

HEC Montréal

Étude de la calibration des anticipations du risque d'infection à la COVID-19

Par

Xavier Dufour-Simard

**Sciences de la gestion
Spécialisation économie appliquée**

*Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences
(M. Sc.)*

Pierre-Carl Michaud
HEC Montréal
Directeur de recherche

Avril 2023

© Xavier Dufour-Simard, 2023

Résumé

Cette étude teste l'hypothèse de rationalité dans la formation des attentes appliquée au risque d'infection à la COVID-19. Nous profitons de données uniques, collectées sur 18 semaines auprès d'un échantillon représentatif de la population québécoise totalisant plus de 54 000 répondants. Chaque répondant a fourni une estimation probabiliste de son risque d'infection à la COVID-19 dans la semaine à venir. Nous estimons ensuite le risque objectif à l'aide des auto-déclarations d'infection élicitées par le questionnaire. Les deux mesures sont agrégées par groupe sociodémographique et comparées. Nous trouvons premièrement que sur l'échantillon entier, la tendance des anticipations de la population québécoise suit assez fidèlement l'évolution de la pandémie, suggérant qu'elle a mise à jour ses anticipations au moins sur une base hebdomadaire. La probabilité d'infection est toutefois considérablement surestimée, la perception moyenne du risque étant à 13.9% alors que le risque objectif moyen se situe à 3.9%. Nous trouvons que les biais (définis comme étant la différence entre la perception et la réalisation du risque) les plus importants sont observés chez les tranches d'âge les plus jeunes, les personnes vaccinées, les travailleurs du secteur de la santé et ceux du secteur de l'enseignement. Chacun de ces biais prend la forme d'une surestimation du risque. Nous trouvons que travailler dans le secteur de l'éducation est associé à une hausse de la perception du risque de 7 points de pourcentage au-dessus de celle d'employés travaillant au service à la clientèle, et ce alors que le risque objectif est sensiblement le même. Similairement, les répondants vaccinés perçoivent le risque d'infection comme étant plus élevé de 8 point de pourcentage à celui perçu par les non-vaccinés, et ce pour un risque objectif équivalent. Différentes stratégies sont employées pour explorer ces biais. Les résultats présentés permettent de rejeter l'hypothèse de rationalité dans la formation des attentes concernant le risque pandémique. L'étude conclue en présentant certaines implications des résultats et en recommandant différentes manières d'incorporer les apprentissages lors d'une prochaine pandémie.

Mots-clés

Recherche quantitative - COVID-19 – Biais de perception – Risque – Anticipations rationnelles – Probabilité subjective – Calibration

Table des matières

Résumé	III
Table des matières	IV
Liste des figures	VI
Liste des tables	VI
Remerciements	VII
1. Introduction	1
2. Revue de littérature	5
2.1 Risque et choix économiques	5
2.2 Anticipations rationnelles	6
2.3 Probabilité subjective	8
2.4 Calibration.....	9
2.5 Perception du risque en santé	10
2.6 COVID-19.....	10
3. Données	13
3.1 Introduction aux données.....	13
3.2 Exploration	16
3.2.1 Composition sociodémographique	16
3.2.2 Mesures de risque.....	18
3.2.3 Statut vaccinal.....	21
4. Méthodologie	24
4.1 Structure économétrique	24
4.2 Variables explicatives.....	27
4.3 Précisions économétriques.....	29
4.4 Incertitude épistémique.....	29
4.5 Groupe de référence : variables omises	31
5. Résultats	32
5.1 Comparaison entre MCO et probit	39
5.2 Analyse de sensibilité.....	42
5.3 Analyse des biais principaux	46
5.3.1 Biais associé au secteur de l'enseignement.....	46
5.3.2 Biais associé au statut vaccinal	51
5.3.3 Sources d'information.....	54
6. Discussion	59
6.1 Discussion sur le biais associé au secteur de l'enseignement	59

6.2 Discussion sur le biais associé au statut vaccinal.....	61
6.3 Mise en perspective des résultats avec la littérature	62
6.4 Discussion sur la pertinence économique	64
6.4.1 Modélisation conceptuelle	64
6.4.2 Implications.....	66
6.5 Limites.....	69
6.6 Pistes futures.....	70
Bibliographie	71

Liste des figures

Figure 3.1 : Distribution de la perception du risque pour l'échantillon non-pondéré	15
Figure 3.2 : Évolution de la composition de l'échantillon sur les 18 semaines à l'étude	17
Figure 3.3 : Évolution des deux mesures de risque sur les 18 semaines à l'étude.....	19
Figure 3.4 : Décomposition de la perception du risque par groupe d'âge	20
Figure 3.5 : Décomposition de la perception du risque conditionnelle à la présence d'un enfant	20
Figure 3.6 : Évolution du statut vaccinal sur les 18 semaines à l'étude	21
Figure 3.7 : Taux de vaccination selon l'occupation principale du répondant.....	22
Figure 3.8 : Taux de vaccination selon les différents groupes d'âge	23
Figure 5.1 : Comparaison de l'effet marginal des variables sociodémographiques sur le risque réel et perçu	34
Figure 5.2 : Comparaison de l'effet marginal de variables sociodémographiques sur le risque réel et perçu	36
Figure 5.3 : Comparaison de l'effet marginal de l'occupation principale sur le risque réel et perçu	37
Figure 5.4 : Comparaison de l'évolution du risque perçu entre les employés du secteur de l'éducation et l'échantillon entier.....	47
Figure 5.5 : Effet marginal capté par le secteur de l'enseignement sur la perception du risque hebdomadaire.....	49
Figure 5.6 : Effet marginal capté par le secteur de l'enseignement sur le risque réel hebdomadaire	50
Figure 5.7 : Estimation du biais de perception hebdomadaire capté par le secteur de l'enseignement... 51	
Figure 5.8 : Comparaison de probabilité subjective entre les vaccinés et non-vaccinés	52
Figure 5.9 : Effet marginal capté par le fait d'être vacciné sur la perception du risque hebdomadaire....	53
Figure 5.10 : Effet marginal capté par le fait d'être vacciné sur le risque objectif hebdomadaire	53
Figure 5.11 : Biais de perception hebdomadaire capté par le fait d'être vacciné.....	54
Figure 5.12 : Estimation hebdomadaire de l'évolution de la pandémie selon différentes sources d'information	57

Liste des tables

Table 3.1 : Statistiques descriptives de l'échantillon représentatif	14
Table 5.1 : Résultats des régressions sur la probabilité subjective et objective (en points de pourcentage)	38
Table 5.2 : Comparaison entre estimations Probit et MCO (en points de pourcentage).....	41
Table 5.3 : Résultats des régressions sur la probabilité subjective (en points de pourcentage)	44
Table 5.4 : Coefficients des différentes sources d'information lorsque régressées sur la probabilité subjective (points de pourcentage)	58

Remerciements

Je tiens à sincèrement remercier Pierre-Carl Michaud pour m'avoir proposé ce sujet de recherche et pour son encadrement. La pertinence de ses commentaires est entièrement garante de la qualité de ce mémoire.

1. Introduction

Le coronavirus a pour une première fois été recensé au Canada à Toronto le 25 janvier 2020, alors que le premier cas dans la province de Québec a été identifié le 27 février 2020¹. Seulement deux semaines plus tard, alors que l'Organisation mondiale de la santé qualifiait l'épidémie de pandémie, le virus fait sa première victime canadienne. Ce fut dans la même semaine que le gouvernement québécois a mis en place les premières mesures visant à contrôler la propagation du virus, faisant notamment de la province la première à fermer ses écoles². La distribution des vaccins au sein des provinces canadiennes a été entamée à la mi-décembre 2020.

Le 29 novembre 2021, presque un an plus tard, le premier cas du variant Omicron est identifié au Québec³. Les caractéristiques de ce variant font en sorte que le vaccin prévient l'apparition de symptômes graves, mais protège moins contre son infection (Abu-Raddad et al., 2022). Ce variant a été responsable de la plus importante vague de cas qu'a dû affronter la province, atteignant 250 000 personnes entre le 22 décembre 2021 et le 7 janvier 2022. Faisant face à une demande accrue pour le dépistage par test PCR⁴, le gouvernement provincial prend la décision le 5 janvier 2022 de réserver l'usage de ces tests aux personnes prioritaires. Cette décision a la conséquence de grandement complexifier l'estimation du nombre de cas confirmés de COVID-19 dans la province. Un projet est alors mis en place par le CIRANO afin de fournir des estimations de l'évolution du nombre de cas sur plusieurs semaines. Les résultats de la première semaine de collecte de données révèlent d'ailleurs que le nombre de cas entre le 13 et 18 janvier 2022 dans la province est de 5 à 8 fois plus élevé que ce que suggèrent les données restreintes provenant des tests PCR⁵.

Un an plus tard, en date du 28 février 2023, il est possible d'affirmer que le Québec a enregistré 1.3 millions d'infections confirmées depuis le début de la pandémie et près de 18 000 morts⁶. La plupart des

¹ The Canadian Press (2021, 24 janvier). « A timeline of COVID-19 in Canada », *The Star*. Récupéré le 28 février 2023 de <https://www.thestar.com/politics/2021/01/24/a-timeline-of-COVID-19-in-canada.html>

² Marin, Stéphanie (2021, 27 février). « Il y a un an, le premier cas de COVID-19 au Québec », *Le devoir*. Récupéré le 28 février 2023 de <https://www.ledevoir.com/societe/sante/596045/il-y-a-un-an-le-premier-cas-de-COVID-19-au-quebec>

³ Compilation Spéciale de l'INSPQ (2022, 5 octobre). « Ligne du temps COVID-19 au Québec », INSPQ. Récupéré le 21 mars 2023 de <https://www.inspq.qc.ca/COVID-19/donnees/ligne-du-temps>

⁴ Test d'analyse moléculaire par réaction en chaîne de la polymérase (*Polymerase Chain Reaction*)

⁵ Lacoursière, Ariane (2022, 21 janvier). « Au moins cinq fois plus élevé que rapporté », *La Presse*. Récupéré le 22 mars 2023 de <https://www.lapresse.ca/actualites/covid-19/2022-01-21/nombre-de-cas/au-moins-cinq-fois-plus-eleve-que-rapporte.php>

⁶ Gouvernement du Canada (2023). *COVID-19 epidemiology update: Current situation*, Canada. Récupéré le 28 février 2023 de https://health-infobase.canada.ca/COVID-19/current-situation.html?stat=num&measure=deaths_total&map=pt#a2

mesures restrictives ont été retirées, laissant entrevoir un retour de la vie à la normale. Ces statistiques témoignent de conséquences concrètes qu'a eues la pandémie sur la population, mais ne relatent pas ce qui fut sa conséquence la plus universelle, soit l'ajout d'un risque d'infection venant complexifier toute forme de prise de décision sur la période.

