number.

#### Independent impact of hypoglycemia

Generalized estimating equations (GEE) models will be used to investigate productivity losses and utility decrements associated with incident hypoglycemic events while controlling for individual patient characteristics. For utility, two separate models will be presented: one for the EQ-5D index and one for the VAS. For productivity losses, one main analysis will be presented for the total number, and separate models for each module will be presented as complementary analysis. Potential model confounding covariates are presented in the Supplementary Table 1. Independent variables were identified based on existing literature as well as previous work from the Alliance for Canadian Health Outcomes Research in Diabetes.[34] Diabetes complications (micro- and macrovascular) to be included in the models are those preconized by HTA agencies for CUA in T2DM.[4,6,7] Based on the current reference study for hypoglycemia utility values,[5] the number of hypoglycemic episodes will be log-transformed to facilitate model fitting. Original scale will be tested in sensitivity analyses. Moreover, transformations of continuous independent variables will also be tested in sensitivity analyses (Supplementary Table 1). Presence of

multicollinearity will be tested, where important collinearity will be defined as a variance inflation factor  $>10$ . Due to the short duration of the study, all covariates, except hypoglycemia, will be considered as time-invariant for model simplicity. If vascular complications occur over the study period (incident cases), history of macrovascular and/or microvascular complications will be treated as a time-varying covariate. Main models will not include interaction terms, as large-scale studies evaluating the impact of diabetes complications on utility have not demonstrated any interaction between vascular event and event history, and between the different vascular events.[35-37] A sensitivity analysis will be performed to test for interaction between hypoglycemic event and determinants, namely history of hypoglycemia and diabetes complications.

Distributions of outcomes will be examined considering the expected nature of productivity data (right-skewed, excess zeros) and utility data (left-skewed, censored, ceiling effect). Although utility is commonly left-skewed, the distribution of utility scores may vary greatly between health conditions depending on the severity. Left-skewness is more common when the studied condition is associated with mild symptoms, whereas severe health states may even lead to negative values representing conditions worse than death.[38] One major advantage of the GEE is the avoidance of distributional assumptions, where several distributions can be tested. Several large-scale longitudinal studies evaluating the impact of diabetes complications on utility, including the landmark UKPDS study, used models fitted under the linear framework.[35-37] Common alternatives to the Gaussian distribution with identity link function include Gaussian distribution with log link, negative binomial distribution with log link, gamma distribution with identity link, and gamma distribution with log link. The gamma distribution requires non-zero positive continuous data, thus utility has to be transposed into disutility (1-utility value). A beta binomial distribution can also be used with a transformation (linear transformation or rescaling) to fit the restrictive open interval (0,1), which excludes the endpoints 0 and 1.[39] For healthcare resource utilization, non-linear options mostly include negative binomial distribution (for resource use), as well as gamma and inverse-Gaussian distribution with log link function (for costs). Raw outcomes data will be explored, where distributions will be depicted using histograms and normal probability plots. Selection of model will be based on predictive performance and goodness-of-fit, with the lowest values for the root mean square error (RMSE) and the mean absolute error (MAE), as well



as the highest values for pseudo- $R^2$ . Moreover, model performance will also be calculated using the quasi-likelihood under the independence model criterion (QIC), where lowest values are best. QIC and  $QIC_u$  will be used for correlation structure selection and variable selection, respectively. Because no distributional assumptions about the response variable are made, the regression parameters cannot be estimated using the method of maximum likelihood. Thus, quasi-likelihood statistics will be used (pseudo- $R^2$ , QIC,  $QIC_u$ ) instead of well-known likelihood statistics ( $R^2$ , Akaike information criterion, Bayesian information criterion).[40] Different variance-covariance structures are available to fit the relationship between responses. In health economics, common choices include exchangeable correlation structure (observations from the same cluster are equally correlated), autoregressive (correlation decreases with time), and unstructured (different and complex correlations). As per ISPOR Good Research Practices for Retrospective Database Analysis Task Force recommendations, if results are similar with different matrices, an exchangeable matrix will be used.[41] Regression diagnostics will be performed to explore the presence of influential observations and outliers (Cook's distance and residuals plots). Exclusion of extreme values will be tested in sensitivity analysis. All results will be expressed as coefficients, standard errors, 95% confidence intervals, and associated p-values (two-sided tests at the 0.05 significance level). Full regression models and adjusted effects will be presented. All statistical analyses will be performed with R.[42]

### Sample size

As per good research practices, sample sizes for utility estimates should be based on precision rather than hypothesis testing.[38,43] Indeed, there is no consensus on the use of minimal important differences for EQ-5D measures for statistical power calculations.[31,44] The sample size was therefore determined to ensure reasonable variability around the utility value for SHE (which also ensures precision for NSHE as the incidence is significantly higher). However, no value for the expected standard deviation was found in the literature to perform a reliable precision-based sample size calculation. A literature review was performed to identify studies estimating utility values for hypoglycemia. However, it is most often difficult to find an expected standard deviation for utility values as they are sensitive to the type of instrument, country, precise definition of outcome, timeframe, etc. Furthermore, most studies do not provide uncertainty values

around the estimates. As discussed in Beaudet *et al.*,[4] utility values for hypoglycemia are limited, and confidence intervals are scarce.

Therefore, the required sample size was determined based on the study by Currie *et al.*,[5] which is used as reference for hypoglycemia utility values by CADTH and NICE.[6,7] In Currie *et al.*, utility estimates were obtained from 68 patients who experienced at least one SHE over a 3-month period. According to Canadian real-world studies, the 3-month incidence of SHE among insulin-treated T2DM patients is approximately 17%.[18,19,45] With a 3-month incidence of 17%, 400 patients will have to be recruited to ensure that 68 patients experience at least one SHE during study follow-up. To account for attrition, the sample size will be increased by 20% for a target number of 500 patients to be enrolled. It is assumed that this sample size will also ensure precision of the productivity estimates.

### Missing data

Missing data will be defined according to outcome-specific guidelines, and a descriptive analysis will be conducted. According to the EQ-5D guidelines, situations that should be treated as missing data include not only unit non-response (no completion of questionnaire), but also item non-response (incomplete questionnaire) and ambiguous values (e.g. two boxes are ticked for a single dimension). When completing the *i*PCQ, some questions can be skipped if not applicable (e.g. when the patient is not employed). If the questions necessary to score a module are incomplete, the module is defined as missing.

If  $\leq 5\%$  of data is missing and there is no significant difference between completers and non-completers, missingness will be assumed to be completely at random (MCAR). Under the MCAR assumption, available case analysis performed with a GEE model yields valid estimates.[40,46-48] If  $>5\%$  of data is missing, the missing data pattern will be categorized as monotonic (i.e. no return after dropout) or non-monotonic (i.e. intermittent missing data).

Moreover, the mechanism by which data is missing will be investigated by examining which baseline covariates and previous measurements predicts missingness of a given outcome. Based on the results of the regression analysis, a specific variable will be determined as a predictor of missingness based on statistical significance and clinical plausibility.[49] In the presence of predictors of missingness, data will be considered as not MCAR, which may bias the GEE estimates. This would lead to the use of the weighted GEE (WGEE), a recently published extension of the GEE incorporating an inverse probability weight matrix.[50] The WGEE is a valid approach for dropout missingness (monotonic missing pattern), which is the typically observed pattern for longitudinal HEOR studies.[51] In addition, HEOR outcomes are most commonly missing as unit non-response (no completion of questionnaire) rather than item non-response (incomplete questionnaire).[49,51] Therefore, overall questionnaires scores (e.g. EQ-5D index score) will be used for missing data analysis. If the use of the WGEE is required, a sensitivity analysis will be conducted to compare the WGEE results to GEE estimates for the available case analysis.

### Unmeasured confounding

The presence of an unknown unmeasured confounder will be explored using E-values, a new standardized approach for sensitivity analysis.[52,53] For effect estimates, the E-value is the minimum strength of association on the risk ratio scale that an unmeasured confounder would need to have with both the exposure and the outcome, above and beyond the measured covariates, to fully explain away the observed association of exposure with the outcome. For the limit of the 95% CI, the E-value denotes the minimum strength of association on the risk ratio scale that an unmeasured confounder would need to have with both the exposure and the outcome, above and beyond the measured covariates, to shift the confidence interval to include the null value. This sensitivity analysis will assess robustness of the associations to unmeasured confounding.

## 9.4 Results

This study is designed to fill a gap in the Canadian evidence on the impact of hypoglycemia on HEOR outcomes. More specifically, it will generate productivity and utility inputs for economic modeling in T2DM. Insulin therapies are expensive, and hypoglycemia is a significant component of economic evaluations. Robust HEOR data may help HTA agencies in future reimbursement decision making.

## 9.5 Discussion

Insulin-induced hypoglycemia is a burden to diabetic patients, and this study will collect HEOR estimates reflecting how SHE and NSHE negatively affect patients' productivity and utility. To our knowledge, this would be the first Canadian real-world study to attempt to longitudinally measure the impact of hypoglycemia on utility and productivity loss, including absenteeism, presenteeism and unpaid work, in insulin-treated T2DM patients. In the present study, hypoglycemia will be categorized into SHE and NSHE, without further subgrouping by severity (e.g. mild, moderate) or time of occurrence (daytime, nocturnal). This classification is preconized by HTA agencies for CUA in T2DM.[6,7] The subjective measurement of hypoglycemia may overestimate the number of episodes, which may underestimate outcome values by event (conservative approach). In addition, the recently published guidelines by the International Hypoglycaemia Study Group recommend that the hypoglycemia threshold be lowered to  $<3.0$  mmol/L ( $<54$  mg/dL) for clinical trials.[54] This threshold was suggested because it is sufficiently low to indicate serious, clinically important hypoglycemia, and also because this level does not occur under physiological conditions in nondiabetic individuals. A joint position statement of the ADA and the European Association for the Study of Diabetes was subsequently published, which approves that the glucose level should be lowered for clinical trials.[54] The use of the official definition of hypoglycemia ( $\leq 3.9$  mmol/L) instead of a lower threshold might lead to overreporting of non-clinically serious events. Nevertheless, this approach is also conservative as it underestimates the outcome values by event. Defining hypoglycemia with symptoms and/or blood glucose measurements is considered a reliable method best reflecting real-world

practice.[18,19] Frequency of assessments (recall period) was also based on previous Canadian studies.[18,19] The use of diary records (if higher than the number reported in the questionnaire) can compensate for potential recall bias, and is a conservative approach as it would decrease the outcome values by event. Considering that longitudinal studies usually have a minimum of three measurements, and that the frequency of assessments is four weeks, the duration of the study is therefore three months.

Outcomes will be measured using validated questionnaires. Although several instruments are available to estimate productivity losses, there is no gold standard.[55] In addition, HTA agencies make different recommendations in terms of types of productivity losses (absenteeism, presenteeism, unpaid work) to include in economic evaluations, as well as methods for valuation/monetization (human capital approach, friction approach).[33] Therefore, productivity losses will be presented as total and type-specific raw scores (number of hours/days), and will not be transposed into monetary value. This approach gives flexibility, allowing future economic evaluations to be fit for purpose. The *i*PCQ scoring manual presents different valuation methods depending on various scenarios (e.g. presence of long-term absences, frequency of measurements), and is fully adaptable to the different HTA requirements.[33] It is recommended that the choice of a work productivity instrument for economic evaluations should be based on purpose, perspective, practicality, population and psychometrics (5‘Ps’).[55] The *i*PCQ meets the 5‘Ps’ criteria as it is a validated generic instrument, that allows monetization, captures all types of productivity losses, and that minimizes the burden to patients (short completion time, easy to understand). Due to its recent development, the use of the *i*PCQ is less extended than older questionnaires, which comes with the advantage that questions were optimized based on three previous validated instruments.[33] Moreover, this promising instrument has a recall period of 4 weeks which matches measurements of hypoglycemia, while limiting recall bias. Although there is no gold standard, the Work Productivity and Activity Impairment (WPAI) questionnaire is the most widely used instrument to assess productivity losses.[56] Like the *i*PCQ, the WPAI questionnaire covers absenteeism, presenteeism and unpaid work, but over a shorter period. Because the recall period is only seven days, it may not capture all productivity losses due to hypoglycemia if events occur outside of the recall period. As for utility, the use of generic preference-based instruments is

recommended, among which the EQ-5D is the most extensively used.[57] The EQ-5D asks patients how they feel today without any recall period. Therefore, for complications associated with transient effects only (e.g. diarrhea), utility should be measured simultaneously (i.e. on the same day). When acute events are followed by persistent fear/anxiety (e.g. stroke), different timing of measurement can be used. If utility is measured shortly after the event, the punctual effect of the complication will be captured. If utility is not concurrently measured, a general effect will be captured. For economic modeling in T2DM, utility values recommended by HTA agencies were all sourced from studies designed to capture the general impact of diabetic complications, including hypoglycemia.[6,7] In the reference study by Currie *et al.*, the EQ-5D captured decrements in utility due to hypoglycemic events that occurred over the last 3 months.[5] Furthermore, in several randomized controlled trials, the impact of SHE on utility was measured for up to 1 year after the event.[36,58,59] A limitation of this approach is the underestimation of the acute physical effects related to the event. Indeed, the current design may capture more accurately the psychological effects of hypoglycemia. Still, this study will estimate the overall impact of this treatment-related adverse effect on utility, providing relevant inputs for T2DM economic modeling.

Confounding is a concern in observational research.[28] Before implementing this study, independent variables were thoroughly identified, and each known variable will be collected to limit confounding. To reflect real-world practice, independent variables that will be extracted from patients' medical files will represent the last available value which may not reflect current unknown value. Moreover, some independent variables will be self-reported potentially leading to residual confounding. Nevertheless, there are no unmeasured known confounders, and the potential impact of an unmeasured unknown confounder will be tested with E-values.[52,53] There is evidence that a cross-sectional design may overestimate the impact of T2DM complications on utility due to underlying heterogeneity across patients.[37] Therefore, a longitudinal design with a time-varying exposure will help protect against time-invariant confounding (natural heterogeneity).[40,46-48] A GEE model will be used to account for the correlation associated with repeated measures from the same individual. Although mixed models provide a flexible framework compared to GEE, they require a large sample size and may be computationally demanding.

Therefore, the simpler GEE method will be used to deal with this non-complex data set [no large-scale data analysis, single level of clustering, and absence of non-stochastic time-varying covariates (e.g. time from baseline)]. In addition, it is acknowledged that a GEE is comparable to a random intercept model for continuous outcomes. A drawback of the GEE is the assumption that data is MCAR.[40,46-48] Therefore, if data are not MCAR, the WGEE will be used to ensure robustness of estimates.[50] Furthermore, reminders and incentives should help minimize the rate of missing data.

Patients will be recruited in several regions throughout Quebec and Ontario to increase the external validity of the results. However, recruitment sites will be limited to urban areas, and may not be representative of rural areas. The real-world design along with the broad eligibility criteria will ensure the HEOR estimates are generalizable to a target population for future reimbursement purposes. It is assumed that recruitment through clinical sites only will not affect the representativeness of the sample since diabetes patients have regular follow-ups with their healthcare providers thereby capturing most eligible patients and not only patients in worse condition. Furthermore, enrolling patients from both medical and diabetes clinics will help be representative of the target population. Although probability sampling is the gold standard to ensure sample representativeness, it is most often not feasible in Canada since many jurisdictions lack electronic patient databases, particularly family practice. Still, a systematic participant recruitment as consecutive sampling using appointments lists also helps minimize selection bias (including oversampling). Participation rates will be recorded to document the risk of selection bias. Of note, the COVID pandemic may affect patient's productivity and utility. However, given the use of the GEE to estimate the independent impact of hypoglycemia on the HEOR estimates, the results are expected to be valid and generalizable.