En effet, la perception du risque joue un rôle crucial dans les décisions économiques. En supposant qu'un individu détermine ses choix suivant une analyse coût-bénéfice, il est raisonnable de penser que l'individu ait mis à jour ses décisions en incorporant à son analyse une estimation du risque que lui présente la COVID-19, en particulier l'apparition d'une nouvelle vague. Cette perception du risque intervient alors dans une multitude de choix de l'individu, allant de la décision d'envoyer son enfant à la garderie à celle de faire son épicerie en ligne, d'aller travailler ou encore de se faire vacciner. Dans un tel contexte, il semble particulièrement important d'évaluer cette perception du risque afin de déterminer sa correspondance avec le risque objectif. Une surestimation pourrait amener un individu à prendre trop de précaution face au virus, alors qu'une sous-estimation pourrait l'amener à prendre trop de risques. Ces biais pourraient s'avérer coûteux en termes de bien-être *ex post*.

Étudier la perception du risque fournit également un outil aidant à juger de l'intervention gouvernementale durant la pandémie. En effet, il est généralement accepté qu'une perception excessive du risque peut mener à la présence de restrictions excessives (Viscusi, 1985). Beaucoup des mesures gouvernementales durant la pandémie ont visé à contrôler la propagation du virus en minimisant le contact humain, souvent au détriment de l'activité économique (Acemoglu et al., 2021). L'émergence généralisée du télétravail est un exemple de mesure mise en place afin de promouvoir la distanciation sociale. En plus de l'implication gouvernementale relative, beaucoup de travailleurs ont fait pression auprès de leur employeur pour gagner / préserver le droit de travailler à distance ou de manière hybride. Étudier la perception et la réalisation du risque pandémique peut être un outil d'analyse permettant l'évaluation de ces changements. Plutôt que de contraindre le comportement, d'autres interventions gouvernementales ont plutôt entrepris d'informer la population sur le virus, de manière à augmenter la compréhension du phénomène et ainsi améliorer la prise de décision au niveau individuel. Il est d'ailleurs maintenant possible d'affirmer que ces mesures furent très importantes car les changements volontaires furent responsables d'une part importante du contrôle de la pandémie (Fernández-Villaverde et Jones, 2020). Ces exemples témoignent de l'importance de la perception du risque d'infection dans un contexte pandémique. Ce point est exacerbé par le fait que la contribution de l'économie aux modèles épidémiologiques (SIR, SIRS) est généralement construite autour de l'ajout de variables de décision pour

l'agent (par exemple le nombre de contacts qu'il décide d'entretenir; Aadland, Finnoff et Huang, 2011), assumant alors souvent que ce dernier forme ses anticipations de manière rationnelle.

Il convient donc d'évaluer la perception du risque de la population, ce qui peut être fait en comparant les anticipations à des anticipations formées rationnellement. L'hypothèse d'anticipations rationnelles, premièrement avancée par Muth (1961), entend que les agents se forment des attentes reflétant leur risque objectif. Le rejet de cette hypothèse implique plutôt un biais dans l'estimation du risque.

Ce mémoire utilise des données uniques représentatives de la population québécoise adulte, récoltées auprès de 3000 à 3027 répondants par semaine sur 18 semaines. Les données ont été récoltées du 13 janvier au 17 mai 2022, capturant notamment la vague Omicron. Plus de 54 000 répondants ont eu à fournir une estimation probabiliste du risque d'infection que leur présentait la COVID-19 dans la semaine à venir. Cette dernière peut ensuite être comparée à une mesure du risque objectif, estimée à l'aide des répondants de l'échantillon étant ou ayant été infectés par le virus dans la semaine précédente. Ce mémoire effectue alors un exercice de calibration, défini par Manski (2004) comme l'évaluation de la validité d'une prédiction probabiliste, réalisée par la comparaison à un taux objectif. Le mémoire vise à savoir s'il existe une différence entre les deux mesures de risque recensées, le cas échéant permettant de rejeter l'hypothèse de rationalité dans la formation des attentes. Un deuxième effort s'intéresse à découvrir chez quels groupes socio-économiques trouve-t-on les biais les plus importants.

Il est premièrement trouvé que sur l'échantillon entier, la perception du risque semble suivre temporellement la tendance de l'évolution de la pandémie (et donc du risque réel), signifiant que la population a mise à jour son estimation du risque au moins sur une base hebdomadaire. Bien que la tendance soit similaire, le risque reste largement surestimé à chaque semaine. Ce résultat est en ligne avec la tendance généralement observée des répondants à surestimer les probabilités à faible occurrence.

La comparaison des risques est ensuite approfondie en évaluant les différences de risque subjectif et objectif entre groupes sociodémographiques. Nous définissons alors le concept de biais de perception, utilisé pour identifier les caractéristiques sociodémographiques relatant une variation significativement différente sur la perception subjective du risque et sur le risque objectif.

Plusieurs biais de perception sont identifiés par ce mémoire, les principaux étant relatifs à l'âge, aux secteurs d'emploi de la santé et de l'enseignement et au statut vaccinal. Afin de mettre leur amplitude en perspective, soulignons que la probabilité moyenne d'infection (le risque objectif) sur l'échantillon est de 3.9%. Il est premièrement trouvé que tous les groupes d'âge sont biaisés dans leur estimation du risque

d'infection (risque subjectif), mais que le biais est particulièrement marqué chez les plus jeunes, où il se manifeste en une surestimation du risque. Similairement, les travailleurs des secteurs de la santé et de l'enseignement entretenaient des biais de perception considérables sur la période de collecte, prenant la forme de surestimations du risque. Pour un risque réel non-différent du groupe contrefactuel, nous recensons dans ces deux secteurs d'emploi une augmentation de la perception du risque de 3.7 et 7 points de pourcentage. Le biais capté par le statut vaccinal prend également la forme d'une surestimation très marquée du risque d'infection que présente le virus, les répondants vaccinés percevant un risque d'infection plus élevé par 8 points de pourcentage à celui perçu par les non-vaccinés. Ce résultat suggère que le choix de certains de ne pas se faire vacciner pourrait être la conséquence d'une faible perception du risque d'infection.

La suite de ce mémoire est divisée en cinq chapitres. Le chapitre 2 aborde la littérature relative aux anticipations rationnelles, à la perception du risque et à la COVID-19. Le troisième chapitre présente plus formellement les données utilisées, suivie du quatrième qui structure la méthodologie employée par le mémoire. Le chapitre 5 présente les résultats. Sont premièrement présentés les résultats généraux, suivis de sous-sections analysant respectivement les biais du secteur de l'enseignement et du vaccin. Le chapitre 6 sert de discussion. Il aborde premièrement les implications concrètes qu'ont pu avoir les biais de perception propres au secteur de l'enseignement et au statut vaccinal. Une courte structuration du problème que pose une pandémie pour une société est ensuite présentée, permettant de comprendre le contexte dans lequel viennent s'imbriquer les biais de perception recensés. Ceci permet également de mettre en perspective la responsabilité du gouvernement face à ces biais, notamment en ce qui a trait à l'externalité de contagion.

2. Revue de littérature

Le risque joue un rôle central dans un grand nombre de décisions économiques. Nous proposons d'abord de voir comment le risque affecte ces décisions. Ensuite, nous abordons la définition des anticipations rationnelles et la méthodologie de calibration pour tester cette hypothèse. Nous introduisons par la suite la notion de risque subjectif en survolant les applications en santé et relatives à la COVID-19.

2.1 Risque et choix économiques

En sciences économiques, le modèle canonique de décision en présence de risque est celui de Von Neumann et Morgenstern (1947), soit la théorie de l'espérance de l'utilité. Supposons deux états de la nature s , tel que $s \in \{0,1\}$. La réalisation 0 représente l'état où l'individu n'est pas infecté de la COVID-19 alors que l'état $s = 1$ représente l'état où l'individu est infecté. L'utilité dans chacun de ces états est dénotée par U_s tq $s \in \{0,1\}$. Le risque que l'état 1 survienne est dénoté par p . Ceci représente le risque objectif, soit le risque que $s = 1$ se réalise. Supposons maintenant que l'agent peut entreprendre une action e qui change l'utilité dans chacun des états. Nous avons alors $U_s(e)$ tq $s \in \{0,1\}$. Si les préférences pour le risque d'un individu respectent certains axiomes, Von Neumann et Morgenstern montrent que le choix optimal de l'individu peut être représenté en utilisant l'espérance de l'utilité :

$$E[U_s] = \max_e (1 - p)U_0(e) + pU_1(e) \quad (1)$$

Il est possible que l'agent forme ses propres anticipations en lien avec le risque p . Le travail de Karni (2005) reprenant le *subjective expected utility model* de Savage (1954) explique le sens double que peut prendre la notion de probabilité. Dans son sens subjectif (p^*), elle réfère à « une mesure du degré de croyance d'un décideur dans la réalisation probable d'un événement », alors que dans son sens objectif (p) elle réfère à « la fréquence relative d'un résultat aléatoire dans des essais répétés »⁷. Savage (1954) propose donc la généralisation de l'espérance de l'utilité en supposant que l'agent utilise ses propres croyances (sa perception subjective du risque) afin de prendre ses décisions. Si p^* est la croyance relative au risque p , alors les préférences de l'agent sont plutôt données par :

$$E[U_s] = \max_e (1 - p^*)U_0(e) + p^*U_1(e) \quad (2)$$

⁷ Traductions libres

Il est donc possible que p^* soit différent de p , tant au niveau individuel qu'au niveau d'un groupe. Ceci mène potentiellement à des choix optimaux de e qui sont différents selon la perception du risque.

Notons finalement une variante de ce problème qui met davantage en lumière l'effet de la perception du risque sur les décisions optimales. Supposons que e mesure un niveau de prévention qui réduit p (Eeckhoudt, Gollier et Schlesinger, 2004). Nous pouvons alors supposer $p(e)$ et $p^*(e)$, avec $p'(e) < 0$. Ce modèle de prévention s'applique bien au contexte de la COVID-19, où e pourrait représenter la décision de porter un masque ou encore de pratiquer la distanciation sociale. L'individu maximise alors son utilité en choisissant e , résolvant le problème suivant :

$$E[U_s] = \max_e (1 - p^*(e))U_0 + p^*(e)U_1 \quad (3)$$

Il s'agit d'un problème de prévention bien commun en science économique, qui illustre comment le risque (ou l'anticipation d'un risque) est un facteur important de la prise de décision.

2.2 Anticipations rationnelles

Une manière d'estimer la validité d'anticipations est de les comparer à des attentes formées rationnellement. Muth (1961) définit ce concept en posant l'hypothèse que les anticipations (probabilités subjectives) tendent à être distribuées suivant la prédiction de la théorie (la probabilité objective). Les anticipations au niveau individuel pourraient être erronées, mais l'espérance de ces dernières reflèterait la tendance que suit la probabilité objective. L'hypothèse se doit alors d'être testée au niveau d'un groupe d'individus ayant des probabilités objectives similaires.

Il devrait être souligné que l'hypothèse de rationalité dans la formation des attentes a déjà été remise en question maintes fois par la littérature. Lovell (1986) aborde plusieurs études dont les résultats convergent vers un rejet statistique de l'hypothèse. Il est par exemple trouvé que certaines entreprises échouent à utiliser adéquatement les données de leurs ventes passées dans la prédiction des ventes futures (Hirsch et al., 1969), ou encore qu'elles omettent d'utiliser pleinement les informations sur la croissance de la masse monétaire dans leurs anticipations d'inflation (De Leeuw, 1981 et McKelvey, 1984).

L'utilisation de l'hypothèse d'anticipations rationnelles rend certaines réalités difficiles à expliquer théoriquement, menant à l'émergence d'énigmes. Une telle situation apparaît lorsqu'on considère le fait que très peu de personnes participent aux marchés financiers, et ce malgré des rendements historiquement intéressants. L'hypothèse d'anticipations rationnelles répond difficilement à cette réalité,

suggérant alors comme explication une très forte aversion au risque. À l'aide d'estimations subjectives, Hurd (2009) suggère une réponse au *equity holding puzzle* en trouvant que les anticipations pour le rendement annuel du *Dow Jones Industrial Average* sont en moyenne beaucoup plus pessimistes que les rendements historiques, au point où l'individu médian (du *Nederlandsche Bank Household Survey*) anticipe un rendement supérieur et moins volatile avec un compte épargne. Ce résultat témoigne du potentiel explicatif des données subjectives, qui se trouvent ici à répondre aux lacunes de l'hypothèse d'anticipations rationnelles.

Une autre limite souvent attribuée à l'hypothèse est qu'elle entend un accès gratuit et instantané à l'information. Un article de Carroll (2001) suggère une modélisation alternative des anticipations, construite autour d'un modèle épidémiologique SIR sans rétablissement (*recovery*), où l'infection représente la transmission de l'information. Il obtient que sa modélisation mène à une meilleure représentation des anticipations de la population relatives à l'inflation et au chômage que l'hypothèse de rationalité, et ce parce que le modèle permet un certain délai pour la transmission de l'information.

Prelec (1998) contribue à cette discussion en répondant théoriquement au problème de l'interprétation des probabilités. Cet auteur présente une alternative au modèle de l'espérance de l'utilité, qui vise à y intégrer une composante nuanciant l'hypothèse sous-jacente qui dicte que les probabilités sont interprétées linéairement. Ce dernier approfondit la notion de *probability weighting function* (PWF), premièrement avancée en 1979 par la *prospect theory* de Kahneman et Tversky. La PWF prévoit la conversion d'une probabilité en un poids décisionnel, qui est ensuite utilisé par l'individu dans sa prise de décision. Le bénéfice de cette conversion est que son processus peut être modélisé, résultant en une fonction représentant de manière plus réaliste la considération du risque par l'individu dans sa prise de décision. Une différence importante de la PWF est relative à ses propriétés. Plutôt que d'être concave, la fonction est régressive, asymétrique, réflexive et en forme de S inversé. Une implication importante dans notre contexte ressort du caractère régressif de cette nouvelle fonction, qui implique que la considération d'une probabilité qui se trouve dans le premier arc du S (généralement pour les probabilités p tel que $p < \frac{1}{e}$, où e correspond au nombre d'Euler) est interprétée différemment d'une probabilité s'y trouvant au-dessus. Plus spécifiquement, cette modélisation implique une surpondération (*overweighting*) des probabilités à faible occurrence contre une sous-pondération des plus hautes probabilités. Cette surpondération mène à une aversion particulièrement marquée aux risques à faible occurrence. L'espérance de la probabilité d'infection à la COVID-19 se situe fort probablement sous le seuil de $1/e$, faisant de cette dernière une probabilité à faible occurrence. Cette structure théorique alternative

suggère alors la possibilité que les répondants aient surestimé le risque hebdomadaire d'infection au virus.

Ces exemples servent à montrer que l'hypothèse d'attentes rationnelles pose d'importantes contraintes sur les capacités analytiques de la population, résultant en des prédictions théoriques parfois défailtantes. Il est toutefois important de tester cette hypothèse dans de multiples contextes. Certaines études se sont intéressées à la calibration des anticipations en lien avec la santé, mais, à notre connaissance, aucune étude ne s'est encore intéressée à la rationalité des anticipations relatives au risque pandémique. L'exercice effectué par cette étude est alors un ajout intéressant à la littérature.

2.3 Probabilité subjective

L'utilisation d'estimations subjectives, bien que souvent critiquée en économie, est une approche méthodologique présentant plusieurs avantages (Manski, 2004). Parmi ces derniers, le plus pertinent dans notre contexte est certainement la possibilité d'effectuer ce que l'auteur définit comme une étude de calibration (*calibration study*), qui consiste à estimer la validité d'une prédiction probabiliste en la comparant à un taux objectif. En juxtaposant les probabilités subjectives et objectives, il nous est possible de tester l'hypothèse d'anticipations rationnelles, ici appliqué au contexte de la pandémie de COVID-19. En décomposant ensuite ces deux mesures de risque parmi différentes sous-populations, il devient possible d'évaluer à quel point chacun des groupes a été précis dans son estimation.

Il est important de noter que l'agent agit selon sa perception du risque (subjectif), mais subit les conséquences que ses actions ont sur le risque objectif. Si l'agent est mal calibré, c'est-à-dire qu'il existe une différence entre son estimation du risque et le risque objectif auquel il s'expose, il se retrouve à encaisser des retombées excessives nuisant à son bien-être. Ceci implique que l'agent tente de minimiser la différence entre le risque qu'il perçoit et le risque auquel il s'expose réellement. Minimiser cette différence lui permet d'agir en étant pleinement conscient des conséquences de ses actions, évitant ainsi de prendre trop de mesures préventives ou au contraire de tomber malade alors qu'il aurait pu l'éviter.

Un point pertinent relatif à l'utilisation de données probabilistes subjectives abordé par Manski et Molinari (2010) est le fait que plusieurs répondants arrondissent leur réponse. Ce problème entraîne une perte de précision, particulièrement marquée dans les estimations de probabilités à faibles occurrences. Les auteurs réussissent à montrer que différentes personnes arrondissent différemment leurs réponses, de manière relativement constante entre les questions. Ils proposent alors de répondre à ce problème empirique en déduisant le type d'arrondissement (*rounding pattern*) employé par chaque répondant, et

en déduisant ensuite l'intervalle de probabilité qu'il camoufle. L'intervalle de probabilité pourrait ensuite être substitué au chiffre arrondi dans l'analyse des données. Bien que cette méthodologie soit intéressante car permettant d'augmenter la fiabilité des estimations probabilistes subjectives, elle requiert une quantité significatives d'estimations par répondant. Les données collectées par le CIRANO n'élucide qu'une seule estimation de probabilité par répondant, ce qui ne nous permet pas de compenser pour l'arrondissement potentiel des réponses. Un autre problème associé à l'usage des probabilités subjectives concerne l'incertitude épistémique. La discussion répondant au problème est présentée à la section 4.4.

2.4 Calibration

Un article de Combs et al. (1978) étudie la calibration de répondants à différentes causes de mortalité. Les participants ont eu à fournir différentes estimations relatives à des couples aléatoires de causes de mortalité, notamment laquelle des deux causes est la plus meurtrière et par quelle proportion. En plus de conclure de manière générale sur une mauvaise calibration des répondants, les auteurs réussissent à établir un biais présent dans l'estimation de la probabilité d'événements à faible occurrence prenant le sens d'une surestimation du risque. Cette tendance à la surestimation du risque est également un élément présent dans le travail de Viscusi (1985).

Le travail de Fischhoff, Lichtenstein et Phillips (1976) aborde plusieurs exercices de calibration présents dans la littérature. Ils trouvent que de manière générale, les répondants sont mal calibrés, et parallèlement trop confiants en leur capacité d'évaluer un risque. Les tests de modèle de Brenner (2003) confirment cet excès de confiance, et abordent le *hard-easy effect*, qui rappelle que le degré d'excès de confiance est généralement plus marqué pour une estimation difficile qu'une estimation facile. Fischhoff et ses collaborateurs suggèrent que l'entraînement peut mener à l'amélioration de la calibration des attentes, quoique peu d'études s'y soient intéressées. Notons également que les auteurs expliquent que pour beaucoup des décisions, une calibration légèrement erronée n'entraîne qu'une faible perte d'utilité. Ils citent toutefois l'exemple de la santé comme étant un domaine où une relativement faible erreur de calibration peut avoir des conséquences très importante sur le bien-être de l'individu.

Le travail de Keren (1987) considère une nuance intéressante, soit l'expérience des répondants avec le risque en question. Les données qu'il utilise ont été récoltées dans deux tournois de Bridge, un regroupant des joueurs experts et l'autre des joueurs amateurs, où ces derniers ont eu à estimer la probabilité que le contrat accepté durant la phase de paris soit réalisé. L'auteur obtient que les joueurs experts sont beaucoup mieux calibrés que les joueurs amateurs, ces derniers étant notamment plus susceptibles

d'estimer les probabilités avec un excès de confiance. Ce résultat suggère qu'il est possible d'entraîner une population afin d'améliorer sa calibration.

2.5 Perception du risque en santé

L'article de Ferrer et Klein (2015) résume certains apports récents à la littérature sur la perception du risque et les comportements en matière de santé. Il est premièrement abordé que l'expérience avec une maladie (soit en l'ayant ou en connaissant quelqu'un étant affecté) affecte à la hausse la perception du risque associée. Les auteurs expliquent également que pour évaluer un risque, de l'information extérieure doit être comprise et utilisée. Ceci fait en sorte que les personnes confortables avec les chiffres (*highly numerate individuals*) sont moins susceptibles d'entretenir des biais dans leur estimation de risque. Un biais particulièrement présent en santé est un biais optimiste (*optimism bias*), stipulant que l'individu se considère personnellement moins à risque qu'autrui. Il est montré que ce type de biais peut mener à une réduction des comportements préventifs.