Robust evidence on productivity and utility associated with insulin-induced hypoglycemia is lacking in Canada. Currently available data on productivity loss were not estimated with a validated questionnaire, thereby increasing the risk of bias.[3,8,9] A systematic review published in 2021 identified forty-two unique instruments to measure productivity, and the authors

recommended the iPCQ for use in economic evaluations.[60] As for the current evidence on utility decrement due to hypoglycemia, Canadian data are limited to vignettes studies (i.e. bespoke descriptions of impaired health states), which are not the preferred source of utility due to their inherent drawbacks.[10-12] This study will generate high-quality HEOR estimates for future economic modeling in T2DM.

## 9.6 References

1. Diabetes Canada Clinical Practice Guidelines Expert, C., Lipscombe, L., Butalia, S., Dasgupta, K., Eurich, D. T., MacCallum, L., et al. (2020). Pharmacologic Glycemic Management of Type 2 Diabetes in Adults: 2020 Update. *Can J Diabetes*, 44(7), 575-591, doi:10.1016/j.jcjd.2020.08.001.
2. Diabetes Canada Clinical Practice Guidelines Expert, C., Yale, J. F., Paty, B., & Senior, P. A. (2018). Hypoglycemia. *Can J Diabetes*, 42 Suppl 1, S104-S108, doi:10.1016/j.jcjd.2017.10.010.
3. O'Reilly, D. J., Burke, N., Tarride, J. E., Hahn, J., & Nurkanovic, L. (2018). Direct Health-Care Costs and Productivity Costs Associated With Hypoglycemia in Adults With Type 1 and Type 2 Diabetes Mellitus That Participated in the Canadian Hypoglycemia Assessment Tool Program. *Can J Diabetes*, 42(6), 659-663, doi:10.1016/j.jcjd.2018.01.010.
4. Beaudet, A., Clegg, J., Thuresson, P. O., Lloyd, A., & McEwan, P. (2014). Review of utility values for economic modeling in type 2 diabetes. *Value Health*, 17(4), 462-470, doi:10.1016/j.jval.2014.03.003.
5. Currie, C. J., Morgan, C. L., Poole, C. D., Sharplin, P., Lammert, M., & McEwan, P. (2006). Multivariate models of health-related utility and the fear of hypoglycaemia in people with diabetes. *Curr Med Res Opin*, 22(8), 1523-1534, doi:10.1185/030079906X115757.
6. Canadian Agency for Drugs and Technologies in Health (CADTH) (2017). New Drugs for Type 2 Diabetes: Second-Line Therapy – Science Report. <https://www.cadth.ca/dv/new-drugs-type-2-diabetes-second-line-therapy-recommendations-report>.
7. National Institute for Health and Care Excellence (NICE) (2015). Type 2 diabetes in adults: management. Appendix F: Full Health Economics Report. [www.nice.org.uk/guidance/ng28](http://www.nice.org.uk/guidance/ng28).
8. Leiter, L. A., Yale, J., Chiasson, J., Harris, S., & Kleinstiver, P. Assessment of the Impact of Fear of Hypoglycemic Episodes on Glycemic and Hypoglycemia Management. In, 2005



9. Brod, M., Wolden, M., Groleau, D., & Bushnell, D. M. (2014). Understanding the economic, daily functioning, and diabetes management burden of non-severe nocturnal hypoglycemic events in Canada: differences between type 1 and type 2. *J Med Econ*, 17(1), 11-20, doi:10.3111/13696998.2013.857676.
10. Levy, A. R., Christensen, T. L., & Johnson, J. A. (2008). Utility values for symptomatic non-severe hypoglycaemia elicited from persons with and without diabetes in Canada and the United Kingdom. *Health Qual Life Outcomes*, 6, 73, doi:10.1186/1477-7525-6-73.
11. Evans, M., Khunti, K., Mamdani, M., Galbo-Jorgensen, C. B., Gundgaard, J., Bogelund, M., et al. (2013). Health-related quality of life associated with daytime and nocturnal hypoglycaemic events: a time trade-off survey in five countries. *Health Qual Life Outcomes*, 11, 90, doi:10.1186/1477-7525-11-90.
12. Harris, S., Mamdani, M., Galbo-Jorgensen, C. B., Bogelund, M., Gundgaard, J., & Groleau, D. (2014). The effect of hypoglycemia on health-related quality of life: Canadian results from a multinational time trade-off survey. *Can J Diabetes*, 38(1), 45-52, doi:10.1016/j.jcjd.2013.09.001.
13. Al Sayah, F., Yeung, R. O., & Johnson, J. A. (2019). Association of Depressive Symptoms and Diabetes Distress With Severe Hypoglycemia in Adults With Type 2 Diabetes. *Can J Diabetes*, 43(5), 316-321, doi:10.1016/j.jcjd.2018.11.002.
14. Basu, S., Yudkin, J. S., Kehlenbrink, S., Davies, J. I., Wild, S. H., Lipska, K. J., et al. (2019). Estimation of global insulin use for type 2 diabetes, 2018-30: a microsimulation analysis. *Lancet Diabetes Endocrinol*, 7(1), 25-33, doi:10.1016/s2213-8587(18)30303-6.
15. Canadian Agency for Drugs and Technologies in Health (CADTH) (2019). Pharmacoeconomic Review Report - Insulin degludec and liraglutide injection (Xultophy). <https://www.cadth.ca/sites/default/files/cdr/pharmacoeconomic/sr0599-xultophy-pharmacoeconomic-review-report.pdf>.
16. Canadian Agency for Drugs and Technologies in Health (CADTH) (2019). Pharmacoeconomic Review Report - Insulin glargine + lixisenatide (Soliqua). <https://www.cadth.ca/sites/default/files/cdr/pharmacoeconomic/sr0564-soliqua-pharmacoeconomic-report.pdf>.
17. Canadian Agency for Drugs and Technologies in Health (CADTH) (2017). Pharmacoeconomic Review Report - Insulin degludec (Tresiba). [https://www.cadth.ca/sites/default/files/cdr/pharmacoeconomic/SR0521\\_Tresiba\\_PE\\_Report.pdf](https://www.cadth.ca/sites/default/files/cdr/pharmacoeconomic/SR0521_Tresiba_PE_Report.pdf).
18. Ratzki-Leewing, A., Harris, S. B., Mequanint, S., Reichert, S. M., Belle Brown, J., Black, J. E., et al. (2018). Real-world crude incidence of hypoglycemia in adults with diabetes: Results of the InHypo-DM Study, Canada. *BMJ Open Diabetes Res Care*, 6(1), e000503, doi:10.1136/bmjdr-2017-000503.

19. Aronson, R., Goldenberg, R., Boras, D., Skovgaard, R., & Bajaj, H. (2018). The Canadian Hypoglycemia Assessment Tool Program: Insights Into Rates and Implications of Hypoglycemia From an Observational Study. *Can J Diabetes*, *42*(1), 11-17, doi:10.1016/j.jcjd.2017.01.007.
20. Seaquist, E. R., Anderson, J., Childs, B., Cryer, P., Dagogo-Jack, S., Fish, L., et al. (2013). Hypoglycemia and diabetes: a report of a workgroup of the American Diabetes Association and the Endocrine Society. *Diabetes Care*, *36*(5), 1384-1395, doi:10.2337/dc12-2480.
21. Bradley, C., Eschwege, E., de Pablos-Velasco, P., Parhofer, K. G., Simon, D., Vandenberghe, H., et al. (2018). Predictors of Quality of Life and Other Patient-Reported Outcomes in the PANORAMA Multinational Study of People With Type 2 Diabetes. *Diabetes Care*, *41*(2), 267-276, doi:10.2337/dc16-2655.
22. Davis, T. M. E., Bruce, D. G., Curtis, B. H., Barraclough, H., & Davis, W. A. (2018). The relationship between intensification of blood glucose-lowering therapies, health status and quality of life in type 2 diabetes: The Fremantle Diabetes Study Phase II. *Diabetes Res Clin Pract*, *142*, 294-302, doi:10.1016/j.diabres.2018.05.047.
23. Brod, M., Rana, A., & Barnett, A. H. (2012). Impact of self-treated hypoglycaemia in type 2 diabetes: a multinational survey in patients and physicians. *Curr Med Res Opin*, *28*(12), 1947-1958, doi:10.1185/03007995.2012.743457.
24. Ostenson, C. G., Geelhoed-Duijvestijn, P., Lahtela, J., Weitgasser, R., Markert Jensen, M., & Pedersen-Bjergaard, U. (2014). Self-reported non-severe hypoglycaemic events in Europe. *Diabet Med*, *31*(1), 92-101, doi:10.1111/dme.12261.
25. EuroQol, G. (1990). EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy*, *16*(3), 199-208, doi:10.1016/0168-8510(90)90421-9.
26. Janssen, M. F., Pickard, A. S., Golicki, D., Gudex, C., Niewada, M., Scalone, L., et al. (2013). Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Qual Life Res*, *22*(7), 1717-1727, doi:10.1007/s11136-012-0322-4.
27. Bouwmans, C., Krol, M., Severens, H., Koopmanschap, M., Brouwer, W., & Hakkaart-van Roijen, L. (2015). The iMTA Productivity Cost Questionnaire: A Standardized Instrument for Measuring and Valuing Health-Related Productivity Losses. *Value Health*, *18*(6), 753-758, doi:10.1016/j.jval.2015.05.009.
28. Patient-Centered Outcomes Research Institute (2019). Antihyperglycemic Therapy and Cardiovascular Risk: Design and Emulation of a Target Trial Using Healthcare Databases.
29. Schmitt, A., Reimer, A., Hermanns, N., Huber, J., Ehrmann, D., Schall, S., et al. (2016). Assessing Diabetes Self-Management with the Diabetes Self-Management Questionnaire (DSMQ) Can Help Analyse Behavioural Problems Related to Reduced Glycaemic Control. *PLoS One*, *11*(3), e0150774, doi:10.1371/journal.pone.0150774.

30. General Assembly of the World Medical, A. (2014). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *J Am Coll Dent*, 81(3), 14-18.
31. Devlin, N., Parkin, D., Janssen, B. (2020). Methods for Analysing and Reporting EQ-5D Data. <https://doi.org/10.1007/978-3-030-47622-9>. Accessed 30 October 2020.
32. Xie, F., Pullenayegum, E., Gaebel, K., Bansback, N., Bryan, S., Ohinmaa, A., et al. (2016). A Time Trade-off-derived Value Set of the EQ-5D-5L for Canada. *Med Care*, 54(1), 98-105, doi:10.1097/MLR.0000000000000447.
33. Institute for Medical Technology Assessment (iMTA) of Erasmus University Rotterdam (2020). Productivity Costs Questionnaire Manual. <https://www.imta.nl/questionnaires/>.
34. Al Sayah, F., Majumdar, S. R., Soprovich, A., Wozniak, L., Johnson, S. T., Qiu, W., et al. (2015). The Alberta's Caring for Diabetes (ABCD) Study: Rationale, Design and Baseline Characteristics of a Prospective Cohort of Adults with Type 2 Diabetes. *Can J Diabetes*, 39 Suppl 3, S113-119, doi:10.1016/j.jcjd.2015.05.005.
35. Hayes, A., Arima, H., Woodward, M., Chalmers, J., Poulter, N., Hamet, P., et al. (2016). Changes in Quality of Life Associated with Complications of Diabetes: Results from the ADVANCE Study. *Value Health*, 19(1), 36-41, doi:10.1016/j.jval.2015.10.010.
36. Shao, H., Yang, S., Fonseca, V., Stoecker, C., & Shi, L. (2019). Estimating Quality of Life Decrements Due to Diabetes Complications in the United States: The Health Utility Index (HUI) Diabetes Complication Equation. *Pharmacoeconomics*, 37(7), 921-929, doi:10.1007/s40273-019-00775-8.
37. Alva, M., Gray, A., Mihaylova, B., & Clarke, P. (2014). The effect of diabetes complications on health-related quality of life: the importance of longitudinal data to address patient heterogeneity. *Health Econ*, 23(4), 487-500, doi:10.1002/hec.2930.
38. Wolowacz, S. E., Briggs, A., Belozeroff, V., Clarke, P., Doward, L., Goeree, R., et al. (2016). Estimating Health-State Utility for Economic Models in Clinical Studies: An ISPOR Good Research Practices Task Force Report. *Value Health*, 19(6), 704-719, doi:10.1016/j.jval.2016.06.001.
39. Hunger, M., Doring, A., & Holle, R. (2012). Longitudinal beta regression models for analyzing health-related quality of life scores over time. *BMC Med Res Methodol*, 12, 144, doi:10.1186/1471-2288-12-144.
40. Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied longitudinal analysis* (2nd ed., Wiley series in probability and statistics). Hoboken, N.J.: Wiley.
41. Johnson, M. L., Crown, W., Martin, B. C., Dormuth, C. R., & Siebert, U. (2009). Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources:

- the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report--Part III. *Value Health*, 12(8), 1062-1073, doi:10.1111/j.1524-4733.2009.00602.x.
42. R Core Team (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>.
  43. Ara, R., Brazier, J., & Young, T. (2017). Recommended Methods for the Collection of Health State Utility Value Evidence in Clinical Studies. *Pharmacoeconomics*, 35(Suppl 1), 67-75, doi:10.1007/s40273-017-0549-6.
  44. International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and Avalon Health Economics (May 23, 2017). Minimal Clinically Important Difference in EQ-5D: We Can Calculate it – But Does That Mean We Should? [https://www.ispor.org/docs/default-source/presentations/1066.pdf?sfvrsn=25feffd6\\_1](https://www.ispor.org/docs/default-source/presentations/1066.pdf?sfvrsn=25feffd6_1). Accessed August 10, 2020.
  45. Khunti, K., Alsifri, S., Aronson, R., Cigrovski Berkovic, M., Enters-Weijnen, C., Forsen, T., et al. (2016). Rates and predictors of hypoglycaemia in 27 585 people from 24 countries with insulin-treated type 1 and type 2 diabetes: the global HAT study. *Diabetes Obes Metab*, 18(9), 907-915, doi:10.1111/dom.12689.
  46. Griffiths, A., Paracha, N., Davies, A., Branscombe, N., Cowie, M. R., & Sculpher, M. (2017). Analyzing Health-Related Quality of Life Data to Estimate Parameters for Cost-Effectiveness Models: An Example Using Longitudinal EQ-5D Data from the SHIFT Randomized Controlled Trial. *Adv Ther*, 34(3), 753-764, doi:10.1007/s12325-016-0471-x.
  47. Gardiner, J. C., Luo, Z., & Roman, L. A. (2009). Fixed effects, random effects and GEE: what are the differences? *Stat Med*, 28(2), 221-239, doi:10.1002/sim.3478.
  48. Hubbard, A. E., Ahern, J., Fleischer, N. L., Van der Laan, M., Lippman, S. A., Jewell, N., et al. (2010). To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology*, 21(4), 467-474, doi:10.1097/EDE.0b013e3181caeb90.
  49. Faria, R., Gomes, M., Epstein, D., & White, I. R. (2014). A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. *Pharmacoeconomics*, 32(12), 1157-1170, doi:10.1007/s40273-014-0193-3.
  50. Xu, C., Li, Z., Xue, Y., Zhang, L., & Wang, M. (2019). An R package for model fitting, model selection and the simulation for longitudinal data with dropout missingness. *Commun Stat Simul Comput*, 48(9), 2812-2829, doi:10.1080/03610918.2018.1468457.
  51. Simons, C. L., Rivero-Arias, O., Yu, L. M., & Simon, J. (2015). Multiple imputation to deal with missing EQ-5D-3L data: Should we impute individual domains or the actual index? *Qual Life Res*, 24(4), 805-815, doi:10.1007/s11136-014-0837-y.

52. Haneuse, S., VanderWeele, T. J., & Arterburn, D. (2019). Using the E-Value to Assess the Potential Effect of Unmeasured Confounding in Observational Studies. *JAMA*, *321*(6), 602-603, doi:10.1001/jama.2018.21554.
53. VanderWeele, T. J., & Ding, P. (2017). Sensitivity Analysis in Observational Research: Introducing the E-Value. *Ann Intern Med*, *167*(4), 268-274, doi:10.7326/M16-2607.
54. International Hypoglycaemia Study, G. (2017). Glucose Concentrations of Less Than 3.0 mmol/L (54 mg/dL) Should Be Reported in Clinical Trials: A Joint Position Statement of the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care*, *40*(1), 155-157, doi:10.2337/dc16-2215.
55. Tang, K. (2015). Estimating productivity costs in health economic evaluations: a review of instruments and psychometric evidence. *Pharmacoeconomics*, *33*(1), 31-48, doi:10.1007/s40273-014-0209-z.
56. Reilly Associates Work Productivity and Activity Impairment Questionnaire. [http://www.reillyassociates.net/WPAI\\_General.html](http://www.reillyassociates.net/WPAI_General.html).
57. Brazier, J., Ara, R., Rowen, D., & Chevrou-Severac, H. (2017). A Review of Generic Preference-Based Measures for Use in Cost-Effectiveness Models. *Pharmacoeconomics*, *35*(Suppl 1), 21-31, doi:10.1007/s40273-017-0545-x.
58. Nauck, M. A., Buse, J. B., Mann, J. F. E., Pocock, S., Bosch-Traberg, H., Frimer-Larsen, H., et al. (2019). Health-related quality of life in people with type 2 diabetes participating in the LEADER trial. *Diabetes Obes Metab*, *21*(3), 525-532, doi:10.1111/dom.13547.
59. Briggs, A. H., Bhatt, D. L., Scirica, B. M., Raz, I., Johnston, K. M., Szabo, S. M., et al. (2017). Health-related quality-of-life implications of cardiovascular events in individuals with type 2 diabetes mellitus: A subanalysis from the Saxagliptin Assessment of Vascular Outcomes Recorded in Patients with Diabetes Mellitus (SAVOR)-TIMI 53 trial. *Diabetes Res Clin Pract*, *130*, 24-33, doi:10.1016/j.diabres.2016.12.019.
60. Hubens, K., Krol, M., Coast, J., Drummond, M. F., Brouwer, W. B. F., Uyl-de Groot, C. A., et al. (2021). Measurement Instruments of Productivity Loss of Paid and Unpaid Work: A Systematic Review and Assessment of Suitability for Health Economic Evaluations From a Societal Perspective. *Value Health*, *24*(11), 1686-1699, doi:10.1016/j.jval.2021.05.002.

**Supplementary Table 1. Potential independent variables for the generalized estimating equations models**

<b>Independent variables</b>	<b>Reference variables</b>
Age	Per additional year <sup>a</sup>
Gender	Dummy: Female
Ethnicity	Dummy: White
Education	Dummy: Post-secondary
Household income	Dummy: $\geq$ \$80,000
Employment status	Dummy: Employed
Living status	Dummy: Married/Living with partner
Alcohol consumption	Dummy: Heavy drinker <sup>b</sup>
Smoking	Dummy: Current smoker
Physical activity	Dummy: Inactive <sup>c</sup>
Physician visits	Per additional visit <sup>d</sup>
Emergency visits and hospitalizations	Per additional visit <sup>e</sup>
Self-care	Per additional score unit <sup>f</sup>
Blood glucose monitoring system	Dummy: Flash (continuous) glucose monitoring
Influenza immunization	Dummy: Annual influenza vaccination
COVID-19 immunization	Dummy: Fully vaccinated <sup>g</sup>
Pneumococcal immunization	Dummy: Vaccinated (ever)
Previous experience of NSHE	Dummy: History <sup>h</sup>
Previous experience of SHE	Dummy: History <sup>i</sup>
Duration of diabetes	Per additional year <sup>j</sup>
HbA1c (%)	Per additional unit <sup>k</sup>
Body mass index	Per additional kg/m <sup>2</sup> <sup>l</sup>
Combination of therapies	Dummy: Combination of insulin and secretagogues
Number of daily injections	Per additional insulin injection <sup>m</sup>
Duration of insulin therapy	Per additional year <sup>n</sup>

Macrovascular complications	Dummy: History <sup>o</sup>
Microvascular complications	Dummy: History <sup>p</sup>
Incident NSHE	Per additional event <sup>q</sup>
Incident SHE	Per additional event <sup>f</sup>

NSHE: non-severe hypoglycemic event; SHE: severe hypoglycemic event

- a. Sensitivity analysis: will be tested as a dummy variable ( $\geq 65$  years old).
- b. Heavy drinker: having  $\geq 4$  drinks in a single day at least once a month in the past year (5 for males).
- c. Physical inactivity:  $< 150$  minutes/week of moderate and vigorous physical activity.
- d. Medical visits: number of visits to primary care doctor and specialists over the last three months.
- e. Emergency visits and hospitalizations: number of emergency room visits and hospital admissions over the last six months.
- f. Self-care: total score of the Diabetes Self-Management Questionnaire (DSMQ).
- g. COVID-19 immunization: Complete vaccination is currently defined as 2 doses.
- h. Previous experience of non-severe hypoglycemic events: occurrence over the last three months.
- i. Previous experience of severe hypoglycemic events: occurrence over the last year.
- j. Sensitivity analysis: will be tested as a dummy variable ( $\geq 10$  years).
- k. Sensitivity analysis: will be tested as a dummy variable ( $\geq 7\%$ ).
- l. Sensitivity analysis: will be tested as a dummy variable ( $\geq 30$  kg/m<sup>2</sup>).
- m. Sensitivity analysis: will be tested as a dummy variable ( $\geq 5$  years).
- n. Sensitivity analysis: will be tested as a dummy variable ( $\geq 3$  injections/day).
- o. Macrovascular complications: myocardial infarction, ischemic heart disease, heart failure, stroke, peripheral vascular disease. Sensitivity analysis: will be tested per additional complication.
- p. Microvascular complications: renal failure, proteinuria, severe vision loss, retinopathy, macular oedema, cataract, amputation, neuropathy, ulcer. Sensitivity analysis: will be tested per additional complication.
- q. Non-severe hypoglycemic events: will be log-transformed. Interpretation of log-transformed independent variables is as follows: A 1% increase in  $x$  changes the dependent variable by coefficient/100. With a 100% increase (equivalent to 1 additional event), the coefficient can be interpreted directly without transformation. Sensitivity analysis: will be tested with original scale.
- r. Severe hypoglycemic events: will be log-transformed. Interpretation of log-transformed independent variables is as follows: A 1% increase in  $x$  changes the dependent variable by coefficient/100. With a 100% increase (equivalent to 1 additional event), the coefficient can be interpreted directly without transformation. Sensitivity analysis: will be tested with original scale.

## CHAPITRE 10 : DISCUSSION

---

Dans un système de santé avec ressources limitées, les évaluations économiques sont essentielles aux prises de décisions des payeurs publics.(3) Un manque de crédibilité des données pharmacoéconomiques peut mener à une décision de remboursement erronée, entraînant une allocation sous-optimale des ressources de la santé. L'objectif global de cette thèse était de comprendre et d'améliorer les pratiques méthodologiques pour la réalisation d'études générant des données pharmacoéconomiques en pratique réelle. Le premier article visait donc à comprendre les méthodologies employées par les chercheurs en pharmacoéconomie pour générer leurs données tout en évaluant la qualité des résultats obtenus. Le deuxième article a répertorié les lignes directrices actuelles sur la génération de données d'utilité afin d'en extraire les recommandations et développer la boîte à outils. Le troisième article visait à mettre en pratique la boîte à outils en rédigeant un protocole d'étude générant des données de pharmacoéconomie en pratique réelle.

La réalisation des différents volets de cette thèse a fait preuve d'une démarche systématique et d'une rigueur méthodologique. La thèse aborde spécifiquement la collecte de données primaires. Malgré l'intérêt et la pertinence des analyses de données secondaires, certains enjeux méthodologiques diffèrent lorsque les études nécessitent le recrutement prospectif de patients. Puisqu'il est plutôt rare que les données pharmacoéconomiques comme l'utilité et la productivité soient collectées de routine, un focus sur la génération de données primaires est plus adapté à ce type de recherche. Également, la thèse se concentre uniquement sur la pratique réelle et ne discute pas de la collecte de données dans le cadre d'essais cliniques qui représente un sujet de recherche à part entière. Enfin, le DMT2 a été choisi comme exemple de cas. Le DMT2 affecte un nombre significatif de patients et est un réel fardeau autant pour les patients eux-mêmes que pour le système de santé. L'ensemble des méthodologies discutées étaient génériques à la pharmacoéconomie et non spécifiques au DMT2. Il est donc attendu que la sélection d'un autre champ thérapeutique n'aurait pas affecté les conclusions de cette thèse.



Le premier volet consistait en une revue systématique de la littérature qui a été conduite selon un protocole prédéfini, utilisant une méthodologie rigoureuse et transparente permettant ainsi la réplication de la revue. La sélection des études, l'extraction des données et l'évaluation de la qualité a été effectuée par deux réviseurs indépendants où les désaccords ont été résolus au moyen de discussions et consultations avec un troisième réviseur lorsque nécessaire. De plus, la stratégie de recherche a été réalisée en utilisant des filtres validés. La revue a inclus 60 études pharmacoéconomiques en DMT2 en pratique réelle dont les méthodologies variaient selon leurs objectifs respectifs. Les études pouvaient inclure une population diabétique générale sans critère d'éligibilité strict ou pouvaient se restreindre à une intervention spécifique ou différents critères tels que la sévérité ou l'âge. Près de la moitié des études décrivait un seul groupe d'étude alors que l'autre moitié comparait au moins deux groupes. Les groupes d'exposition étaient majoritairement des états de santé ou des interventions de santé. La plupart des études avaient un devis transversal. Un des résultats importants de l'extraction des méthodologies a été le manque de transparence. En effet, les auteurs ne rapportaient pas toujours toutes les informations pertinentes à la conduite de l'étude, telles que la méthode de mesure de l'exposition, la source pour la sélection des participants, le nombre de sites d'études, la durée de recrutement, le calcul de la taille d'échantillon, la méthode d'échantillonnage, le taux de participation, le taux de données manquantes, l'approbation par un comité éthique, l'obtention du consentement du patient, la déclaration des conflits d'intérêt et la source de financement. Un autre résultat important de l'étude a été la démonstration d'une faible qualité méthodologique. En effet, le quart des issues analytiques ont été jugées à risque modéré de biais, la moitié à risque sérieux et le quart à risque critique. Le quart des issues descriptives avaient un score de qualité sous les 50% et plus du tiers avaient un score sous les 75%. La plus grande problématique rencontrée dans le volet 1 était l'interprétation des études. En effet, tel que mentionné ci-dessus, les articles manquaient de transparence et des informations essentielles étaient manquantes, notamment les méthodes de mesure de l'exposition et de sélection des patients. Ceci affectait la compréhension et l'évaluation des articles, où la consultation d'un troisième réviseur a été nécessaire à quelques reprises. De plus, puisque certaines informations essentielles étaient manquantes, le risque de biais ne pouvait pas toujours être établi pour certains domaines spécifiques. Cela signifie que la qualité des études est potentiellement surestimée. Lors de l'évaluation des biais à l'aide de l'outil ROBINS(55), si par exemple, le taux de données manquantes n'est pas mentionné dans la publication, le risque de biais associé au

domaine « données manquantes » est considéré comme non évaluable. Un domaine non évaluable n'est pas tenu en compte dans l'évaluation globale du risque de biais, ce qui peut potentiellement surestimer la qualité réelle de l'étude. Une autre limite de la revue est que la littérature grise n'a pas été considérée et que la recherche a été limitée aux articles publiés en anglais ou en français. Par contre, puisque le but de la revue n'était pas de répertorier l'ensemble des résultats sur le sujet de recherche, mais plutôt de démontrer les tendances méthodologiques, il est improbable que la conclusion aurait changé en incluant des études identifiées dans la littérature grise. La période de publication pour la sélection des études était de courte durée. Il est possible qu'une période plus longue aurait permis d'identifier d'autres méthodologies que celles répertoriées dans la revue. L'extension de la période de sélection aurait également mené à la même conclusion en identifiant un plus grand nombre de méthodologies hétérogènes menant encore à la démonstration d'un manque d'uniformité et de crédibilité. En effet, puisque ce domaine de recherche a beaucoup évolué au cours des années, reculer plus loin dans le temps aurait uniquement renforcé la présente conclusion. La restriction à une plus courte période a donc permis d'établir un portrait de la situation actuelle. Aussi, dans le cadre d'une même étude, les méthodologies peuvent varier selon les différentes issues mesurées. La revue était limitée aux issues pharmacoéconomiques et pourrait donc ne pas refléter les méthodologies employées pour la collecte d'issues cliniques. Il convient également de mentionner que les résultats des publications n'ont pas été extraits, seulement les méthodologies en lien avec la question de recherche. Ceci explique pourquoi cette revue systématique a été écrite en suivant les recommandations PRISMA, mais n'a pas l'appellation précise qui aurait impliqué l'extraction des résultats des études individuelles.(184) Les conclusions de l'article 1 illustrent le manque de rigueur méthodologique de certaines études pharmacoéconomiques signifiant que les résultats doivent être interprétés avec prudence. La faible qualité met en doute la pertinence des résultats à la réalisation d'évaluations économiques crédibles. Le manque de transparence et de crédibilité des études met également de l'avant le besoin de standardisation des études générant des données pharmacoéconomiques. Un cadre de recherche peut aider à la conduite d'études moins à risque de biais, ce qui a mené un deuxième volet de cette thèse qui visait à développer une boîte à outils.

Il n'existe pas de cadre de recherche reconnu pour guider les chercheurs dans la conduite d'études pharmacoéconomiques en pratique réelle. Les publications des agences d'évaluation des technologies de la santé et des agences professionnelles de pharmacoéconomie ont été identifiées afin d'en extraire les recommandations sur la génération de données pharmacoéconomiques, plus précisément les scores d'utilité. Les recommandations extraites visaient particulièrement le choix d'instrument de mesure et la sélection des répondants. Étant donné le manque de recommandations pour éviter les biais dans la conception d'une étude, une deuxième recherche a identifié les publications sur les bonnes pratiques pour la recherche observationnelle en général indépendamment du type d'issues. Un constat important est le manque de guide unique pour les chercheurs qui désirent se référer à des lignes directrices. Le manque de transparence et de crédibilité des études générant des données pharmacoéconomiques en pratique réelle met de l'avant le besoin d'un cadre de recherche qui servirait de référence à suivre. Le deuxième volet visait donc à créer une boîte à outils, un premier pas vers un cadre de recherche officiel, à partir de l'ensemble des recommandations existantes pour la génération de données d'utilité et la recherche en pratique réelle. La boîte inclut différentes étapes méthodologiques afin de supporter les chercheurs dans la conception et conduite d'études tout en limitant les biais de confusion, sélection et information. Étant donné l'impact direct sur les résultats d'une évaluation économique, la génération des données pharmacoéconomiques de plus grande qualité aiderait à limiter l'incertitude associée au ICER d'une intervention thérapeutique. La boîte à outils a également dû être restreinte en termes d'issues. Bien que cette thèse discute d'un ensemble de données pharmacoéconomiques, le volet 2 se concentre uniquement sur la collecte des scores d'utilité associée à des états de santé. Initialement, les intentions étaient que la boîte à outils aborde l'ensemble des données pharmacoéconomiques, mais un seul article devenait rapidement surchargé perdant ainsi sa facilité de lecture et pertinence d'utilisation. La boîte à outils initiale avec les recommandations des agences professionnelles de pharmacoépidémiologie, recherche évaluative et pharmacoéconomie ont été tout de même présentées à la section 6.2, bien que non publiées. Également, la boîte à outils couvre uniquement les recommandations liées aux biais et ne touche pas à d'autres enjeux tels que l'éthique et la transparence. De plus, la boîte à outils vise uniquement la génération de données et n'aborde pas les techniques de modélisation économiques. La boîte à outils n'est ni une liste de vérification pour la rédaction de protocoles ni un outil d'évaluation de la qualité. Enfin, la boîte à outils n'aborde pas en profondeur de chaque biais, donc

les chercheurs intéressés par une discussion en détail sur un certain aspect devront se tourner vers les références sources ou vers des ouvrages complémentaires. À l’opposé, le survol plus large des principaux biais (confusion, sélection, information) permet de couvrir les points majeurs dans un seul et même outil. La revue conduite pour le développement de la boîte à outils n’était pas une revue systématique de la littérature qui par définition tente d’identifier toutes les preuves sur un sujet de recherche dans le but d’avoir une vue d’ensemble objective et transparente sur les résultats répondant à une question de recherche précise.(185) Une revue ciblée a été préférée puisque le but n’était pas de comptabiliser des résultats en tant que tel, par exemple les effets d’une intervention de santé, ni de répertorier toutes les publications sur sujet, mais plutôt un ensemble de lignes directrices. Une liste d’agences professionnelles de recherche (pharmacoépidémiologie, recherche évaluative, pharmacoéconomie) en lien avec l’élaboration de la boîte à outils a plutôt été ciblée. La méthodologie demeure rigoureuse, impliquant l’aide d’un deuxième réviseur et établissant la stratégie de recherche à partir d’une revue publiée.(125)