L'étude de Borrelli et al. (2010) arrive à une conclusion similaire en évaluant la perception du risque de fumer chez des fumeurs malades. Ils obtiennent qu'avoir une plus haute perception du risque augmente les chances d'arrêter de fumer. Au contraire, avoir une plus faible perception du risque est associé avec un biais optimiste plus important. Les résultats suggèrent que venir augmenter la perception du risque (*perceived vulnerability*) pourrait diminuer le biais optimiste et encourager l'arrêt de consommation de tabac.

Ropeik (2012) discute des dangers d'entretenir un biais de perception (*perception gap*), un point bien illustré par l'exemple précédent du tabagisme. L'auteur explique que ce biais entraîne trois types de conséquences nocives : des comportements personnels dangereux (pour soi et/ou autrui), un stress chronique (entraînant ses propres ramifications nuisibles) et des politiques publiques sous-optimales. La conclusion de l'auteur souligne le besoin de discerner les dangers qu'entraînent les biais de perception de risque. Elle reconnaît l'importance de dépasser le cadre des anticipations rationnelles, de manière à modéliser la perception du risque de manière réaliste. Toute perception du risque est par nature subjective, et donc intrinsèquement ouverte aux biais. L'auteur défend que, afin qu'ils soient résolus, ces biais doivent être extensivement étudiés plutôt qu'ignorés par la théorie.

2.6 COVID-19

Malgré le fait que la pandémie de COVID-19 soit un événement très récent, plusieurs études s'y sont déjà attardées. Plusieurs d'entre elles mettent à profit des données récoltées au début de 2020, le travail de

Leigh et al. (2022) en étant un bon exemple. Les auteurs se sont intéressés à la perception, aux comportements et aux types d'information consultés en lien avec la pandémie en se basant sur les résultats d'une enquête réalisée au Canada au printemps 2020, seulement deux mois après le premier cas recensé au Québec. Ils obtiennent que la pandémie affecte différemment la population. Par exemple, ils montrent que les personnes plus jeunes et les femmes perçoivent une dégradation plus importante de leur état de santé (décomposée en 4 mesures : mentale, physique, spirituelle et sociale) que les personnes plus âgées et les hommes. Cette relation a été approfondie par Han et al. (2021), qui trouvent qu'une plus haute perception du risque relatif à la COVID-19 est associée avec plus d'émotions négatives, et est conséquemment inversement reliée à la santé mentale.

Les résultats de Leigh et al. montrent également que les personnes âgées et celles à plus haut revenu avaient une plus haute intention d'aller se faire vacciner que leur contrepartie respective. De plus, les auteurs identifient des différences significatives pertinentes quant à la recherche d'information. Ils obtiennent que dans leur recherche d'information en lien avec la pandémie, les personnes moins éduquées considèrent moins de sources d'information distinctes, et sont plus propices à se fier à des commentaires de leurs familles et amis sur les réseaux sociaux. Au contraire, les gens plus éduqués sont susceptibles de vérifier l'information à l'aide de plusieurs sources qu'ils jugent comme crédibles. Ces observations montrent que déjà en début de pandémie, il semblait y avoir une certaine hétérogénéité dans la population quant aux perceptions relatives à la maladie, la vaccination et la recherche d'information.

Joslyn, Sinatra et Morrow (2021) effectuent une revue de littérature sur les thèmes de la perception de risque, la prise de décision en temps de risque et la communication du risque, le tout appliqué au contexte de la pandémie. Quelques points particulièrement intéressants pour notre contexte ressortent. Le premier est qu'à travers les travaux de Garfin et al. (2021) et de Reinoltz et al. (2021), nous savons qu'une plus forte perception du risque est associée avec une plus faible exposition aux situations potentiellement dangereuses (les exemples de manger en salle dans un restaurant et de pratiquer la distanciation sociale sont utilisés). Ainsi, la perception du risque est une variable d'intérêt qui prédit le comportement. Toutefois, Luckman et al. (2021) démontre qu'accomplir une action préventive (telle que porter un masque) peut réduire l'intention d'agir de manière sécuritaire par la suite. Il y aurait alors des retombées négatives (*negative spillover*) liées aux actes de prévention.

Plusieurs études ont par la suite identifié l'importance de la structure de l'information dans la perception du message, soulignant ici la vulnérabilité de l'interprétation humaine. Par exemple, Reinholtz et al. (2021)

montre qu'en temps de déclin dans le nombre de cas, montrer un graphique du nombre de cas cumulatif (avec une baisse de la hausse) mène à une perception du risque plus élevée que montrer un graphique du nombre de nouveaux cas (décroissant). Il a ensuite été montré que les gens exposés au deuxième graphique avaient l'intention d'agir de manière plus risquée. Similairement, utiliser l'expression « symptômes légers » versus explicitement nommer ces symptômes légers mène à une diminution de l'intention de compliance avec les mesures gouvernementales suggérées (Kostopoulou et Schwartz, 2021). Ces biais de compréhension sont utilisés par les auteurs pour souligner la susceptibilité de la population à la forme que prend le message.

Une autre étude, cette fois-ci de Garfin et al. (2021), a utilisé des données récoltées en mars et avril 2020 pour évaluer la perception du risque à travers un échantillon représentatif de la population américaine. Ils obtiennent que les répondants, de manière générale, surestiment hautement leur chance de mourir de la COVID-19. Les auteurs réussissent aussi à isoler un biais optimiste, les répondants estimant qu'ils ont en moyenne moins de chance d'être infectés par la COVID-19 que la personne moyenne. Un autre fait est trouvé, soit qu'être exposé à la COVID-19 (soit en l'attrapant ou en connaissant quelqu'un l'ayant attrapé) influence de manière significative les estimations de susceptibilité à l'infection. Ceci est utilisé par les auteurs comme argument démontrant que la population forme ses attentes au moins en partie selon son environnement et ses expériences.

Bhuiya et al. (2021) commencent leur étude en rappelant que la perception du risque est un prédicteur important de la décision personnelle d'entreprendre des mesures préventives. Les auteurs s'attardent ensuite à étudier l'existence de biais de perception. Les biais relatés dans cette étude concernent l'accord ou le désaccord du répondant avec certaines affirmations, telle que « la grippe saisonnière est tout aussi dangereuse que le coronavirus ». Ils progressent ensuite à montrer que, sur un échantillon combinant les États-Unis, le Canada et le Royaume-Uni, un faible niveau d'éducation et un faible revenu prédisent l'existence de biais de perception de ce type. La présence de ce biais laisse supposer que les personnes moins éduquées ont généralement une moins bonne compréhension du phénomène entourant la COVID-19 et du risque s'y rattachant.

3. Données

3.1 Introduction aux données

Les données utilisées dans ce mémoire ont été collectées par le Centre interuniversitaire de recherche en analyse des organisations (CIRANO)⁸. La collecte de données s'est faite en ligne par le biais de panel de *Léger Opinion*. L'enquête a été réalisée sur 6 jours, du jeudi au mardi de chaque semaine, et ce pour 18 semaines consécutives. La première semaine de collecte s'est étendue du 13 au 18 janvier 2022, la deuxième du 20 au 25 et ainsi de suite jusqu'à la dernière du 12 au 17 mai. Toutes les personnes questionnées étaient âgées d'au moins 18 ans. Le nombre de répondants retenu par semaine oscille entre 3000 et 3027, pour un échantillon complet de 54 155 personnes.

Léger Opinion a utilisé les résultats du recensement de Statistique Canada de 2016 pour pondérer les données de cette enquête de manière à obtenir un échantillon représentatif de la population adulte québécoise. Les statistiques descriptives de l'échantillon représentatif sur les 18 semaines de collecte sont présentées à la table 3.1. On y voit que l'échantillon représentatif est composé majoritairement de femmes (51%). L'âge moyen se situe à 47 ans. L'éducation de niveau collégiale est la plus communément atteinte, avec 44% de l'échantillon, suivi du secondaire (30%) et de l'universitaire (25%). Plus de la moitié de l'échantillon travaille (56%), près du tiers est à la retraite (29%), 7% sont étudiants et 7% sont sans emploi.

⁸ Boisclair, David, Roxanne Borgès Da Silva, Vincent Boucher, Nathalie De Marcellis-Warin, Pierre-Carl Michaud et Ingrid Peignier (2022, 28 janvier). « Combien de personnes ont développés des symptômes ou contracté la COVID-19 au Québec? Une étude exploratoire », *CIRANO*. Récupéré le 15 février 2023 de <https://www.cirano.qc.ca/fr/sommaires/2022s-03>

Table 3.1 : Statistiques descriptives de l'échantillon représentatif

Variable	
Femme	51.2%
Âge moyen	46.72 (17.21)
Revenu par ménage	
<20k	6.97%
20-40k	15.72%
40-60k	19.26%
60-80k	15.75%
80-100k	14.61%
>100k	27.69%
Éducation	
Primaire	0.9%
Secondaire	29.88%
Collégiale	43.95%
Universitaire	25.26%
Personne par ménage	2.41 (1.27)
Ménage avec enfant mineur	25%
Région de résidence	
Grand Montréal	50%
Grand Québec	10%
Autres régions	40%
Situation d'emploi	
Employé temps plein	49.89%
Employé temps partiel	6.1%
Étudiant	7.11%
Retraité	29.12%
Au foyer / sans emploi	7.05%
Vaccin (1 dose et plus)	92.5%

En plus des caractéristiques sociodémographiques listées plus haut, quelques variables sont particulièrement utilisées dans notre analyse. Considérons premièrement la question suivante : *Selon vous, sur une échelle de 0 à 100, quelles sont les chances que vous développiez des symptômes associés à la COVID-19 dans les 7 prochains jours?* Il s'agit d'une anticipation que le répondant produit sur la base des informations qu'il détient. La réponse à cette question correspond à la perception de la probabilité d'attraper la COVID-19 dans la semaine à venir. Nous référerons désormais à cette variable comme étant la probabilité subjective (ou la perception du risque) d'attraper la COVID-19.