Afin de mettre de l’avant la pertinence de la boîte à outils à supporter les chercheurs en pharmacoéconomie, le troisième volet consistait à rédiger un protocole en respectant les standards de la boîte à outils. Le troisième volet visait tout d’abord à identifier un besoin pour la génération de données pharmacoéconomiques. Les données d’utilité et de perte de productivité associées à l’hypoglycémie chez les patients DMT2 traités à l’insuline sont limités, particulièrement au Canada. Un protocole de recherche a donc été rédigé en suivant les recommandations publiées de la boîte à outils et également celles identifiées mais non publiées. Les résultats futurs de cette étude pourront combler en besoin en fournissant des données pharmacoéconomiques de qualité pour la modélisation des traitements à l’insuline en DMT2. Le volet 3 vise à répondre un besoin concret et est conçu pour assurer sa pertinence pour la modélisation économique. L’étude mesure l’impact de l’hypoglycémie sur plusieurs issues pharmacoéconomiques, incluant l’utilité, l’absentéisme, le présentéisme et le travail non rémunéré (activités courantes). Le protocole du volet 3 suit les recommandations de la boîte à outils et tente de limiter le risque de biais tout en respectant les contraintes, notamment budgétaires, des chercheurs. Le biais de confusion tente d’être limité par le devis d’étude et la mesure des confondants connus. Une étude de sensibilité est incluse au protocole pour la possibilité de confusion résiduelle. Un plan statistique pour la gestion des

données manquantes est également inclus en cas de besoin. Également, la classification de l'exposition répond aux exigences des agences d'évaluation des technologies de la santé. La fréquence et les outils de mesure tentent de limiter le biais de rappel. Les issues pharmacoéconomiques sont mesurées à l'aide d'instruments validés. La population recrutée vise à assurer la représentativité d'un modèle économique. Des démarches systématiques et préventives sont mises en place pour limiter le biais de sélection au recrutement et durant l'étude (pertes au suivi). Le devis d'étude visait des patients prévalents plutôt que des nouveaux utilisateurs. L'inclusion de patients prévalents peut mener à un biais de sélection si par exemple un traitement d'intérêt est arrêté prématurément chez les patients non-répondants ou intolérants, où la fréquence de l'issue d'intérêt pourrait être sous-estimée.(55) Ce biais est un enjeu important dans la recherche d'efficacité comparative.(55) Dans la présente étude, les issues pourraient être surestimées si par exemple les patients souffrant d'hypoglycémie ont prématurément arrêté leur traitement à l'insuline, où l'échantillon à l'étude serait représentatif d'un sous-groupe plus sain. Ce scénario est cliniquement improbable.(186, 187) En effet, l'hypoglycémie est un effet à long terme de l'insuline, c'est-à-dire que les événements hypoglycémiques sont rares dans la première année d'utilisation, puis peuvent graduellement augmenter au fil du temps.(186, 187) À cet effet, un devis incluant uniquement des patients prévalents a été utilisé dans deux études canadiennes à grande échelle, l'étude HAT (179) et *InHypo-DM*(180), qui avaient pour objectif de mesurer la fréquence d'évènements hypoglycémiques chez les patients traités à l'insuline. Un devis avec patients prévalents a donc été choisi afin d'être en mesure d'évaluer l'impact de l'hypoglycémie, considérant que les événements hypoglycémiques sont rares durant la première année d'initiation de l'insuline.(186, 187) Puisque le devis inclut des patients prévalents, les questionnaires ne sont pas complétés à l'entrée dans l'étude, considérant que la date de recrutement ne représente pas la date de début de traitement (*baseline*). La complétion de questionnaires à l'entrée aurait exigé de mesurer les événements hypoglycémiques rétrospectivement avant l'entrée dans l'étude, ce qui aurait augmenté le risque de biais de rappel.(188)

La conception de l'étude permettra la collecte de données robustes pour de futures évaluations économiques. La conduite de l'étude au-delà de la conception du protocole ne faisait pas partie des objectifs de cette thèse. Dans le contexte de la pandémie COVID-19, la réalisation

de ce type d'étude prospective était simplement irréaliste dans le cadre d'un projet de doctorat non financé. Cette décision a été prise en accord avec le comité avisur. L'objectif du volet 3 était donc de mettre en pratique la boîte à outils pour la conception d'une étude rigoureuse et non la conduite de l'étude. Le développement d'un protocole de recherche a permis d'atteindre l'objectif spécifique du troisième volet ainsi que l'objectif global de cette thèse qui encadre uniquement les pratiques méthodologiques. Le cadre de recherche explique également pourquoi la revue de la littérature du volet 1 n'a pas extrait les résultats des études incluses, mais uniquement les méthodologies. Au même titre, l'absence de résultats du projet de recherche du volet 3 n'empêche aucunement de discuter des pratiques méthodologiques, puisqu'elles sont définies a priori dans le protocole. Par contre, malgré la robustesse du protocole, il faut s'attendre à faire face à des défis lors de la conduite de l'étude. Dans les études auxquelles j'ai participé (NCT02315365, NCT02139215, NCT02228421, NCT02135757), le recrutement de patients était probablement le plus gros défi. La responsabilité d'identifier les patients potentiels et de les recruter dans l'étude revient aux sites participant au projet de recherche. Le choix des sites d'étude est donc clé au succès du projet. Tout d'abord, les sites sélectionnés doivent avoir un bassin potentiel de patients suffisant pour répondre au besoin de recherche. Une étude de faisabilité est habituellement conduite afin d'évaluer la capacité de recrutement. Par contre, sélectionner les plus gros centres de recherche n'est pas toujours la meilleure option. En effet, il est commun que les grands centres participent déjà à de nombreuses études cliniques et projets de recherche, réduisant significativement le bassin de patients disponibles, mais également les ressources pour la conduite du projet, le personnel de recherche étant déjà surchargé. À cet effet, il est donc très important de sélectionner des sites de recherche motivés. Le développement d'une bonne relation avec l'équipe de recherche est essentiel afin d'assurer une motivation continue. Lorsque disponibles, les performances passées de recrutement peuvent aussi aider à sélectionner des sites. La motivation de l'équipe de recherche doit être maintenue tout au long de l'étude, des appels et/ou visites régulières doivent donc être planifiés. L'envoi d'infolettres est aussi une technique pour garder la motivation et même créer une compétition amicale entre les sites sur la performance de recrutement. Il est aussi important d'utiliser des outils permettant de suivre et de prédire le taux de recrutement. Dans le cas où la vitesse de recrutement ne répond pas aux attentes en termes d'échéancier, une option pourrait être par exemple d'ajouter un ou plusieurs autres sites participants. De par mon expérience personnelle, le taux de participation des patients recrutés dans

des cliniques/centres de recherche est beaucoup plus élevé que par exemple des panels en ligne. Le risque de données manquantes est un autre aspect à prévoir. Un plan de gestion des données manquantes doit bien sûr être développé, mais la prévention est clé, où des rappels aux patients peuvent s'avérer grandement utile pour assurer la complétion des questionnaires à temps. Enfin, un autre enjeu préalablement rencontré dans d'autres projets de recherche était des difficultés lors de la complétion des formulaires d'extraction (*case report form*) par l'équipe de recherche. Il est possible que les formulaires d'extraction (*case report form*) manquent de clarté sur certaines variables ou présentent des problèmes techniques lorsque complétés électroniquement. Il peut donc s'avérer utile de vérifier la qualité de complétion après le recrutement par exemple des dix premiers patients et de réajuster au besoin.

## CHAPITRE 11 : CONTRIBUTIONS ET PERSPECTIVES

---

Le but de cette thèse était d'étudier et contribuer à l'amélioration des pratiques méthodologiques pour la génération de données pharmacoéconomiques en pratique réelle. Considérant que les décisions de remboursement sont entre autres basées sur les évaluations économiques, la qualité des ACU a un impact direct sur l'allocation des ressources du système de santé canadien. Les évaluations économiques doivent donc être réalisées de façon rigoureuse, expliquant le nombre important de lignes directrices sur la modélisation économique, incluant notamment celles des agences d'évaluation des technologies de la santé(3, 4) et celles de l'agence professionnelle de pharmacoéconomie ISPOR.(189) L'incertitude d'une évaluation économique englobe différents aspects tels que l'incertitude méthodologique, l'incertitude structurelle et l'incertitude paramétrique.(3) L'incertitude paramétrique, c'est-à-dire l'incertitude associée à la justesse des paramètres, peut être introduite à plusieurs niveaux débutant par la source originale (étude générant de nouvelles données), la recherche et sélection des sources originales, puis la synthèse des données.(161) Bien que des recommandations sur la synthèse et l'utilisation des données aient été publiées, les lignes directrices sur la génération de nouvelles données (études originales) sont limitées.(190) De plus, l'incertitude paramétrique peut être introduite par tous les différents types de paramètres incluent dans une ACU, notamment les données cliniques (efficacité et innocuité) et les données pharmacoéconomiques, telles que les données d'utilité et de coûts (directs et indirects). En effet, l'impact sur les résultats n'est pas exclusif aux paramètres cliniques, où le ICER peut également être très sensible à des changements au niveau des paramètres pharmacoéconomiques.(191) Considérant que les paramètres biaisés peuvent mener à une conclusion erronée quant à la l'efficience des interventions en santé, il est surprenant qu'il y ait un manque flagrant de lignes directrices sur la génération de nouvelle données pharmacoéconomiques. Pourtant, le risque de biais associé à la génération de données pharmacoéconomiques est discuté depuis bien longtemps. Par exemple, cela fait plus d'une décennie que le simple choix d'un modèle de régression pour les scores d'utilité fait débat.(163, 164, 167, 176, 192) Il en va de même pour le risque de biais durant la collecte et l'analyse des données de coûts aux fins d'évaluations économiques.(38, 39, 41-43) Quant au choix du devis d'étude, une publication largement citée par Alva et al. a démontré à l'aide des données de l'étude



UKPDS en DMT2 qu'un devis transversal peut générer des données d'utilité biaisées dû à l'hétérogénéité entre les patients.(165) Par contre, un devis longitudinal est associé à un risque de biais dû aux données manquantes et ce, malgré les différentes méthodes proposées pour en tenir compte.(52, 177, 193) Bien au-delà du choix de l'instrument de mesure en soi, il y a de nombreuses autres décisions méthodologiques (devis, analyse statistique, gestion des données manquantes, etc.) qui influenceront la qualité d'une étude et des paramètres pharmacoéconomiques générés aux fins de réalisation d'une ACU. Corroborant les constats de la littérature, l'article 1 de cette thèse a démontré que les craintes face à la présence de biais sont bien d'actualité, où la qualité de récentes publications en DMT2 était faible, voire critique. Les conclusions de l'article 1 mettent ainsi de l'avant que l'ensemble des ouvrages sur les biais demeure insuffisant pour assurer une rigueur méthodologique et qu'il y a un besoin concret des chercheurs en pharmacoéconomie d'être mieux encadrés. L'article 2 est un premier pas vers ce cadre de recherche et l'article 3 est un exemple de protocole robuste et crédible suivant ces bonnes pratiques de recherche, servant d'exemple de son applicabilité et de sa pertinence pour la conception d'étude en pharmacoéconomie. Comme perspective future, la conduite de l'étude permettra de générer des résultats servant d'évidence complémentaire lors d'évaluations économiques canadiennes visant les patients atteints de DMT2 et traités à l'insuline. Également, il serait intéressant de tester la boîte à outils dans divers champs thérapeutiques.

Au cœur de ce projet de recherche, le volet 2 a démontré le manque de cadre de référence unique pour les chercheurs voulant générer des données de pharmacoéconomie. À partir des recommandations actuelles, le volet 2 a développé une boîte à outils spécifiquement destinée aux chercheurs tentant de conceptualiser une étude robuste et pertinente à la modélisation économique. À ce jour, il n'existe aucune ligne directrice spécifique à la génération de données pharmacoéconomiques en milieu de pratique réelle. La création de la boîte à outils sert de premier pas pour répondre au besoin de standardisation. Ce projet de recherche ne visait pas à réinventer les bonnes pratiques de recherche, mais plutôt à mettre de l'avant l'applicabilité des recommandations existantes aux études générant des données de pharmacoéconomie en pratique réelle. De plus, la boîte à outils se veut être un apport concret pour les investigateurs en répertoriant les recommandations dans un seul et même cadre de référence, visant ainsi à contribuer à

l'amélioration des bonnes pratiques méthodologiques. L'amélioration des méthodologies serait d'un apport significatif pour les payeurs en aidant à produire des données crédibles réduisant les incertitudes entourant les prises de décision de remboursement. Bien que l'étendue de la boîte à outils ait été limitée dans l'article 2, cette thèse contribue plus largement en répertoriant les recommandations d'un nombre significatif d'agences professionnelles de recherche en santé et de pharmacoeconomie. Cette recherche initiale n'était pas limitée aux données d'utilité, permettant de couvrir les différents objectifs de cette thèse, notamment la collecte de données de perte de productivité.

Pour les chercheurs qui désirent réaliser une évaluation économique, il existe des lignes directrices servant d'encadrement, telles que les recommandations des agences d'évaluations des technologies de la santé.(3, 4) Également, ISPOR a publié une liste de vérification pour améliorer la transparence lors de la publication d'évaluations économiques.(194) Au fil des années, la *CHEERS checklist* a permis de standardiser les évaluations économiques et plusieurs journaux l'exige désormais aux fins de publication.(194) Il y a également ce même besoin d'uniformité pour les études générant les données pharmacoéconomiques et non seulement pour la modélisation économique en tant que tel. La boîte à outils ne consiste pas en des lignes directrices officielles, bien que ses fondements soient basés sur les bonnes pratiques publiées. La boîte à outils sert de premier pas pour sensibiliser les chercheurs au risque de biais lors de la conduite d'études et pourrait susciter l'intérêt de développer de lignes directrices servant de guide de référence. Comme perspective future, d'autres recherches pourraient être réalisées afin de couvrir certains aspects complémentaires, tels que l'éthique, la transparence, la reproductibilité, etc.

## CHAPITRE 12 : CONCLUSION

---

Cette thèse de doctorat est au cœur d'un sujet d'actualité chez les décideurs en santé qui font face à l'utilisation croissante de données en pratique réelle. Les risques de biais associés à ce type de recherche sont bien connus à ce jour et la communauté scientifique est sensibilisée à l'impact potentiel de ces biais sur l'interprétation des résultats. Alors pourquoi, tel que démontré dans cette thèse, persiste-t-il un manque de rigueur méthodologique lors de la génération de données pharmacoéconomiques? Pourtant, une évaluation économique erronée peut directement affecter l'optimisation de nos ressources de la santé. L'accès aux technologies les plus efficaces est primordial pour les patients et plus globalement pour notre système de soins.

Considérant l'émergence de nombreuses thérapies avec prix premium, il serait temps que les bonnes pratiques de recherche passent de la théorie à la pratique. Mais comment? En sensibilisant la communauté de chercheurs en pharmacoéconomie à viser un niveau hiérarchique d'évidence plus élevé, ce qui se veut être la contribution de cette thèse. Cette thèse de doctorat corrobore les craintes sur la recherche en pratique réelle en pharmacoéconomie et propose plusieurs pistes de solutions, rassemblées dans une boîte à outils. Bien que les notions théoriques sur les biais soient bien connues, la standardisation des méthodologies pour générer des données pharmacoéconomiques en pratique réelle est un besoin qui demeure non comblé. Afin de mobiliser les chercheurs et réellement augmenter la crédibilité des études pharmacoéconomiques en pratique réelle, des lignes directrices sont essentielles.

Ultimement, un cadre de recherche officiel permettrait de monter la barre en termes de crédibilité. Ceci un appel aux experts en pharmacoéconomie, à quand la prochaine *ISPOR checklist*?

## BIBLIOGRAPHIE

---

1. Santé Canada. Ligne directrice - Gestion des présentations de drogues et demandes [updated 2019 Jul 25; cited 2021 Jan 26]. Available from: <https://www.canada.ca/content/dam/hc-sc/documents/services/drugs-health-products/drug-products/applications-submissions/guidance-documents/management-drug-submissions/industry/management-drug-submissions-gestion-presentations-demandes-drogues-fra.pdf>.
2. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10(1):37-48.
3. Agence canadienne des médicaments et des technologies de la santé. Lignes directrices de l'évaluation économique des technologies de la santé au Canada, 4e édition [updated 2017 Mar; cited 2021 Jan 26]. Available from: [https://www.cadth.ca/sites/default/files/pdf/guidelines\\_for\\_the\\_economic\\_evaluation\\_of\\_health\\_technologies\\_canada\\_4th\\_ed\\_f.pdf](https://www.cadth.ca/sites/default/files/pdf/guidelines_for_the_economic_evaluation_of_health_technologies_canada_4th_ed_f.pdf).
4. Institut national d'excellence en santé et en services sociaux. Guide de soumission d'une demande à l'INESSS [updated 2018 Sep; cited 2021 Jan 26]. Available from: [https://www.inesss.qc.ca/fileadmin/doc/INESSS/Inscription\\_medicaments/Fiches\\_inscription/Guide\\_soumission.pdf](https://www.inesss.qc.ca/fileadmin/doc/INESSS/Inscription_medicaments/Fiches_inscription/Guide_soumission.pdf).
5. Garrison LP, Jr., Neumann PJ, Erickson P, Marshall D, Mullins CD. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. *Value Health*. 2007;10(5):326-35.
6. Makady A, de Boer A, Hillege H, Klungel O, Goettsch W. What Is Real-World Data? A Review of Definitions Based on Literature and Stakeholder Interviews. *Value Health*. 2017;20(7):858-65.
7. Diabetes Canada. Diabetes in Canada [updated 2020 Feb; cited 2021 Jan 26]. Available from: [https://diabetes.ca/DiabetesCanadaWebsite/media/Advocacy-and-Policy/Backgrounder/2020\\_Backgrounder\\_Canada\\_English\\_FINAL.pdf](https://diabetes.ca/DiabetesCanadaWebsite/media/Advocacy-and-Policy/Backgrounder/2020_Backgrounder_Canada_English_FINAL.pdf).
8. Brazier J, Ara R, Rowen D, Chevrou-Severac H. A Review of Generic Preference-Based Measures for Use in Cost-Effectiveness Models. *Pharmacoeconomics*. 2017;35(Suppl 1):21-31.
9. Torrance GW, Feeny D, Furlong W. Visual analog scales: do they have a role in the measurement of preferences for health states? *Med Decis Making*. 2001;21(4):329-34.
10. Rowen D, Azzabi Zouraq I, Chevrou-Severac H, van Hout B. International Regulations and Recommendations for Utility Data for Health Technology Assessment. *Pharmacoeconomics*. 2017;35(Suppl 1):11-9.
11. Sintonen H. The 15D instrument of health-related quality of life: properties and applications. *Ann Med*. 2001;33(5):328-36.

12. Richardson J, Iezzi A, Khan MA, Maxwell A. Validity and reliability of the Assessment of Quality of Life (AQoL)-8D multi-attribute utility instrument. *Patient*. 2014;7(1):85-96.
13. Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095-108.
14. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727-36.
15. Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Med Care*. 2002;40(2):113-28.
16. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ*. 2002;21(2):271-92.
17. Kaplan RM, Anderson JP. A general health policy model: update and applications. *Health Serv Res*. 1988;23(2):203-35.
18. Canadian Agency for Drugs and Technologies in Health (CADTH). New Drugs for Type 2 Diabetes: Second-Line Therapy – Science Report [updated Sep 2017; cited 2021 Jan 26]. Available from: [https://cadth.ca/sites/default/files/pdf/TR0012\\_T2D\\_Science\\_Report.pdf](https://cadth.ca/sites/default/files/pdf/TR0012_T2D_Science_Report.pdf).
19. Xie F, Pullenayegum E, Gaebel K, Bansback N, Bryan S, Ohinmaa A, et al. A Time Trade-off-derived Value Set of the EQ-5D-5L for Canada. *Med Care*. 2016;54(1):98-105.
20. Canadian Agency for Drugs and Technologies in Health (CADTH). Using Canadian administrative databases to derive economic data for health technology assessments. [updated Mar 2009; cited 2022 Mar 11]. Available from: [https://www.cadth.ca/media/pdf/H0483\\_Canadian\\_Admin\\_Databases\\_mg\\_e.pdf](https://www.cadth.ca/media/pdf/H0483_Canadian_Admin_Databases_mg_e.pdf).
21. Tang K. Estimating productivity costs in health economic evaluations: a review of instruments and psychometric evidence. *Pharmacoeconomics*. 2015;33(1):31-48.
22. Drummond M, Barbieri M, Cook J, Glick HA, Lis J, Malik F, et al. Transferability of economic evaluations across jurisdictions: ISPOR Good Research Practices Task Force report. *Value Health*. 2009;12(4):409-18.
23. van Roijen L, Essink-Bot ML, Koopmanschap MA, Bonsel G, Rutten FF. Labor and health status in economic evaluation of health care. The Health and Labor Questionnaire. *Int J Technol Assess Health Care*. 1996;12(3):405-15.
24. Kessler RC, Ames M, Hymel PA, Loeppke R, McKenas DK, Richling DE, et al. Using the World Health Organization Health and Work Performance Questionnaire (HPQ) to evaluate the indirect workplace costs of illness. *J Occup Environ Med*. 2004;46(6 Suppl):S23-37.

25. Kumar RN, Hass SL, Li JZ, Nickens DJ, Daenzer CL, Wathen LK. Validation of the Health-Related Productivity Questionnaire Diary (HRPQ-D) on a sample of patients with infectious mononucleosis: results from a phase 1 multicenter clinical trial. *J Occup Environ Med.* 2003;45(8):899-907.
26. Jacob-Tacke KH, Koopmanschap MA, Meerding WJ, Severens JL. Correcting for compensating mechanisms related to productivity costs in economic evaluations of health care programmes. *Health Econ.* 2005;14(5):435-43.
27. Brouwer WB, Koopmanschap MA, Rutten FF. Productivity losses without absence: measurement validation and empirical evidence. *Health Policy.* 1999;48(1):13-27.
28. Turpin RS, Ozminkowski RJ, Sharda CE, Collins JJ, Berger ML, Billotti GM, et al. Reliability and validity of the Stanford Presenteeism Scale. *J Occup Environ Med.* 2004;46(11):1123-33.
29. Zhang W, Bansback N, Kopec J, Anis AH. Measuring time input loss among patients with rheumatoid arthritis: validity and reliability of the Valuation of Lost Productivity questionnaire. *J Occup Environ Med.* 2011;53(5):530-6.
30. Stewart WF, Ricci JA, Leotta C. Health-related lost productive time (LPT): recall interval and bias in LPT estimates. *J Occup Environ Med.* 2004;46(6 Suppl):S12-22.
31. Lerner D, Amick BC, 3rd, Rogers WH, Malspeis S, Bungay K, Cynn D. The Work Limitations Questionnaire. *Med Care.* 2001;39(1):72-85.
32. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. *Pharmacoeconomics.* 1993;4(5):353-65.
33. Goetzel RZ, Ozminkowski RJ, Long SR. Development and reliability analysis of the Work Productivity Short Inventory (WPSI) instrument measuring employee health and productivity. *J Occup Environ Med.* 2003;45(7):743-62.
34. Bouwmans C, Krol M, Severens H, Koopmanschap M, Brouwer W, Hakkaart-van Roijen L. The iMTA Productivity Cost Questionnaire: A Standardized Instrument for Measuring and Valuing Health-Related Productivity Losses. *Value Health.* 2015;18(6):753-8.
35. Hubens K, Krol M, Coast J, Drummond MF, Brouwer WBF, Uyl-de Groot CA, et al. Measurement Instruments of Productivity Loss of Paid and Unpaid Work: A Systematic Review and Assessment of Suitability for Health Economic Evaluations From a Societal Perspective. *Value Health.* 2021;24(11):1686-99.
36. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2015.
37. Sterne JAC, Hernán, M.A., McAleenan, A., Reeves, B.C., Higgins, J.P.T. Chapter 25: Assessing risk of bias in a non-randomized study. In: Higgins JPT, Thomas, J., Chandler, J.,

Cumpston, M., Li, T., Page, M.J., et al., editor. Cochrane Handbook for Systematic Reviews of Interventions version 60 Available from: [www.trainingcochrane.org/handbook2019](http://www.trainingcochrane.org/handbook2019).

38. Clement Nee Shrive FM, Ghali WA, Donaldson C, Manns BJ. The impact of using different costing methods on the results of an economic evaluation of cardiac care: microcosting vs gross-costing approaches. *Health Econ.* 2009;18(4):377-88.
39. Evans CJ, Crawford B. Data collection methods in prospective economic evaluations: how accurate are the results? *Value Health.* 2000;3(4):277-86.
40. Faries D, Peng X, Pawaskar M, Price K, Stamey JD, Seaman JW, Jr. Evaluating the impact of unmeasured confounding with internal validation data: an example cost evaluation in type 2 diabetes. *Value Health.* 2013;16(2):259-66.
41. Franklin M, Lomas J, Walker S, Young T. An Educational Review About Using Cost Data for the Purpose of Cost-Effectiveness Analysis. *Pharmacoeconomics.* 2019;37(5):631-43.
42. Frappier J, Tremblay G, Charny M, Cloutier LM. Costing bias in economic evaluations. *J Med Econ.* 2015;18(8):596-9.
43. Fukuda H, Imanaka Y. Assessment of transparency of cost estimates in economic evaluations of patient safety programmes. *J Eval Clin Pract.* 2009;15(3):451-9.
44. Guertin JR, Bowen JM, De Rose G, O'Reilly DJ, Tarride JE. Illustration of the Impact of Unmeasured Confounding Within an Economic Evaluation Based on Nonrandomized Data. *MDM Policy Pract.* 2017;2(1):2381468317697711.
45. Guertin JR, Conombo B, Langevin R, Bergeron F, Holbrook A, Humphries B, et al. A Systematic Review of Methods Used for Confounding Adjustment in Observational Economic Evaluations in Cardiology Conducted between 2013 and 2017. *Med Decis Making.* 2020;40(5):582-95.
46. Handorf EA, Heitjan DF, Bekelman JE, Mitra N. Estimating cost-effectiveness from claims and registry data with measured and unmeasured confounders. *Stat Methods Med Res.* 2019;28(7):2227-42.
47. Harkanen T, Maljanen T, Lindfors O, Virtala E, Knekt P. Confounding and missing data in cost-effectiveness analysis: comparing different methods. *Health Econ Rev.* 2013;3(1):8.
48. Kreif N, Grieve R, Sadique MZ. Statistical methods for cost-effectiveness analyses that use observational data: a critical appraisal tool and review of current practice. *Health Econ.* 2013;22(4):486-500.
49. Sackett DL. Bias in analytic research. *J Chronic Dis.* 1979;32(1-2):51-63.
50. The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on Methodological Standards in Pharmacoepidemiology (Revision 8)

EMA/95098/2010. [cited 2021 Jan 26]. Available from: [http://www.encepp.eu/standards\\_and\\_guidances](http://www.encepp.eu/standards_and_guidances).

51. Berger ML, Martin BC, Husereau D, Worley K, Allen JD, Yang W, et al. A questionnaire to assess the relevance and credibility of observational studies to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. *Value Health*. 2014;17(2):143-56.
52. Faria R, Gomes M, Epstein D, White IR. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. *Pharmacoeconomics*. 2014;32(12):1157-70.
53. Leurent B, Gomes M, Carpenter JR. Missing data in trial-based cost-effectiveness analysis: An incomplete journey. *Health Econ*. 2018;27(6):1024-40.
54. Halme AS, Tannenbaum C. Performance of a Bayesian Approach for Imputing Missing Data on the SF-12 Health-Related Quality-of-Life Measure. *Value Health*. 2018;21(12):1406-12.
55. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:i4919.
56. Lévesque LE, Hanley JA, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *BMJ*. 2010;340:b5087.
57. Diabetes Canada Clinical Practice Guidelines Expert C, Lipscombe L, Butalia S, Dasgupta K, Eurich DT, MacCallum L, et al. Pharmacologic Glycemic Management of Type 2 Diabetes in Adults: 2020 Update. *Can J Diabetes*. 2020;44(7):575-91.
58. Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. *Int J Epidemiol*. 2017;46(2):756-62.
59. Burcu M, Oehrlein E.M.,. Time-Dependent Confounders: Are They All the Same? [updated 2016 Apr; cited 2021 Jan 26]. Available from: [https://www.ispor.org/docs/default-source/publications/value-outcomes-spotlight/march-april-2016/vos-time-dependent-confounders.pdf?sfvrsn=ca518f95\\_2](https://www.ispor.org/docs/default-source/publications/value-outcomes-spotlight/march-april-2016/vos-time-dependent-confounders.pdf?sfvrsn=ca518f95_2).
60. Angrist JD, Pischke Jr-S. *Mastering 'metrics : the path from cause to effect*. Princeton ; Oxford: Princeton University Press; 2015. xv, 282 pages p.
61. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res*. 2011;46(3):399-424.
62. Hanley JA, Dendukuri N. Efficient sampling approaches to address confounding in database studies. *Stat Methods Med Res*. 2009;18(1):81-105.



63. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*. 2009;20(4):512-22.
64. Berger ML, Dreyer N, Anderson F, Towse A, Sedrakyan A, Normand SL. Prospective observational studies to assess comparative effectiveness: the ISPOR good research practices task force report. *Value Health*. 2012;15(2):217-30.
65. Duke Margolis Center for Health Policy. White Paper - Understanding the Need for Non-Interventional Studies Using Secondary Data to Generate Real-World Evidence for Regulatory Decision Making, and Demonstrating Their Credibility [updated 2019 Nov 25; cited 2021 Jan 26]. Available from: <https://healthpolicy.duke.edu/publications/understanding-need-non-interventional-studies-using-secondary-data-generate-real-world>.
66. Haneuse S, VanderWeele TJ, Arterburn D. Using the E-Value to Assess the Potential Effect of Unmeasured Confounding in Observational Studies. *JAMA*. 2019;321(6):602-3.
67. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Ann Intern Med*. 2017;167(4):268-74.
68. Dalal D, Dubreuil M, Peloquin C, Neogi T, Zhang Y, Choi H, et al. Meloxicam and risk of myocardial infarction: a population-based nested case-control study. *Rheumatology international*. 2017;37(12):2071-8.
69. Diabetes Canada Clinical Practice Guidelines Expert C, Yale JF, Paty B, Senior PA. Hypoglycemia. *Can J Diabetes*. 2018;42 Suppl 1:S104-S8.
70. Canadian Agency for Drugs and Technologies in Health (CADTH). Pharmacoeconomic Review Report - Insulin degludec (Tresiba) 2017 [Available from: [https://www.cadth.ca/sites/default/files/cdr/pharmacoeconomic/SR0521\\_Tresiba\\_PE\\_Report.pdf](https://www.cadth.ca/sites/default/files/cdr/pharmacoeconomic/SR0521_Tresiba_PE_Report.pdf)].
71. Canadian Agency for Drugs and Technologies in Health (CADTH). Pharmacoeconomic Review Report - Insulin degludec and liraglutide injection (Xultophy) 2019 [Available from: <https://www.cadth.ca/sites/default/files/cdr/pharmacoeconomic/sr0599-xultophy-pharmacoeconomic-review-report.pdf>].
72. Canadian Agency for Drugs and Technologies in Health (CADTH). Pharmacoeconomic Review Report - Insulin glargine + lixisenatide (Soliqua) 2019 [Available from: <https://www.cadth.ca/sites/default/files/cdr/pharmacoeconomic/sr0564-soliqua-pharmacoeconomic-report.pdf>].
73. Diabetes Canada Clinical Practice Guidelines Expert C, Imran SA, Agarwal G, Bajaj HS, Ross S. Targets for Glycemic Control. *Can J Diabetes*. 2018;42 Suppl 1:S42-S6.
74. Action to Control Cardiovascular Risk in Diabetes Follow-On Eye Study G, the Action to Control Cardiovascular Risk in Diabetes Follow-On Study G. Persistent Effects of Intensive Glycemic Control on Retinopathy in Type 2 Diabetes in the Action to Control Cardiovascular Risk in Diabetes (ACCORD) Follow-On Study. *Diabetes Care*. 2016;39(7):1089-100.