Un histogramme de la distribution de la probabilité subjective est présenté à la **Figure 3.1**. On y voit une forte concentration à la probabilité 0 (22 419 personnes), ce qui pourrait refléter qu'une proportion importante de l'échantillon a déjà eu la maladie et s'y considère immunisée, ou simplement que plusieurs personnes sont optimistes quant à leur chance de contracter la COVID-19 (Garfin et al., 2021). On voit également un amoncellement important (6595 personnes) à la probabilité 50. Ceci pourrait refléter une incertitude importante quant au risque perçu par les répondants (ceux-ci estimant les probabilités à 50-50 suivant l'expression populaire). La littérature nomme ce phénomène *l'incertitude épistémique* (Bruine de Bruin et al., 2000). Une discussion concernant ce problème est présentée à la section 4.4. Les personnes ayant la COVID-19 au moment de répondre au questionnaire ont été codées comme ayant une probabilité subjective de 100%. Sur l'échantillon entier, cette probabilité a une moyenne de 17.62%, alors qu'en éliminant les répondants ayant la COVID-19 au moment de répondre la moyenne est de 13.9%.

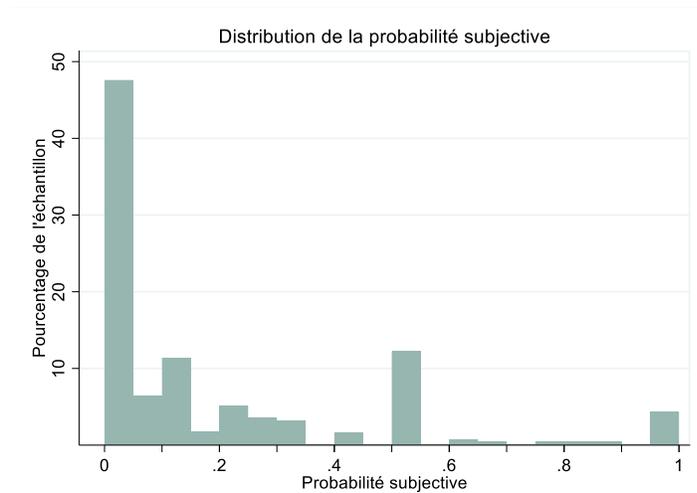


Figure 3.1 : Distribution de la perception du risque pour l'échantillon non-pondéré

Une autre variable particulièrement importante dans notre contexte est construite autour de la question suivante : *Dans les 7 derniers jours incluant aujourd'hui, avez-vous eu un résultat de test positif à la COVID-19?* Les choix de réponse étant 1. *oui*, 2. *non* ou 3. *Pas de résultat de test, mais je crois que je l'ai contractée (autodiagnostic en fonction de mes symptômes des 7 derniers jours)*. Les réponses à cette question ont été utilisées pour construire la variable qui sera dorénavant référée comme étant la probabilité objective (ou risque réel) de contracter la COVID-19. Cette variable binaire est égale à 1 si le répondant a reçu un résultat de test positif à la COVID-19 ou s'il s'est auto diagnostiqué, et égal à 0 autrement.

Dans le contexte d'une étude où est comparée la probabilité objective à la probabilité subjective, il semble adéquat de considérer les autodiagnostic comme étant des cas de COVID-19 confirmés. Ceci aide également à pallier le fait que depuis le 4 janvier 2022, les tests de dépistage PCR n'étaient plus accessibles à la population générale du Québec (ceux-ci étant réservés aux travailleurs essentiels et personnes prioritaires) (Boisclair et al., 2022). La population a donc dû compter exclusivement sur les résultats de tests rapides. Ces tests n'étaient pas en grande disponibilité dans les premières semaines de la collecte de données. Finalement, cette définition plus large permet de se rapprocher de la définition utilisée dans la question subjective portant sur les symptômes. Notons toutefois qu'inclure les autodiagnostic reste une approximation pouvant être source d'erreurs. En divisant le nombre total de personnes ayant répondu avoir la COVID-19 par le nombre de personnes dans l'échantillon, nous estimons que la probabilité objective moyenne d'attraper la COVID-19 est de 3.94%. Cette probabilité a largement varié d'une semaine à l'autre, allant d'un minimum de 2.2% à la semaine 9 (10 au 15 mars) à un maximum de 6.28% à la semaine 13 (7 au 12 avril).

Afin d'encourager la capture par l'enquête de tendances de sous-populations qui auraient autrement pu être délaissées, plusieurs questions concernant le réseau du répondant ont été incluses dans le questionnaire. Ceci permet de mettre à profit la méthode d'amplificateur par réseau⁹. Le présent mémoire concentre ses estimations directement sur l'échantillon représentatif colligé par *Léger Opinion*, et seul un usage limité est fait des variables relatives au réseau du répondant.

3.2 Exploration

3.2.1 Composition sociodémographique

⁹ Pour plus de détails sur la méthode d'amplificateur par réseau, voir Boisclair et al. (2022)

Cette sous-section est dédiée à l'analyse de l'évolution temporelle de certaines des variables d'intérêt à l'étude. Afin d'analyser des tendances dans les variables d'intérêts concernant la COVID-19, il est important de valider que la composition de l'échantillon demeure constante dans le temps. Il peut premièrement être montré que la composition démographique de l'échantillon représentatif est très stable à travers les 18 semaines de collecte de données. En effet, les graphiques de la **Figure 3.2** démontrent l'évolution temporelle de la distribution du sexe, des niveaux d'éducation et de l'âge au sein de notre échantillon. On y voit que la composition en termes de genre et d'éducation est constante. Le léger bruit qui peut être perçu sur le graphique de la distribution de l'âge est sans doute le résultat d'une plus fine granularité dans la décomposition de la variable.

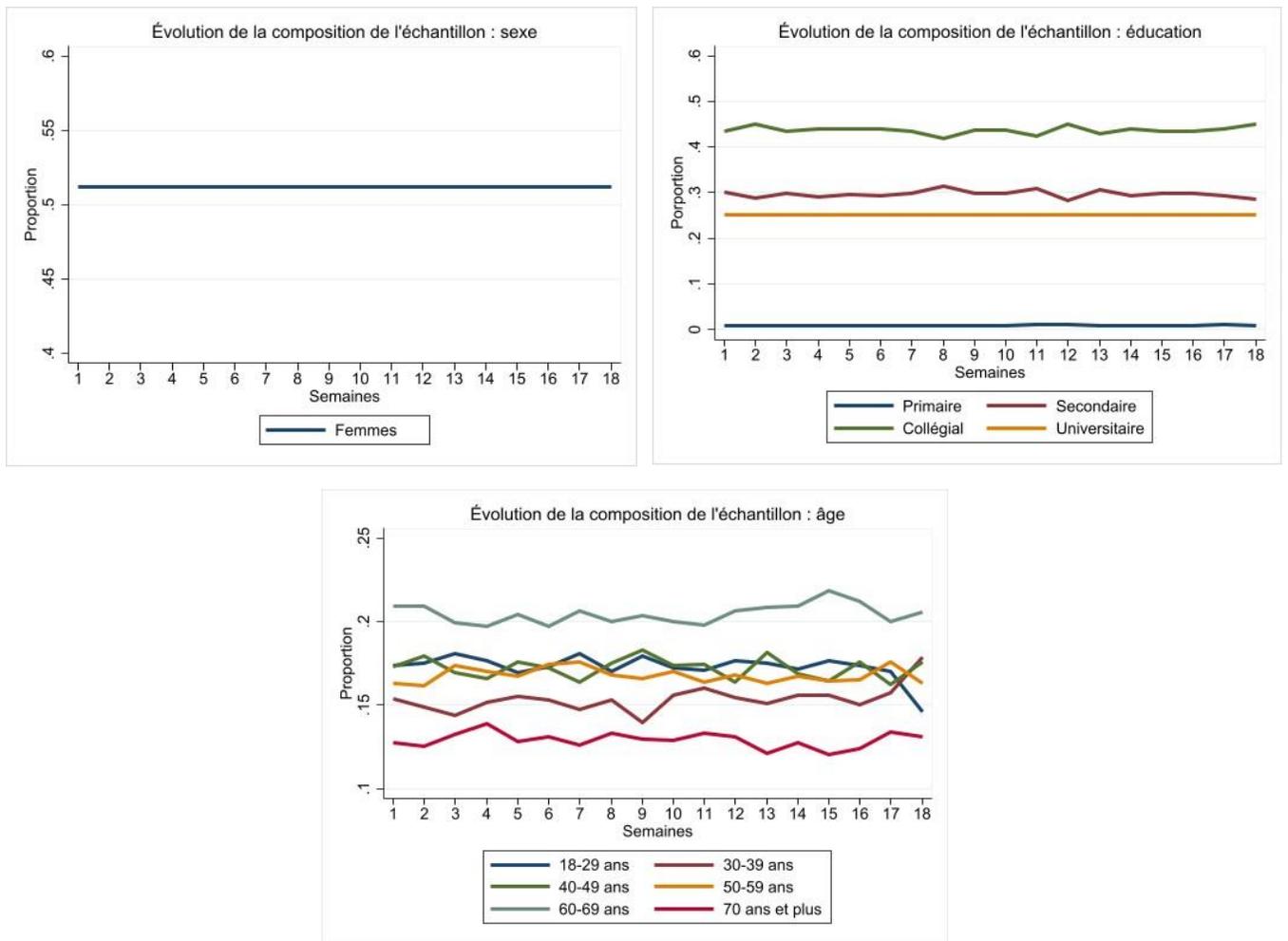


Figure 3.2 : Évolution de la composition de l'échantillon sur les 18 semaines à l'étude

3.2.2 Mesures de risque

Le graphique suivant (**Figure 3.3**) présente la probabilité subjective moyenne et la probabilité objective moyenne d'attraper la COVID-19, et ce pour chacune des 18 semaines de l'échantillon. Il est important de noter que sont comparés à chacune des semaines le risque objectif, tel qu'estimé par le nombre de personnes ayant la COVID-19 dans la dernière semaine, et le risque subjectif pour la semaine à *venir*. Ce décalage temporel explique pourquoi les probabilités subjectives réagissent si rapidement au risque réel. Le graphique nous montre que la perception du risque est nettement supérieure au risque objectif, et ce pour chacune des semaines. On retrouve toutefois la même tendance dans les deux courbes, soient une pente descendante dans les premières semaines, suivie d'une vague débutant à la semaine 10, plafonnant à la semaine 13 et décroissante ensuite. Le fait que les deux courbes suivent une évolution similaire suggère que les gens ont mis à jour leur perception du risque au moins sur une base hebdomadaire, et ce de manière relativement précise. On peut aussi voir que les répondants semblent surréagir à la hausse de cas causant la vague des semaines 10 à 13. En effet, alors que le risque réel n'augmente qu'au plus de 3 à 4 points de pourcentage, la perception du risque sur la même période augmente de 9 à 10 points de pourcentage. Nous verrons plus formellement dans la section *Résultats* que ce type de surréaction est commun dans l'échantillon. Il est possible que la forte perception du risque ait contribué à diminuer les la probabilité objective par l'entremise d'une diminution des comportements risqués, suivant le mécanisme identifié par l'étude de Borrelli et al. (2010). Ce mécanisme aurait alors encouragé une mauvaise calibration.