75. Hayward RA, Reaven PD, Wiitala WL, Bahn GD, Reda DJ, Ge L, et al. Follow-up of glycemic control and cardiovascular outcomes in type 2 diabetes. *N Engl J Med.* 2015;372(23):2197-206.
76. Zoungas S, Chalmers J, Neal B, Billot L, Li Q, Hiraakawa Y, et al. Follow-up of blood-pressure lowering and glucose control in type 2 diabetes. *N Engl J Med.* 2014;371(15):1392-406.
77. Nathan DM, Group DER. The diabetes control and complications trial/epidemiology of diabetes interventions and complications study at 30 years: overview. *Diabetes care.* 2014;37(1):9-16.
78. Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HAW. 10-Year Follow-up of Intensive Glucose Control in Type 2 Diabetes. *New England Journal of Medicine.* 2008;359(15):1577-89.
79. Hayes AJ, Leal J, Gray AM, Holman RR, Clarke PM. UKPDS outcomes model 2: a new version of a model to simulate lifetime health outcomes of patients with type 2 diabetes mellitus using data from the 30 year United Kingdom Prospective Diabetes Study: UKPDS 82. *Diabetologia.* 2013;56(9):1925-33.
80. Hoerger TJ, Segel, J.E., Zhang, P., and Sorensen, S.W. Validation of the CDC-RTI Diabetes Cost-Effectiveness Model. RTI Press publication No. MR-0013-0909. Research Triangle Park, NC: RTI International.2009.
81. McEwan P, Foos V, Palmer JL, Lamotte M, Lloyd A, Grant D. Validation of the IMS CORE Diabetes Model. *Value Health.* 2014;17(6):714-24.
82. McEwan P, Peters JR, Bergenheim K, Currie CJ. Evaluation of the costs and outcomes from changes in risk factors in type 2 diabetes using the Cardiff stochastic simulation cost-utility model (DiabForecaster). *Curr Med Res Opin.* 2006;22(1):121-9.
83. Willis M, Asseburg C, He J. Validation of economic and health outcomes simulation model of type 2 diabetes mellitus (ECHO-T2DM). *J Med Econ.* 2013;16(8):1007-21.
84. Ye W, Brandle M, Brown MB, Herman WH. The Michigan Model for Coronary Heart Disease in Type 2 Diabetes: Development and Validation. *Diabetes Technol Ther.* 2015;17(10):701-11.
85. Mount Hood Diabetes Challenge Network [April 2022]. Available from: <https://www.mthooddiabeteschallenge.com/>.
86. Nichols GA, Brown JB. Validating the Framingham Offspring Study equations for predicting incident diabetes mellitus. *Am J Manag Care.* 2008;14(9):574-80.
87. Davis WA, Peters KE, Makepeace A, Griffiths S, Bundell C, Grant SFA, et al. Prevalence of diabetes in Australia: insights from the Fremantle Diabetes Study Phase II. *Intern Med J.* 2018;48(7):803-9.

88. Hallgren Elfgren IM, Grodzinsky E, Törnvall E. The Swedish National Diabetes Register in clinical practice and evaluation in primary health care. *Prim Health Care Res Dev.* 2016;17(6):549-58.
89. Evans M, Khunti K, Mamdani M, Galbo-Jorgensen CB, Gundgaard J, Bogelund M, et al. Health-related quality of life associated with daytime and nocturnal hypoglycaemic events: a time trade-off survey in five countries. *Health Qual Life Outcomes.* 2013;11:90.
90. Harris S, Mamdani M, Galbo-Jorgensen CB, Bogelund M, Gundgaard J, Groleau D. The effect of hypoglycemia on health-related quality of life: Canadian results from a multinational time trade-off survey. *Can J Diabetes.* 2014;38(1):45-52.
91. Levy AR, Christensen TL, Johnson JA. Utility values for symptomatic non-severe hypoglycaemia elicited from persons with and without diabetes in Canada and the United Kingdom. *Health Qual Life Outcomes.* 2008;6:73.
92. Currie CJ, Morgan CL, Poole CD, Sharplin P, Lammert M, McEwan P. Multivariate models of health-related utility and the fear of hypoglycaemia in people with diabetes. *Curr Med Res Opin.* 2006;22(8):1523-34.
93. National Institute for Health and Care Excellence (NICE). Type 2 diabetes in adults: management. Appendix F: Full Health Economics Report 2015 [updated 2 December 2015; 12 March 2021]. Available from: [www.nice.org.uk/guidance/ng28](http://www.nice.org.uk/guidance/ng28).
94. Alvarez-Guisasola F, Yin DD, Nocea G, Qiu Y, Mavros P. Association of hypoglycemic symptoms with patients' rating of their health-related quality of life state: a cross sectional study. *Health Qual Life Outcomes.* 2010;8:86.
95. Bradley C, Eschwege E, de Pablos-Velasco P, Parhofer KG, Simon D, Vandenberghe H, et al. Predictors of Quality of Life and Other Patient-Reported Outcomes in the PANORAMA Multinational Study of People With Type 2 Diabetes. *Diabetes Care.* 2018;41(2):267-76.
96. Davis RE, Morrissey M, Peters JR, Wittrup-Jensen K, Kennedy-Martin T, Currie CJ. Impact of hypoglycaemia on quality of life and productivity in type 1 and type 2 diabetes. *Curr Med Res Opin.* 2005;21(9):1477-83.
97. Lundkvist J, Berne C, Bolinder B, Jonsson L. The economic and quality of life impact of hypoglycemia. *Eur J Health Econ.* 2005;6(3):197-202.
98. Marrett E, Radican L, Davies MJ, Zhang Q. Assessment of severity and frequency of self-reported hypoglycemia on quality of life in patients with type 2 diabetes treated with oral antihyperglycemic agents: A survey study. *BMC Res Notes.* 2011;4:251.
99. Marrett E, Stargardt T, Mavros P, Alexander CM. Patient-reported outcomes in a survey of patients treated with oral antihyperglycaemic medications: associations with hypoglycaemia and weight gain. *Diabetes Obes Metab.* 2009;11(12):1138-44.

100. McCoy RG, Van Houten HK, Ziegenfuss JY, Shah ND, Wermers RA, Smith SA. Self-report of hypoglycemia and health-related quality of life in patients with type 1 and type 2 diabetes. *Endocr Pract.* 2013;19(5):792-9.
101. Meneghini LF, Lee LK, Gupta S, Preblich R. Association of hypoglycaemia severity with clinical, patient-reported and economic outcomes in US patients with type 2 diabetes using basal insulin. *Diabetes Obes Metab.* 2018;20(5):1156-65.
102. Pagkalos E, Thanopoulou A, Sampanis C, Bousboulas S, Melidonis A, Tentolouris N, et al. The Real-Life Effectiveness and Care Patterns of Type 2 Diabetes Management in Greece. *Exp Clin Endocrinol Diabetes.* 2018;126(1):53-60.
103. Pratipanawatr T, Satirapoj B, Ongphiphadhanakul B, Suwanwalaikorn S, Nitiyanant W. Impact of Hypoglycemia on Health-Related Quality of Life among Type 2 Diabetes: A Cross-Sectional Study in Thailand. *J Diabetes Res.* 2019;2019:5903820.
104. Sheu WH, Ji LN, Nitiyanant W, Baik SH, Yin D, Mavros P, et al. Hypoglycemia is associated with increased worry and lower quality of life among patients with type 2 diabetes treated with oral antihyperglycemic agents in the Asia-Pacific region. *Diabetes Res Clin Pract.* 2012;96(2):141-8.
105. Shi L, Shao H, Zhao Y, Thomas NA. Is hypoglycemia fear independently associated with health-related quality of life? *Health Qual Life Outcomes.* 2014;12:167.
106. Solli O, Stavem K, Kristiansen IS. Health-related quality of life in diabetes: The associations of complications with EQ-5D scores. *Health Qual Life Outcomes.* 2010;8:18.
107. Tabaei BP, Shill-Novak J, Brandle M, Burke R, Kaplan RM, Herman WH. Glycemia and the quality of well-being in patients with diabetes. *Qual Life Res.* 2004;13(6):1153-61.
108. Takahara M, Katakami N, Shiraiwa T, Abe K, Ayame H, Ishimaru Y, et al. Evaluation of health utility values for diabetic complications, treatment regimens, glycemic control and other subjective symptoms in diabetic patients using the EQ-5D-5L. *Acta Diabetol.* 2019;56(3):309-19.
109. Terauchi Y, Ozaki A, Zhao X, Teoh C, Jaffe D, Tajima Y, et al. Humanistic and economic burden of cardiovascular disease related comorbidities and hypoglycaemia among patients with type 2 diabetes in Japan. *Diabetes Res Clin Pract.* 2019;149:115-25.
110. Vexiau P, Mavros P, Krishnarajah G, Lyu R, Yin D. Hypoglycaemia in patients with type 2 diabetes treated with a combination of metformin and sulphonylurea therapy in France. *Diabetes Obes Metab.* 2008;10 Suppl 1:16-24.
111. Wasem J, Bramlage P, Gitt AK, Binz C, Krekler M, Deeg E, et al. Co-morbidity but not dysglycaemia reduces quality of life in patients with type-2 diabetes treated with oral mono- or dual combination therapy--an analysis of the DiaRegis registry. *Cardiovasc Diabetol.* 2013;12:47.

112. Yfantopoulos J, Chantzaras A. Health-related quality of life and health utilities in insulin-treated type 2 diabetes: the impact of related comorbidities/complications. *Eur J Health Econ.* 2020;21(5):729-43.
113. Zhang Y, Wu J, Chen Y, Shi L. EQ-5D-3L Decrements by Diabetes Complications and Comorbidities in China. *Diabetes Ther.* 2020;11(4):939-50.
114. Leiter LA, Yale J, Chiasson J, Harris S, Kleinstiver P, editors. *Assessment of the Impact of Fear of Hypoglycemic Episodes on Glycemic and Hypoglycemia Management* 2005.
115. Brod M, Wolden M, Groleau D, Bushnell DM. Understanding the economic, daily functioning, and diabetes management burden of non-severe nocturnal hypoglycemic events in Canada: differences between type 1 and type 2. *J Med Econ.* 2014;17(1):11-20.
116. O'Reilly DJ, Burke N, Tarride JE, Hahn J, Nurkanovic L. Direct Health-Care Costs and Productivity Costs Associated With Hypoglycemia in Adults With Type 1 and Type 2 Diabetes Mellitus That Participated in the Canadian Hypoglycemia Assessment Tool Program. *Can J Diabetes.* 2018;42(6):659-63.
117. Araki A, Umegaki H, Sakurai T, Mizuno Y, Miyao M, Imori M, et al. Determinants and impact of physical impairment in patient-reported outcomes among older patients with type 2 diabetes mellitus in Japan. *Curr Med Res Opin.* 2021;37(3):393-402.
118. DiBonaventura M, Link C, Pollack MF, Wagner JS, Williams SA. The relationship between patient-reported tolerability issues with oral antidiabetic agents and work productivity among patients having type 2 diabetes. *J Occup Environ Med.* 2011;53(2):204-10.
119. Pawaskar M, Iglay K, Witt EA, Engel SS, Rajpathak S. Impact of the severity of hypoglycemia on health - Related quality of life, productivity, resource use, and costs among US patients with type 2 diabetes. *J Diabetes Complications.* 2018;32(5):451-7.
120. Pawaskar M, Witt EA, Engel SS, Rajpathak SN, Iglay K. Severity of hypoglycaemia and health-related quality of life, work productivity and healthcare costs in patients with type 2 diabetes in Europe. *Endocrinol Diabetes Metab.* 2018;1(2):e00011.
121. Rhoads GG, Orsini LS, Crown W, Wang S, Getahun D, Zhang Q. Contribution of hypoglycemia to medical care expenditures and short-term disability in employees with diabetes. *J Occup Environ Med.* 2005;47(5):447-52.
122. Canadian Agency for Drugs and Technologies in Health (CADTH). *Strings Attached: CADTH Search Filters Database* [Available from: <https://www.cadth.ca/finding-evidence-literature-searching-tools-support-systematic-reviews-0>].
123. Scottish Intercollegiate Guidelines Network. *SIGN Search Filters* [Available from: <https://www.sign.ac.uk/our-guidelines/>].
124. Lockwood C SR, Oh EG, . *Synthesizing quantitative evidence.* Wilkins LW, editor 2011.

125. Schaumberg DA, McDonald L, Shah S, Stokes M, Nordstrom BL, Ramagopalan SV. Evaluation of comparative effectiveness research: a practical tool. *J Comp Eff Res*. 2018;7(5):503-15.
126. Bell H, Wailoo, A.J., Hernandez, M., Grieve, R., Faria, R., Gibson, L., et al. The use of real world data for the estimation of treatment effects in NICE decision making [updated 2016 Dec 12; cited 2021 Jan 26]. Available from: <http://nicedsu.org.uk/wp-content/uploads/2018/05/RWD-DSU-REPORT-Updated-DECEMBER-2016.pdf>.
127. Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med*. 2015;12(10):e1001885.
128. Berger ML, Mamdani M, Atkins D, Johnson ML. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report--Part I. *Value Health*. 2009;12(8):1044-52.
129. Berger ML, Sox H, Willke RJ, Brixner DL, Eichler HG, Goettsch W, et al. Good Practices for Real-World Data Studies of Treatment and/or Comparative Effectiveness: Recommendations from the Joint ISPOR-ISPE Special Task Force on Real-World Evidence in Health Care Decision Making. *Value Health*. 2017;20(8):1003-8.
130. Cox E, Martin BC, Van Staa T, Garbe E, Siebert U, Johnson ML. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report--Part II. *Value Health*. 2009;12(8):1053-61.
131. Dreyer NA, Bryant A, Velentgas P. The GRACE Checklist: A Validated Assessment Tool for High Quality Observational Studies of Comparative Effectiveness. *J Manag Care Spec Pharm*. 2016;22(10):1107-13.
132. Dreyer NA, Schneeweiss S, McNeil BJ, Berger ML, Walker AM, Ollendorf DA, et al. GRACE principles: recognizing high-quality observational studies of comparative effectiveness. *Am J Manag Care*. 2010;16(6):467-71.
133. Duke Margolis Center for Health Policy. White Paper - Characterizing RWD Quality and Relevancy for Regulatory Purposes 2018 [Available from: <https://healthpolicy.duke.edu/sites/default/files/2020-08/Characterizing%20RWD%20for%20Regulatory%20Use.pdf>].
134. Duke Margolis Center for Health Policy. White Paper - A Framework for Regulatory Use of Real-World Evidence 2019 [Available from: [https://healthpolicy.duke.edu/sites/default/files/2020-08/rwe\\_white\\_paper\\_2017.09.06.pdf](https://healthpolicy.duke.edu/sites/default/files/2020-08/rwe_white_paper_2017.09.06.pdf)].

135. Duke Margolis Center for Health Policy. White Paper - Determining Real-World Data's Fitness for Use and the Role of Reliability 2019 [Available from: [https://healthpolicy.duke.edu/sites/default/files/2019-11/rwd\\_reliability.pdf](https://healthpolicy.duke.edu/sites/default/files/2019-11/rwd_reliability.pdf)].
136. Duke Margolis Center for Health Policy. White Paper - Adding Real-World Evidence to a Totality of Evidence Approach for Evaluating Marketed Product Effectiveness 2019 [Available from: <https://healthpolicy.duke.edu/sites/default/files/2020-08/Totality%20of%20Evidence%20Approach.pdf>].
137. Duke Margolis Center for Health Policy. White Paper - A Roadmap for Developing Study Endpoints in Real-World Settings 2020 [Available from: <https://healthpolicy.duke.edu/sites/default/files/2020-08/Real-World%20Endpoints.pdf>].
138. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). ENCePP Checklist for Study Protocols (Revision 4). EMA/540136/2009 2018 [Available from: [http://www.encepp.eu/standards\\_and\\_guidances/checkListProtocols.shtml](http://www.encepp.eu/standards_and_guidances/checkListProtocols.shtml)].
139. Faria R, Alava, M.H., Manca, A., Wailoo, A.J.,. NICE DSU Technical Support Document 17: The use of observational data to inform estimates of treatment effectiveness for Technology Appraisal: Methods for comparative individual patient data. 2015. 2015 [Available from: <http://www.nicedsu.org.uk>].
140. Food and Drug Administration. Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment 2005 [Available from: <https://www.fda.gov/files/drugs/published/Good-Pharmacovigilance-Practices-and-Pharmacoepidemiologic-Assessment-March-2005.pdf>].
141. Food and Drug Administration. Framework for FDA's Real-World Evidence Program 2018 [Available from: <https://www.fda.gov/media/120060/download>].
142. Food and Drug Administration. Use of Electronic Health Record Data in Clinical Investigations. Guidance for Industry 2018 [Available from: <https://www.fda.gov/media/97567/download>].
143. Food and Drug Administration (FDA). Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices. Guidance for Industry and Food and Drug Administration Staff 2017 [Available from: <https://www.fda.gov/media/99447/download>].
144. Gliklich RE LM, Dreyer NA (sr eds),. Registries for Evaluating Patient Outcomes: A User's Guide. 4th ed. (Prepared by L&M Policy Research, LLC, under Contract No. 290-2014-00004-C with partners OM1 and IQVIA) AHRQ Publication No. 19(20)- EHC020. Rockville, MD: Agency for Healthcare Research and Quality. <https://doi.org/10.23970/AHRQEPREGISTRIES4>. Accessed October 30 20202020.
145. Green Park Collaborative. RWE Decoder Tool - A Practical Tool for Assessing Relevance and Rigor of Real World Evidence: User's Guide 2017 [Available from: [http://www.cmtptnet.org/docs/resources/RWE\\_Decoder\\_Users\\_Guide.pdf](http://www.cmtptnet.org/docs/resources/RWE_Decoder_Users_Guide.pdf)].

146. International Society of Pharmacoepidemiology (ISPE). Guidelines for good pharmacoepidemiology practice (GPP). *Pharmacoepidemiol Drug Saf.* 25(1):2-10.
147. Johnson ML, Crown W, Martin BC, Dormuth CR, Siebert U. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report--Part III. *Value Health.* 2009;12(8):1062-73.
148. Motheral B, Brooks J, Clark MA, Crown WH, Davey P, Hutchins D, et al. A checklist for retrospective database studies--report of the ISPOR Task Force on Retrospective Databases. *Value Health.* 2003;6(2):90-7.
149. Patient-Centered Outcomes Research Institute (PCORI). Methodological Standards and Patient-Centeredness in Comparative Effectiveness Research: The PCORI Perspective. *JAMA.* 2012;307(15):1636-40.
150. Patient-Centered Outcomes Research Institute (PCORI). A RoadMap for State Policymakers to Use Comparative Effectiveness and Patient-Centered Outcomes Research to Inform Decision Making 2014 [Available from: [https://www.nashp.org/wp-content/uploads/2014/07/corected-roadmap-8\\_23\\_2018.pdf](https://www.nashp.org/wp-content/uploads/2014/07/corected-roadmap-8_23_2018.pdf)].
151. Patient-Centered Outcomes Research Institute (PCORI). Methodology Standards 2019 [Available from: <https://www.pcori.org/sites/default/files/PCORI-Methodology-Standards.pdf>].
152. Velentgas P, Dreyer, N.A., Nourjah, P., Smith, S.R., Torchia, M.M. (eds). Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. AHRQ Publication No. 12(13)-EHC099. Rockville, MD: Agency for Healthcare Research and Quality; 2013 [cited 2021 Jan 26]. Available from: [www.effectivehealthcare.ahrq.gov](http://www.effectivehealthcare.ahrq.gov).
153. von Elm E, Altman DG, Egger M, Pocock SJ, Gotsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med.* 2007;147(8):573-7.
154. Wang SV, Schneeweiss S, Berger ML, Brown J, de Vries F, Douglas I, et al. Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0. *Pharmacoepidemiol Drug Saf.* 2017;26(9):1018-32.
155. Bell H, Wailoo, A.J., Hernandez, M., Grieve, R., Faria, R., Gibson, L., et al. The use of real world data for the estimation of treatment effects in NICE decision making 2016 [Available from: <http://nicedsu.org.uk/wp-content/uploads/2018/05/RWD-DSU-REPORT-Updated-DECEMBER-2016.pdf>].
156. Brazier JE, Rowen D. NICE DSU Technical Support Document 11: Alternatives to EQ-5D for generating health state utility values. The National Institute for Health and Care Excellence (2011) [Available from: <http://www.nicedsu.org.uk>].



157. Kearns B, Ara R, Wailoo AG. A Review of the Use of Statistical Regression Models to Inform Cost Effectiveness Analyses within the NICE Technology Appraisals Programme: Report by the Decision Support Unit. The National Institute for Health and Care Excellence (2012) [Available from: [www.nicedsu.org.uk](http://www.nicedsu.org.uk)].
158. Papaioannou D, Brazier JE, Paisley S. NICE DSU Technical Support Document 9: The identification, review and synthesis of health state utility values from the literature. The National Institute for Health and Care Excellence (2011) [Available from: <http://www.nicedsu.org.uk>].
159. Rowen D, Brazier JE, Wong R, Wailoo AG. Measuring and valuing health-related quality of life when sufficient EQ-5D data is not available. The National Institute for Health and Care Excellence (2020) [Available from: <http://nicedsu.org.uk>].
160. Brazier J. Valuing health States for use in cost-effectiveness analysis. *Pharmacoeconomics*. 2008;26(9):769-79.
161. Brazier J, Ara R, Azzabi I, Busschbach J, Chevrou-Severac H, Crawford B, et al. Identification, Review, and Use of Health State Utilities in Cost-Effectiveness Models: An ISPOR Good Practices for Outcomes Research Task Force Report. *Value Health*. 2019;22(3):267-75.
162. Wolowacz SE, Briggs A, Belozeroff V, Clarke P, Doward L, Goeree R, et al. Estimating Health-State Utility for Economic Models in Clinical Studies: An ISPOR Good Research Practices Task Force Report. *Value Health*. 2016;19(6):704-19.
163. Devlin N, Parkin, D., Janssen, B. *Methods for Analysing and Reporting EQ-5D Data*: Springer; 2020 [Available from: <https://doi.org/10.1007/978-3-030-47622-9>].
164. Pullenayegum EM, Tarride JE, Xie F, Goeree R, Gerstein HC, O'Reilly D. Analysis of health utility data when some subjects attain the upper bound of 1: are Tobit and CLAD models appropriate? *Value Health*. 2010;13(4):487-94.
165. Alva M, Gray A, Mihaylova B, Clarke P. The effect of diabetes complications on health-related quality of life: the importance of longitudinal data to address patient heterogeneity. *Health Econ*. 2014;23(4):487-500.
166. Griffiths A, Paracha N, Davies A, Branscombe N, Cowie MR, Sculpher M. Analyzing Health-Related Quality of Life Data to Estimate Parameters for Cost-Effectiveness Models: An Example Using Longitudinal EQ-5D Data from the SHIFT Randomized Controlled Trial. *Adv Ther*. 2017;34(3):753-64.
167. Hunger M, Doring A, Holle R. Longitudinal beta regression models for analyzing health-related quality of life scores over time. *BMC Med Res Methodol*. 2012;12:144.
168. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol Drug Saf*. 2006;15(5):291-303.

169. McCandless LC, Gustafson P, Levy A. Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Stat Med.* 2007;26(11):2331-47.
170. Vanderweele TJ, Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology.* 2011;22(1):42-52.
171. McCandless LC, Gustafson P. A comparison of Bayesian and Monte Carlo sensitivity analysis for unmeasured confounding. *Stat Med.* 2017;36(18):2887-901.
172. Steenland K, Greenland S. Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. *Am J Epidemiol.* 2004;160(4):384-92.
173. Phillips CV. Quantifying and reporting uncertainty from systematic errors. *Epidemiology.* 2003;14(4):459-66.
174. Bray I. Lecture Notes: Epidemiology, Evidence-Based Medicine and Public Health, Lecture Notes, 6th Edition. *Journal of Public Health.* 2013;36(1):177-.
175. Salazar LF, Crosby RA, DiClemente RJ. *Research methods in health promotion* 2015.
176. Hunter RM, Baio G, Butt T, Morris S, Round J, Freemantle N. An educational review of the statistical issues in analysing utility data for cost-utility analysis. *Pharmacoeconomics.* 2015;33(4):355-66.
177. Leurent B, Gomes M, Carpenter JR. Missing data in trial-based cost-effectiveness analysis: An incomplete journey. *Health economics.* 2018;27(6):1024-40.
178. Brunetti M, Shemilt I, Pregno S, Vale L, Oxman AD, Lord J, et al. GRADE guidelines: 10. Considering resource use and rating the quality of economic evidence. *J Clin Epidemiol.* 2013;66(2):140-50.
179. Khunti K, Alsifri S, Aronson R, Cigrovski Berkovic M, Enters-Weijnen C, Forsen T, et al. Rates and predictors of hypoglycaemia in 27 585 people from 24 countries with insulin-treated type 1 and type 2 diabetes: the global HAT study. *Diabetes Obes Metab.* 2016;18(9):907-15.
180. Ratzki-Leewing A, Harris SB, Mequanint S, Reichert SM, Belle Brown J, Black JE, et al. Real-world crude incidence of hypoglycemia in adults with diabetes: Results of the InHypo-DM Study, Canada. *BMJ Open Diabetes Res Care.* 2018;6(1):e000503.
181. Goldstein H. *Longitudinal Studies and The Measurement of Change.* *Journal of the Royal Statistical Society Series D (The Statistician).* 1968;18(2):93-117.
182. Seaquist ER, Anderson J, Childs B, Cryer P, Dagogo-Jack S, Fish L, et al. Hypoglycemia and diabetes: a report of a workgroup of the American Diabetes Association and the Endocrine Society. *Diabetes Care.* 2013;36(5):1384-95.

183. Al Sayah F, Majumdar SR, Soprovich A, Wozniak L, Johnson ST, Qiu W, et al. The Alberta's Caring for Diabetes (ABCD) Study: Rationale, Design and Baseline Characteristics of a Prospective Cohort of Adults with Type 2 Diabetes. *Can J Diabetes*. 2015;39 Suppl 3:S113-9.
184. Radua J. PRISMA 2020 - An updated checklist for systematic reviews and meta-analyses. *Neurosci Biobehav Rev*. 2021;124:324-5.
185. Cumpston MS, McKenzie JE, Welch VA, Brennan SE. Strengthening systematic reviews in public health: guidance in the Cochrane Handbook for Systematic Reviews of Interventions, 2nd edition. *J Public Health (Oxf)*. 2022.
186. Khunti K, Caputo S, Damci T, Dzida GJ, Ji Q, Kaiser M, et al. The safety and efficacy of adding once-daily insulin detemir to oral hypoglycaemic agents in patients with type 2 diabetes in a clinical practice setting in 10 countries. *Diabetes Obes Metab*. 2012;14(12):1129-36.
187. Roumie CL, Min JY, Greevy RA, Grijalva CG, Hung AM, Liu X, et al. Risk of hypoglycemia following intensification of metformin treatment with insulin versus sulfonylurea. *Cmaj*. 2016;188(6):E104-e12.
188. Aronson R, Goldenberg R, Boras D, Skovgaard R, Bajaj H. The Canadian Hypoglycemia Assessment Tool Program: Insights Into Rates and Implications of Hypoglycemia From an Observational Study. *Can J Diabetes*. 2018;42(1):11-7.
189. International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Good Practices for Outcomes Research [cited 2021 Jan 26]. Available from: <https://www.ispor.org/heor-resources/good-practices-for-outcomes-research>.
190. Kristensen FB, Husereau D, Huic M, Drummond M, Berger ML, Bond K, et al. Identifying the Need for Good Practices in Health Technology Assessment: Summary of the ISPOR HTA Council Working Group Report on Good Practices in HTA. *Value Health*. 2019;22(1):13-20.
191. Ara R, Hill H, Lloyd A, Woods HB, Brazier J. Are Current Reporting Standards Used to Describe Health State Utilities in Cost-Effectiveness Models Satisfactory? *Value Health*. 2020;23(3):397-405.
192. Pullenayegum EM, Tarride JE, Xie F, O'Reilly D. Calculating utility decrements associated with an adverse event: marginal Tobit and CLAD coefficients should be used with caution. *Med Decis Making*. 2011;31(6):790-9.
193. Ara R, Brazier J, Young T. Recommended Methods for the Collection of Health State Utility Value Evidence in Clinical Studies. *Pharmacoeconomics*. 2017;35(Suppl 1):67-75.
194. Husereau D, Drummond M, Augustovski F, de Bekker-Grob E, Briggs AH, Carswell C, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. *BMJ*. 2022;376:e067975.

## ANNEXE 1. STRATÉGIE DE RECHERCHE EN DIABÈTE DE TYPE 2

---

### Stratégie de recherche pour les valeurs d'utilité en diabète de type 2

#### Ovid MEDLINE database

- 1 diabetes mellitus/
- 2 NIDDM.ti,ab.
- 3 MODY.ti,ab.
- 4 (late onset adj diabet\$).ti,ab.
- 5 (maturity onset adj diabet\$).ti,ab.
- 6 (non insulin\$ depend\$ or noninsulin\$ depend\$ or non insulin?depend\$ or noninsulin?depend\$).ti,ab.
- 7 (typ\$ 2 adj6 diabet\$).ti,ab.
- 8 (typ\$ II adj6 diabet\$).ti,ab.
- 9 T2DM.ti,ab.
- 10 glucose intoleran\$.ti,ab.
- 11 Diabetes Mellitus, Type 2/
- 12 or/1-11
- 13 hypoglycemia/
- 14 hypoglyc?emia.ti,ab.
- 15 (hypoglycaemi\$ or hypoglycemi\$).ti,ab.
- 16 or/13-15
- 17 (utilit\$ approach\$ or health gain or hui or hui2 or hui 2 or hui3 or hui 3).ti,ab.
- 18 (health measurement\$ scale\$ or health measurement\$ questionnaire\$).ti,ab.
- 19 (standard gamble\$ or categor\$ scal\$ or linear scal\$ or linear analog\$ or visual scal\$ or magnitude estimat\$).ti,ab.
- 20 (time trade off\$ or rosser\$ classif\$ or rosser\$ matrix or rosser\$ distress\$ or hrqol).ti,ab.
- 21 (index of wellbeing or quality of wellbeing or qwb).ti,ab.
- 22 (multiattribute\$ health ind\$ or multi attribute\$ health ind\$).ti,ab.
- 23 (health utilit\$ index or health utilit\$ indices).ti,ab.
- 24 (multiattribute\$ theor\$ or multi attribute\$ theor\$ or multiattribute\$ analys\$ or multi attribute\$ analys\$).ti,ab.
- 25 (health utilit\$ scale\$ or classification of illness state\$ or 15 dimension).ti,ab.
- 26 health state\$ utilit\$.ti,ab.
- 27 well year\$.ti,ab.
- 28 (multiattribute\$ utilit\$ or multi attribute\$ utilit\$).ti,ab.
- 29 health utilit\$ scale\$.ti,ab.
- 30 (euro qol or euro qual or eq-5d or eq5d or eq 5d or euroqol or euroqual).ti,ab.
- 31 (qualy or qaly or qualys or qalys or quality adjusted life year\$).ti,ab.