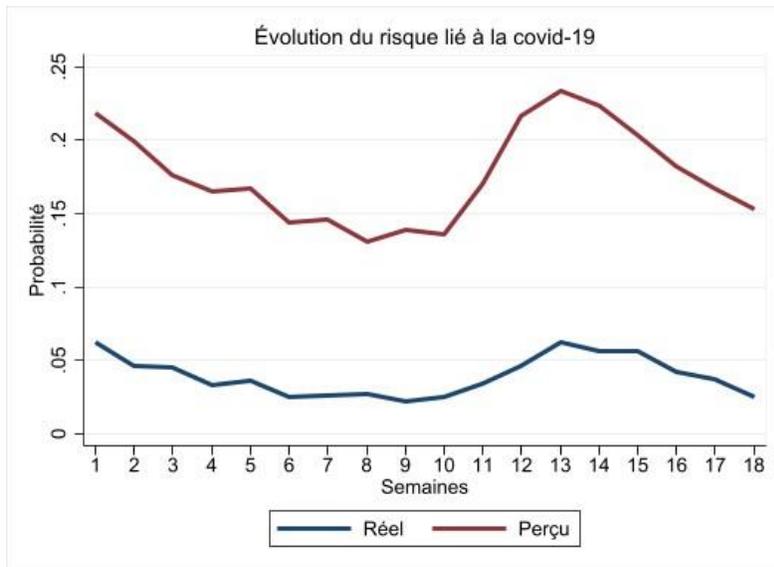


Figure 3.3 : Évolution des deux mesures de risque sur les 18 semaines à l'étude

La **Figure 3.4** décompose la perception du risque parmi les 6 groupes d'âge recensés. On y voit que les perceptions sont très similaires pour les trois premiers groupes (de 18 à 49 ans). Ensuite, chacun des groupes d'âge subséquents se retrouve à percevoir un risque légèrement plus élevé que le groupe d'âge suivant. Ainsi, les 50-59 ans perçoivent plus de risque que les 60-69 ans, qui en perçoivent à leur tour plus que les 70 ans et plus. Il n'est pas surprenant d'obtenir ce type de relation puisque les groupes d'âge plus élevés font sans doute face à un moindre risque d'infection, supposant qu'ils entretiennent de plus petits cercles sociaux et ont moins de chances de faire partie de la population active.

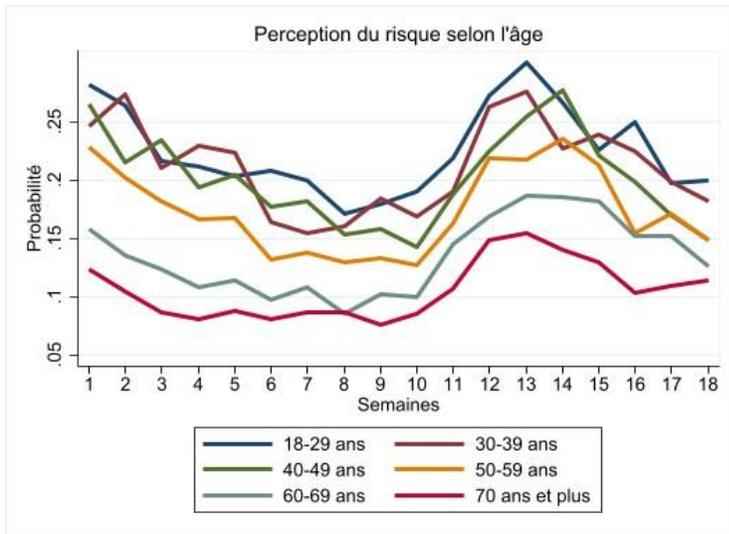


Figure 3.4 : Décomposition de la perception du risque par groupe d'âge

La probabilité subjective peut à nouveau être décomposée, cette fois-ci selon si au moins une personne mineure fait partie du ménage du répondant (**Figure 3.5**). Il a été obtenu que la perception du risque est constamment plus élevée pour les ménages ayant un enfant que pour l'échantillon total. Ceci pourrait être expliqué par le fait que les enfants et les écoles ont souvent été identifiés comme étant des vecteurs de propagation du virus. De manière générale, ce court survol de la variable de probabilité subjective semble indiquer qu'elle est instructive quant à la situation du répondant, et reflète au moins en partie le risque réel auquel il fait face.

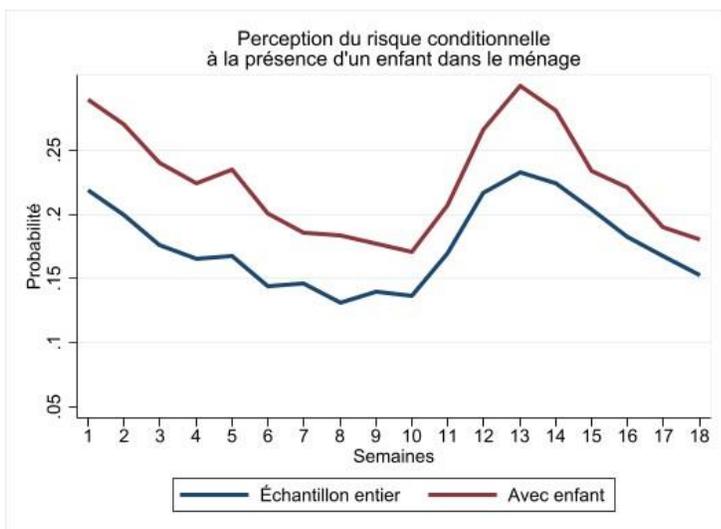


Figure 3.5 : Décomposition de la perception du risque conditionnelle à la présence d'un enfant

3.2.3 Statut vaccinal

En termes de statut vaccinal, seulement 2% de l'échantillon avait reçu quatre doses au moment d'être questionné, 67% avait reçu trois doses, 23% en avait reçu deux et 1% n'en avait reçu qu'une seule. 6% de l'échantillon représentatif n'est pas vacciné. Bien que le nombre de doses de vaccin augmentent parmi les gens vaccinés au fil des semaines de l'échantillon, la proportion de gens vaccinés reste très constante, tournant aux alentours de 93% (**Figure 3.6**). Ceci semble indiquer que le choix de certains d'être non-vacciné est demeuré inchangé sur la période d'étude.

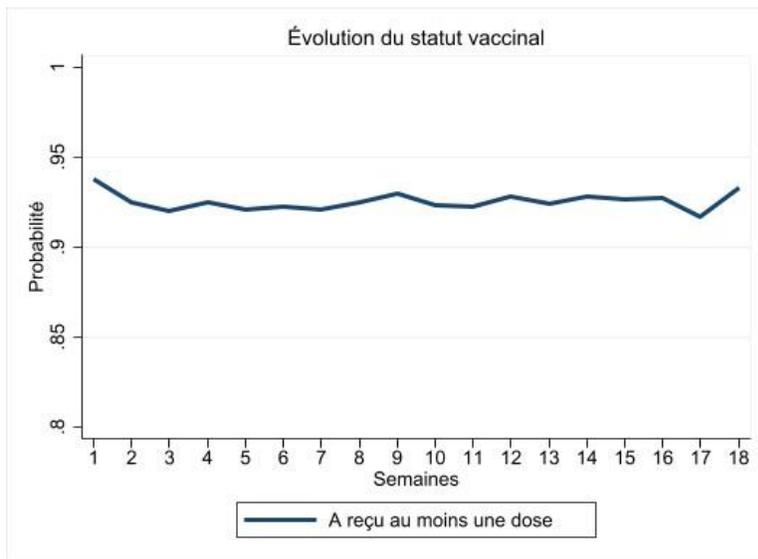


Figure 3.6 : Évolution du statut vaccinal sur les 18 semaines à l'étude

En décomposant le statut vaccinal parmi les différents niveaux d'éducation, une relation évidente ressort. 12.8% des gens n'ayant qu'une éducation primaire ne sont pas vaccinés, contre 7.8% pour le secondaire, 6% au collégial et 3.5% au niveau universitaire. Il semble alors que les gens non-vaccinés soient majoritairement concentrés dans les plus faibles niveaux d'éducation. La décomposition en termes de revenu du ménage révèle une tendance similaire, les plus hauts taux de vaccination correspondant aux plus hautes tranches de revenu. 54.7% des personnes non-vaccinées sont des hommes. Cette distribution du vaccin semble correspondre aux résultats de Leigh et al. (2022), qui prévoyaient déjà en 2020 une intention de vaccination plus importante chez les gens plus éduqués et/ou à plus haut revenu.

La **Figure 3.7** illustre le taux de vaccination selon l'occupation principale du répondant, l'emploi étant ici divisé en 7 secteurs. Une définition détaillée de chacun des secteurs est offerte à la section 4.2. La ligne

rouge horizontale sur le graphique représente le taux de vaccination moyen de l'échantillon représentatif entier. Cette décomposition par occupation nous montre que le vaccin est plus prévalent chez les retraités (97.5%), les étudiants (94.9%) et les employés à temps partiel / temps plein (moyenne pondérée de 91%). Le groupe chez qui le vaccin est le moins populaire est celui regroupant les personnes sans emploi / au foyer, avec un taux de vaccination de 84%. Parmi les répondants employés, ceux du secteur de la construction et de l'exploitation des ressources naturelles détiennent le plus faible taux de vaccination, avec une moyenne de 84.4%. En décomposant le taux de vaccination par groupes d'âge (**Figure 3.8** **Figure 3.8**), on obtient une courbe en U. On y voit que les jeunes adultes et les personnes âgées sont hautement vaccinés, alors que le taux de vaccination chute de quelques points de pourcentage pour les personnes âgées de 30 à 45 ans.