32 (hye or hyes or health\$ year\$ equivalent\$).ti,ab.

33 (person trade off\$ or person tradeoff\$ or time tradeoff\$ or time trade off\$).ti,ab.

34 theory utilit\$.ti,ab.

35 (sf 6d or short from 6d).ti,ab.

36 (sf6 or sf 6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six or shortform6 or short form6).ti,ab,kf.

37 Quality-Adjusted Life Years/  
38 quality adjusted life.ti,ab,kf.

39 (qaly\* or qald\* or qale\* or qtime\* or life year or life years).ti,ab,kf.

40 (health adj3 utilit\*).ti,ab,kf.

41 (utilit\* adj3 (valu\* or measur\* or health or life or estimat\* or elicit\* or disease or score\* or weight)).ti,ab,kf.

42 (preference\* adj3 (valu\* or measur\* or health or life or estimat\* or elicit\* or disease or score\* or instrument or instruments)).ti,ab,kf.

43 disutilit\*.ti,ab,kf.

44 standard gamble\*.ti,ab,kf.

45 tto.ti,ab,kf.

46 (hui or hui1 or hui2 or hui3).ti,ab,kf.

47 (eq or euroqol or euro qol or eq5d or eq 5d or euroqual or euro qual).ti,ab,kf.

48 rosser.ti,ab,kf.

49 (health\* adj2 year\* adj2 equivalent\*).ti,ab,kf.

50 (quality of wellbeing or quality of well being or index of wellbeing or index of well being or qwb).ti,ab,kf.

51 (time trade off or time tradeoff).ti,ab,kf.

52 or/17-51

53 (systematic review or meta-analysis).pt.

54 meta-analysis/ or systematic review/ or systematic reviews as topic/ or meta-analysis as topic/ or "meta analysis (topic)"/ or "systematic review (topic)"/ or exp technology assessment, biomedical/ or network meta-analysis/  
55 ((systematic\* adj3 (review\* or overview\*)) or (methodologic\* adj3 (review\* or overview\*))).ti,ab,kf,kw.

56 ((quantitative adj3 (review\* or overview\* or synthes\*)) or (research adj3 (integrati\* or overview\*))).ti,ab,kf,kw.

57 ((integrative adj3 (review\* or overview\*)) or (collaborative adj3 (review\* or overview\*)) or (pool\* adj3 analy\*)).ti,ab,kf,kw.

58 (data synthes\* or data extraction\* or data abstraction\*).ti,ab,kf,kw.

59 (handsearch\* or hand search\*).ti,ab,kf,kw.

60 (mantel haenszel or peto or der simonian or dersimonian or fixed effect\* or latin square\*).ti,ab,kf,kw.

61 (met analy\* or metanaly\* or technology assessment\* or HTA or HTAs or technology overview\* or technology appraisal\*).ti,ab,kf,kw.

62 (meta regression\* or metaregression\*).ti,ab,kf,kw.

63 (meta-analy\* or metaanaly\* or systematic review\* or biomedical technology assessment\* or bio-medical technology assessment\*).mp,hw.

64 (medline or cochrane or pubmed or medlars or embase or cinahl).ti,ab,hw.  
65 (cochrane or (health adj2 technology assessment) or evidence report).jw.  
66 (comparative adj3 (efficacy or effectiveness)).ti,ab,kf,kw.  
67 (outcomes research or relative effectiveness).ti,ab,kf,kw.  
68 ((indirect or indirect treatment or mixed-treatment or bayesian) adj3 comparison\*).ti,ab,kf,kw.  
69 (meta-analysis or systematic review).md.  
70 (multi\* adj3 treatment adj3 comparison\*).ti,ab,kf,kw.  
71 (mixed adj3 treatment adj3 (meta-analy\* or metaanaly\*)).ti,ab,kf,kw.  
72 umbrella review\*.ti,ab,kf,kw.  
73 (multi\* adj2 paramet\* adj2 evidence adj2 synthesis).ti,ab,kw,kf.  
74 (multiparamet\* adj2 evidence adj2 synthesis).ti,ab,kw,kf.  
75 (multi-paramet\* adj2 evidence adj2 synthesis).ti,ab,kw,kf.  
76 or/53-75  
77 12 and 16 and 52 and 76

### Ovid EMBASE database

1 non insulin dependent diabetes mellitus/  
2 (MODY or NIDDM).ti,ab.  
3 diabetes mellitus, Type II/  
4 insulin resistance/  
5 insulin\$ resistenc\$.ti,ab.  
6 (obes\$ adj diabet\$).ti,ab.  
7 impaired glucose toleranc\$.ti,ab.  
8 glucose intoleranc\$.ti,ab.  
9 (non insulin\$ depend\$ or noninsulin\$ depend\$ or noninsulin?depend\$ or non insulin?depend\$).ti,ab.  
10 ((typ\$ 2 or typ\$ II) adj diabet\$).ti,ab.  
11 ((adult\$ or matur\$ or late or slow or stabl\$) adj diabet\$).ti,ab.  
12 t2dm.ti,ab.  
13 or/1-12  
14 hypoglycemia/  
15 hypoglyc?emia.ti,ab.  
16 (hypoglycaemi\$ or hypoglycemi\$).ti,ab.  
17 or/14-16  
18 (utilit\$ approach\$ or health gain or hui or hui2 or hui 2 or hui3 or hui 3).ti,ab.  
19 (health measurement\$ scale\$ or health measurement\$ questionnaire\$).ti,ab.  
20 (standard gamble\$ or categor\$ scal\$ or linear scal\$ or linear analog\$ or visual scal\$ or magnitude  
estimat\$).ti,ab.  
21 (time trade off\$ or rosser\$ classif\$ or rosser\$ matrix or rosser\$ distress\$ or hrqol).ti,ab.

22 (index of wellbeing or quality of wellbeing or qwb).ti,ab.

23 (multiattribute\$ health ind\$ or multi attribute\$ health ind\$).ti,ab.

24 (health utilit\$ index or health utilit\$ indices).ti,ab.

25 (multiattribute\$ theor\$ or multi attribute\$ theor\$ or multiattribute\$ analys\$ or multi attribute\$ analys\$).ti,ab.

26 (health utilit\$ scale\$ or classification of illness state\$ or 15 dimension).ti,ab.

27 health state\$ utilit\$.ti,ab.

28 well year\$.ti,ab.

29 (multiattribute\$ utilit\$ or multi attribute\$ utilit\$).ti,ab.

30 health utilit\$ scale\$.ti,ab.

31 (euro qol or euro qual or eq-5d or eq5d or eq 5d or euroqol or euroqual).ti,ab.

32 (qaly or qaly or qualys or qalys or quality adjusted life year\$).ti,ab.

33 (hye or hyes or health\$ year\$ equivalent\$).ti,ab.

34 (person trade off\$ or person tradeoff\$ or time tradeoff\$ or time trade off\$).ti,ab.

35 theory utilit\$.ti,ab.

36 (sf 6d or short from 6d).ti,ab.

37 Quality-Adjusted Life Year/  
38 quality adjusted life.ti,ab,kw.

39 (qaly\* or qald\* or qale\* or qtime\* or life year or life years).ti,ab,kw.

40 (sf6 or sf 6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six or shortform6 or short form6).ti,ab,kw.

41 (health\* adj2 year\* adj2 equivalent\*).ti,ab,kw.

42 (health adj3 utilit\*).ti,ab,kw.

43 (utilit\* adj3 (valu\* or measur\* or health or life or estimat\* or elicit\* or disease or score\* or weight)).ti,ab,kw.

44 (preference\* adj3 (valu\* or measur\* or health or life or estimat\* or elicit\* or disease or score\* or instrument or instruments)).ti,ab,kw.

45 disutilit\*.ti,ab,kw.

46 rosser.ti,ab,kw.

47 standard gamble\*.ti,ab,kw.

48 tto.ti,ab,kw.

49 (time trade off or time tradeoff).ti,ab,kw.

50 (hui or hui1 or hui2 or hui3).ti,ab,kw.

51 (eq or euroqol or euro qol or eq5d or eq 5d or euroqual or euro qual).ti,ab,kw.

52 (quality of wellbeing or quality of well being or index of wellbeing or index of well being or qwb).ti,ab,kw.

53 or/18-52

54 (systematic review or meta-analysis).pt.  
55 meta-analysis/ or systematic review/ or systematic reviews as topic/ or meta-analysis as topic/ or "meta analysis (topic)"/ or "systematic review (topic)"/ or exp technology assessment, biomedical/ or network meta-analysis/  
56 ((systematic\* adj3 (review\* or overview\*)) or (methodologic\* adj3 (review\* or overview\*))).ti,ab,kf,kw.

57 ((quantitative adj3 (review\* or overview\* or synthes\*)) or (research adj3 (integrati\* or  
overview\*))).ti,ab,kf,kw.

58 ((integrative adj3 (review\* or overview\*)) or (collaborative adj3 (review\* or overview\*)) or (pool\*  
adj3 analy\*)).ti,ab,kf,kw.

59 (data synthes\* or data extraction\* or data abstraction\*).ti,ab,kf,kw.

60 (handsearch\* or hand search\*).ti,ab,kf,kw.

61 (mantel haenszel or peto or der simonian or dersimonian or fixed effect\* or latin  
square\*).ti,ab,kf,kw.

62 (met analy\* or metaanaly\* or technology assessment\* or HTA or HTAs or technology overview\* or  
technology appraisal\*).ti,ab,kf,kw.

63 (meta regression\* or metaregression\*).ti,ab,kf,kw.

64 (meta-analy\* or metaanaly\* or systematic review\* or biomedical technology assessment\* or bio-  
medical technology assessment\*).mp,hw.

65 (medline or cochrane or pubmed or medlars or embase or cinahl).ti,ab,hw.

66 (cochrane or (health adj2 technology assessment) or evidence report).jw.

67 (comparative adj3 (efficacy or effectiveness)).ti,ab,kf,kw.

68 (outcomes research or relative effectiveness).ti,ab,kf,kw.

69 ((indirect or indirect treatment or mixed-treatment or bayesian) adj3 comparison\*).ti,ab,kf,kw.

70 (meta-analysis or systematic review).md.

71 (multi\* adj3 treatment adj3 comparison\*).ti,ab,kf,kw.

72 (mixed adj3 treatment adj3 (meta-analy\* or metaanaly\*)).ti,ab,kf,kw.

73 umbrella review\*.ti,ab,kf,kw.

74 (multi\* adj2 paramet\* adj2 evidence adj2 synthesis).ti,ab,kw,kf.

75 (multiparamet\* adj2 evidence adj2 synthesis).ti,ab,kw,kf.

76 (multi-paramet\* adj2 evidence adj2 synthesis).ti,ab,kw,kf.

77 or/54-76

78 13 and 17 and 53 and 77



## Stratégie de recherche pour les valeurs de productivité en diabète de type 2

### Ovid MEDLINE database

1 diabetes mellitus/  
2 NIDDM.ti,ab.  
3 MODY.ti,ab.  
4 (late onset adj diabet\$.ti,ab.  
5 (maturity onset adj diabet\$.ti,ab.  
6 (non insulin\$ depend\$ or noninsulin\$ depend\$ or non insulin?depend\$ or  
noninsulin?depend\$.ti,ab.  
7 (typ\$ 2 adj6 diabet\$.ti,ab.  
8 (typ\$ II adj6 diabet\$.ti,ab.  
9 T2DM.ti,ab.  
10 glucose intoleran\$.ti,ab.  
11 Diabetes Mellitus, Type 2/  
12 or/1-11  
13 hypoglycemia/  
14 hypoglyc?emia.ti,ab.  
15 (hypoglycaemi\$ or hypoglycemi\$.ti,ab.  
16 or/13-15  
17 (productivity loss or productivity losses).mp.  
18 absenteeism.mp. or exp Absenteeism/  
19 presenteeism.mp. or exp Presenteeism/  
20 work productivity.mp.  
21 activity impairment.mp.  
22 (work loss or work\*loss).mp.  
23 work disability.mp.  
24 Workplace Activity Limitations Scale.mp.  
25 Work Limitations Questionnaire.mp.  
26 Work Ability Index.mp.  
27 Work Productivity Survey.mp.  
28 (Work Productivity and Activity Impairment).mp.  
29 WPAI.mp.  
30 Endicott Work Productivity Scale.mp.  
31 (Health and Labour Questionnaire).mp.  
32 (Health and Work Performance Questionnaire).mp.  
33 (Health and Work Questionnaire).mp.  
34 (Lam Employment Absence and Productivity Scale).mp.  
35 Sheehan Disability Scale.mp.

- 36 Stanford Presenteeism Scale.mp.
- 37 (Work and Health Interview).mp.
- 38 Worker Productivity Index.mp.
- 39 Work Productivity Short Inventory.mp.
- 40 (sick leave or medical leave).mp.
- 41 Work Capacity Evaluation.mp. or exp work capacity/
- 41 or/17-41
- 42 12 and 16 and 41

### Ovid EMBASE database

- 1 non insulin dependent diabetes mellitus/
- 2 (MODY or NIDDM).ti,ab.
- 3 diabetes mellitus, Type II/
- 4 insulin resistance/
- 5 insulin\$ resistenc\$.ti,ab.
- 6 (obes\$ adj diabet\$).ti,ab.
- 7 impaired glucose toleranc\$.ti,ab.
- 8 glucose intoleranc\$.ti,ab.
- 9 (non insulin\$ depend\$ or noninsulin\$ depend\$ or noninsulin?depend\$ or non  
insulin?depend\$).ti,ab.
- 10 ((typ\$ 2 or typ\$ II) adj diabet\$).ti,ab.
- 11 ((adult\$ or matur\$ or late or slow or stabl\$) adj diabet\$).ti,ab.
- 12 t2dm.ti,ab.
- 13 or/1-12
- 14 hypoglycemia/
- 15 hypoglyc?emia.ti,ab.
- 16 (hypoglycaemi\$ or hypoglycemi\$).ti,ab.
- 17 or/14-16
- 18 (productivity loss or productivity losses).mp.
- 19 absenteeism.mp. or exp Absenteeism/
- 20 presenteeism.mp. or exp Presenteeism/
- 21 work productivity.mp.
- 22 activity impairment.mp.
- 23 (work loss or work\*loss).mp.
- 24 work disability.mp.
- 25 Workplace Activity Limitations Scale.mp.
- 26 Work Limitations Questionnaire.mp.

27	Work Ability Index.mp.
28	Work Productivity Survey.mp.
29	(Work Productivity and Activity Impairment).mp.
30	WPAI.mp.
31	Endicott Work Productivity Scale.mp.
32	(Health and Labour Questionnaire).mp.
33	(Health and Work Performance Questionnaire).mp.
34	(Health and Work Questionnaire).mp.
35	(Lam Employment Absence and Productivity Scale).mp.
36	Sheehan Disability Scale.mp.
37	Stanford Presenteeism Scale.mp.
38	(Work and Health Interview).mp.
39	Worker Productivity Index.mp.
40	Work Productivity Short Inventory.mp.
41	(sick leave or medical leave).mp.
42	Work Capacity Evaluation.mp. or exp work capacity/
43	or/18-42
44	13 and 17 and 42