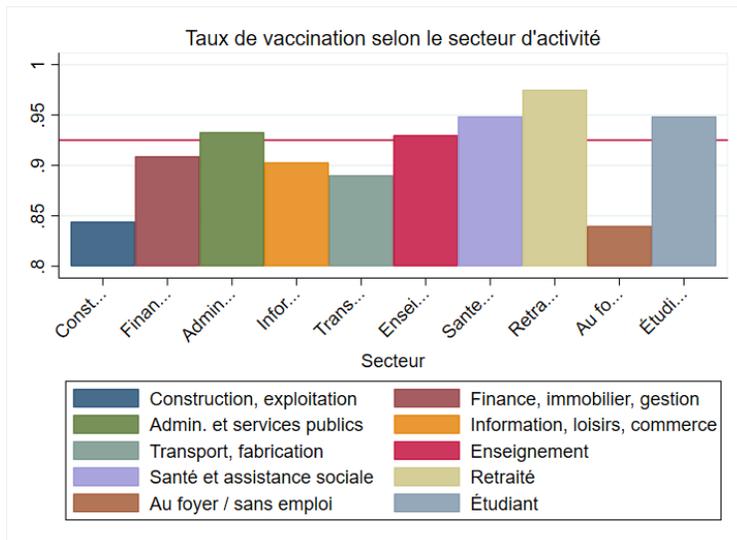


Figure 3.7 : Taux de vaccination selon l'occupation principale du répondant

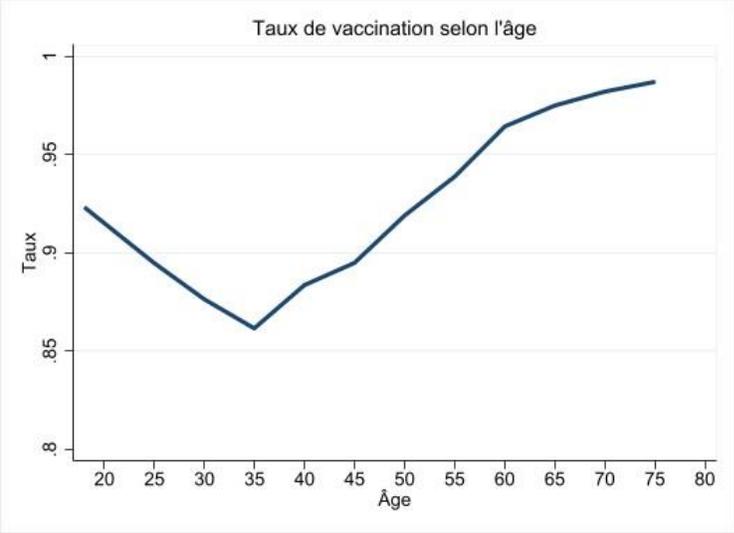


Figure 3.8 : Taux de vaccination selon les différents groupes d'âge

4. Méthodologie

Le survol des données montre à travers des comparaisons naïves qu'il existe un écart important dans la calibration du risque des répondants. Afin de creuser cette relation, l'analyse qui suit fait usage de régressions pour décomposer les différentes mesures de risque à travers plusieurs caractéristiques sociodémographiques. Ceci permet d'estimer l'effet marginal liant une caractéristique à chacune des mesures de risque. Les caractéristiques regroupant des sous-populations particulièrement mal calibrées peuvent ainsi être isolées, et ce en compensant pour certaines différences entre individu.

4.1 Structure économétrique

Notre objectif est donc de tester la calibration du risque subjectif au risque objectif. Pour ce faire, nous avons recours à un modèle économétrique. Commençons par dénoter par n le nombre de répondants par semaine de collecte de données, et t semaines de collecte. Le nombre total de répondants est ainsi résumé par $\sum_{t=1}^{18} n_t$. Malgré que la collecte de données se soit échelonnée sur 18 semaines, la forte majorité de l'analyse de cette étude se fait en coupe transversale, regroupant toutes les semaines en une seule période. La modélisation qui suit ne considère alors pas l'aspect temporel des données. Toutefois, afin de contrôler pour l'état constamment changeant de la pandémie, il semble pertinent d'inclure une constante pour chacune des semaines de collecte. On s'attend donc à ce que ces constantes capturent l'effet que pourrait avoir une forte hausse (ou baisse) de cas sur les variables dépendantes. Toute chose étant égale par ailleurs, les coefficients relatifs aux variables explicatives d'intérêt devraient ainsi refléter une tendance relativement constante dans le temps. Nous analysons cette conjecture plus tard dans l'analyse.

Considérons premièrement la situation du risque objectif d'infection à la COVID-19 encouru par l'agent. La réalisation de l'infection est recensée par la variable Y , elle-même définie de manière binaire pour chaque répondant i ($Y_i = 1$ si le répondant a la COVID-19, 0 sinon). Cette dernière est une variable aléatoire de Bernoulli, et peut donc être présentée comme suit, où X réfère à un vecteur de variables explicatives sociodémographiques et l'indice o dénote la référence au risque objectif (réel) :

$$E_o[Y_i|X_i] = 1 * Pr(Y_i = 1 | X_i) + 0 * Pr(Y_i = 0 | X_i) \quad (4)$$

La nature dichotomique de cette variable dépendante fait en sorte qu'elle devrait être estimée à l'aide d'un modèle de régression particulier (probit ou logit), qui considère la distribution binomiale de Y_i et permet l'obtention d'estimateurs non-biaisés relatant l'effet de X_i sur cette probabilité. Toutefois, dans

l'optique de faciliter la comparaison entre les coefficients obtenus de cette régression et ceux obtenus de la régression sur la probabilité subjective, nous posons ici une hypothèse de linéarité sur la forme fonctionnelle de la relation liant Y_i aux caractéristiques X_i . Afin de permettre l'évaluation de cette hypothèse, la comparaison des effets marginaux probit et des coefficients obtenus par estimation linéaire est présentée à la section 5.1 du travail. Les légères différences de coefficients relevées sont sans doute le résultat d'effets marginaux se situant dans une section moins linéaire de la CDF. Aucune de ces différences n'est suffisante pour occasionner un changement de signe ou de significativité dans nos principales variables d'analyse (enseignement, santé et vaccin). Puisqu'il peut y être vu que les coefficients sont d'amplitude similaires, nous poursuivons l'analyse avec le modèle linéaire

L'hypothèse de linéarité imposée à Y_i implique que l'espérance conditionnelle du risque réel est de la forme suivante, où β_o réfère à un vecteur de coefficients relatant l'impact de différentes caractéristiques X_i sur le risque réel :

$$E_o[Y_i|X_i] = X_i\beta_o \quad (5)$$

Le terme d'erreur $\varepsilon_{i,o}$ est alors défini comme étant la différence entre la réalisation de l'infection à la COVID-19 et l'espérance de l'infection conditionnelle au vecteur X_i . Ce terme d'erreur est de moyenne 0 par construction.

$$\varepsilon_{i,o} = Y_i - E_o[Y_i|X_i] = Y_i - X_i\beta_o \quad (6)$$

Posons maintenant les bases de la régression sur la probabilité subjective. Chaque répondant a en tête une estimation du risque d'infection auquel il fait face, défini par l'expression $E_s[Y_i|X_i]$, où l'indice s réfère à l'aspect subjectif de cette estimation. Nous posons l'hypothèse que cette espérance est égale au produit d'un vecteur de caractéristiques sociodémographique X_i et de coefficients β_s .

$$E_s[Y_i|X_i] = X_i\beta_s \quad (7)$$

Ce mémoire a la chance de bénéficier d'une évaluation directe de $E_s[Y_i|X_i]$ fournie par le répondant, prenant la forme d'une estimation probabiliste dénotée P_i .

$$E_s[Y_i|X_i] = E[P_i|X_i] = X_i\beta_s \quad (8)$$

Ainsi, sur la base de l'observation de X_i , l'espérance conditionnelle subjective est la moyenne des probabilités subjectives des répondants possédant les caractéristiques X_i . L'équation 6 définit quant à elle le terme d'erreur $\varepsilon_{i,s}$, de moyenne 0 par construction.

$$\varepsilon_{i,s} = E_s[Y_i|X_i] - P_i \quad (9)$$

La comparaison des équations (5) et (8) suggère qu'on peut définir la rationalité des anticipations par l'égalité des paramètres subjectifs et objectifs (assumant la présence d'une constante). Débutons par former un système d'équations linéaires.

$$P_i = X_i\beta_s + \varepsilon_{i,s} \quad (10)$$

$$Y_i = X_i\beta_o + \varepsilon_{i,o} \quad (11)$$

Le coefficient β_j correspond à l'effet marginal¹⁰ qu'entraîne la caractéristique $X_{i,j}$ sur le risque du répondant i . Ainsi, un $\beta_{s,j} > 0$ indique que le groupe ayant la caractéristique $x_j = 1$ présente une estimation du risque qui est plus élevée que celle des répondants ne faisant pas partie du groupe (ayant $x_j = 0$).¹¹ Ainsi, les caractéristiques démographiques contenues dans le vecteur X_i peuvent dans ce contexte être vues comme des facteurs augmentant ou diminuant le risque de l'agent, et ce à l'échelle des coefficients β . Puisqu'il est peu probable qu'en présence d'une variable dépendante binaire le terme d'erreur soit normalement distribué, les écarts-types de ces deux régressions sont calculés de manière robuste.

L'hypothèse de rationalité dans la formation des attentes entend que l'espérance des anticipations pour chaque groupe X_i doit être équivalente à la réalisation. Dans notre contexte, la validité de cette hypothèse peut être évaluée en effectuant un test d'égalité des coefficients entre les deux régressions (Benitez-Silva et al., 2004), puisque :

$$\begin{aligned} E[P_i - Y_i|X_i] &= 0 \\ \Leftrightarrow E[X_i\beta_s - X_i\beta_o|X_i] + E[\varepsilon_{i,s} - \varepsilon_{i,o}|X_i] &= 0 \end{aligned}$$

¹⁰ Notons que nous comprenons l'expression « effet marginal » dans le sens d'association, ici utilisée suivant la convention économétrique. Nous n'entendons en aucun cas que les associations identifiées impliquent causalité.

¹¹ Similairement, un $\beta_{o,j} > 0$ implique que la caractéristique X_j regroupe une sous-population faisant face à un risque plus élevé d'être infecté par la COVID-19.

$$\begin{aligned} \Leftrightarrow X_i(\beta_s - \beta_o) &= 0 \\ \Leftrightarrow \beta_s &= \beta_o \end{aligned} \quad (12)$$

L'hypothèse testée peut alors être résumée à l'équivalence des coefficients inter-régression, soit $\beta_{j,s} = \beta_{j,o}$. La présence d'une constante dans le modèle définit un groupe de référence. Si ce groupe de référence a des anticipations rationnelles, alors le respect de l'équation (12) nous permettrait de conclure en faveur de la rationalité des anticipations de la sous-population ayant la caractéristique $x_j = 1$. Autrement, si on observe deux coefficients significativement différents, il serait possible de rejeter l'hypothèse nulle de rationalité des anticipations et de conclure qu'il existe un biais significatif dans l'estimation du risque, capté par la variable x_j . Le cas où $\beta_{j,s} > \beta_{j,o}$ correspond alors à une surestimation du risque d'infection par le groupe ayant la caractéristique j . Il pourrait alors être dit que ce groupe entretient un biais positif.¹² De manière générale, si on suppose que la différence des coefficients de la variable j est égale à b_j ($\beta_{j,s} - \beta_{j,o} = b_j$) et que b_j est différent de 0, alors il peut être affirmé que la population ayant la caractéristique j entretient un biais de perception de l'échelle de b_j points de pourcentage.

L'égalité entre chaque couple de coefficients est estimée à l'aide d'un test de Wald prenant la forme suivante :

$$W_j = \frac{(\beta_{j,o} - \beta_{j,s})^2}{\text{var}(\beta_{j,o}) + \text{var}(\beta_{j,s}) - 2 * \text{cov}(\beta_{j,o}, \beta_{j,s})} \sim \chi_j^2 \quad (13)$$

Dans ce contexte, avec un seul degré de liberté, obtenir un $\chi^2 > 3.841$ permet de rejeter l'hypothèse nulle d'égalité des coefficients au seuil de significativité de 5%. Tout test satisfaisant cette condition permet alors de conclure en faveur du rejet de l'hypothèse de rationalité la formation des anticipations.

4.2 Variables explicatives

Le vecteur de variables explicatives X_i est composé de plusieurs variables sociodémographiques. Ceci permet de sectionner la population en plusieurs sous-groupes et de tester à quel point les estimations de chaque groupe sont précises. Chacune des variables est construite de manière dichotomique, les variables catégoriques ont ainsi été décomposées en variables binaires. Ceci rend l'estimation des coefficients

¹² À l'inverse, le cas où $\beta_{j,s} < \beta_{j,o}$ implique la présence d'un biais négatif, soit une situation où un groupe sous-estime le risque auquel il fait face.

simple : ils sont directement interprétables comme étant des effets marginaux influençant la probabilité par β points de pourcentage. Il doit être noté que cette structure fait en sorte que chacun des coefficients exprime une différence entre le groupe ayant la caractéristique j ($x_j = 1$) et celui ne l'ayant pas ($x_j = 0$).

Voici la liste des variables utilisées dans cette analyse, avec entre parenthèses chacun des choix de réponses considérés :

- le sexe (homme, femme),
- l'âge (divisé en 6 tranches : 18-29 ans, 30-39, 40-49, 50-59, 60-69, 70+),
- le revenu du ménage (divisé en 6 tranches : <20k, 20-40k, 40-60k, 60-80k, 80-100k, >100k),
- le niveau d'éducation (primaire, secondaire, collégial ou universitaire),
- une variable indiquant si au moins un enfant mineur réside dans le ménage,
- le nombre de personne dans le ménage (1, 2, 3, 4, 5+),
- la région de résidence (région du grand Montréal, région de Québec, autres régions),
- le statut vaccinal (ici simplifié à si le répondant a reçu au moins une dose de vaccin),
- le secteur d'activité :
 - secteur 1 : Agriculture, foresterie, pêche et chasse, extraction minière, exploitation en carrière, extraction de pétrole et de gaz, construction (N = 2 256)
 - secteur 2 : Finance et assurances, services immobiliers, services de location et de location à bail, services professionnels, scientifiques et techniques, gestion de sociétés et d'entreprises, services administratifs, services de soutien, services de gestion des déchets et services d'assainissement, autres services (N = 9 257)
 - secteur 3 : Administrations publiques et services publics (N = 4 931)
 - secteur 4 : Industrie de l'information, industrie culturelle, commerce de détail, arts, spectacles, loisirs, services d'hébergement et de restauration (N = 4 411)
 - secteur 5 : Transport, entreposage, commerce de gros, fabrication (N = 4 048)
 - secteur 6 : Services d'enseignement (N = 3 166)
 - secteur 7 : Soins de santé et assistance sociale (N = 3 687)
 - secteur 8 : Retraité (N = 16 050)
 - secteur 9 : Au foyer ou sans emploi (N = 3 357)
 - secteur 10 : Étudiant sans emploi (N = 2 639)
- une constante par semaine

Certains répondants ont omis de fournir une réponse à certaines des questions, laissant quelques variables sans réponse. Plutôt que d'omettre ces répondants de l'échantillon, une variable binaire additionnelle a été créée et incluse pour regrouper les non-réponses et potentiellement contrôler pour une répartition non-aléatoire des omissions. Plus précisément, cette méthodologie a été appliquée aux variables du revenu (5615 non-réponses), de l'éducation (257), de la présence d'enfant (250), du nombre de personne dans le ménage (250), du secteur d'occupation (353) et du statut vaccinal (745).

4.3 Précisions économétriques

En utilisant un système de régression basé sur la même population, il est fort probable que les termes d'erreur soient corrélés d'une régression à l'autre. Les erreurs seraient indépendantes entre observations d'une même régression mais seraient corrélées pour une même personne à travers les deux régressions ($\rho(\varepsilon_{i,s}, \varepsilon_{i,o}) \neq 0$). Ceci pose problème puisque pour tester l'hypothèse des anticipations rationnelles ($\beta_{j,s} = \beta_{j,o}$), des coefficients obtenus de régressions distinctes doivent être comparés. Afin de compenser pour cette corrélation, il est possible d'utiliser le modèle économétrique de *Seemingly Unrelated Regression* (SUR) tel qu'établi par Zellner (1962). Il est connu qu'en utilisant exactement les mêmes variables explicatives dans chacune des régressions, le modèle de SUR offre des estimations et erreurs standards (non-robustes) équivalentes à l'estimation par moindres carrés ordinaires (MCO). Malgré cela, le présent mémoire fait usage du modèle de SUR puisqu'il offre également l'avantage considérable de permettre de tester conjointement deux coefficients obtenus de régressions distinctes. Il est important de noter que, sauf indication contraire, chacun des résultats présentés dans ce papier est obtenu en pondérant les observations de manière à retrouver l'échantillon représentatif de la population québécoise. Tous les calculs sont effectués sur la version 17.0 de Stata.

4.4 Incertitude épistémique

Dans la section 4.1, nous avons posé quelques hypothèses permettant de structurer empiriquement ce travail. L'une de ces hypothèses, résumée par l'égalité $E_s[Y_i|X_i] = E[P_i|X_i]$, stipule qu'à défaut de pouvoir observer l'espérance subjective du risque d'infection, cette dernière sera déduite à partir de l'estimation probabiliste subjective du risque fournie par les répondants (P_i). En utilisant cette dernière, ce mémoire ouvre la porte au biais de l'incertitude épistémique. Ce problème bien connu des probabilités subjectives survient lorsqu'un répondant inscrit 50 comme probabilité, non pas parce que c'est à ce niveau qu'il estime la probabilité, mais parce qu'il ne sait pas quoi répondre (Bruine de Bruin et al., 2000). La probabilité 50 est alors vue comme étant une manière de dire que l'un ou l'autre des résultats est plausible, et que le répondant ne saurait trancher.

L'étude de Bruine de Bruin avance d'ailleurs plusieurs facteurs qui prédisent une plus forte tendance à éliciter la probabilité 50 pour son ambiguïté plutôt que sa valeur. Parmi ceux-ci se retrouvent les questions ouvertes, qui demandent directement au répondant d'estimer une probabilité. Les auteurs proposent une manière alternative de formuler la question, qui serait d'offrir la possibilité au répondant de répondre soit par une probabilité ou par une étendue de probabilités¹³. Le travail de Manski (2004) sur les probabilités subjectives reconnaît la valeur que peut avoir cette formulation plus permissive et encourage son usage. L'enquête sur laquelle se base ce travail formule toutefois la question de manière ouverte, encourageant indirectement l'incertitude épistémique.

Un autre facteur identifié par Bruine de Bruin et ses collaborateurs est le degré de familiarité qui unit le répondant et le risque qu'il lui est demandé d'estimer. Les auteurs avancent que les événements sur lesquels le répondant a peu de contrôle ou pour lesquels il est difficile de déterminer une probabilité entraînent plus d'incertitude épistémique. Estimer sa vulnérabilité à l'infection d'un virus dans la prochaine semaine rentre certainement dans la catégorie des événements dont la probabilité est difficile à estimer.

L'incertitude épistémique pose surtout problème dans un contexte de faibles probabilités comme le nôtre ($\bar{P} = 13.9\%$), puisqu'elle se retrouve à augmenter la moyenne du risque perçu et donc à biaiser à la hausse les résultats. Il ne serait toutefois pas adéquat de simplement retirer tous les répondants ayant défini leur risque perçu à 50% car nous viendrions biaiser notre échantillon représentatif de la population québécoise, spécialement parce qu'il est généralement accepté que les personnes plus jeunes et moins éduquées ont plus souvent recours à la probabilité 50 dans les enquêtes probabilistes (Bruine et Bruin, 2000). Également, le fait que 780 personnes aient évalué subjectivement leur risque à 40% et 361 personnes à 60% indiquent que certainement des répondants ont choisi 50% car cela représentait leur réelle estimation du risque perçu.

La solution à cet enjeu sera premièrement de présenter les résultats sur l'échantillon entier représentatif de la population québécoise, sans regard particulier pour l'incertitude épistémique. Ces résultats demeurent intéressants car leur rôle principal est d'isoler les caractéristiques sociodémographiques prédictives de biais. L'estimation de la valeur réelle que prennent ces biais est un ajout pertinent mais secondaire à cette analyse. La robustesse des résultats sera ensuite évaluée à l'aide d'une analyse de sensibilité, où sont omis de l'échantillon les répondants ayant fourni une estimation du risque d'infection

¹³ Le répondant incertain aurait alors la possibilité d'estimer la probabilité comme étant entre 0 et 100 plutôt que fixée à 50.

à 50%. La valeur réelle que prend le biais se trouvera alors sans doute entre les estimations sur l'échantillon entier et celles sur l'échantillon tronqué. Il pourra y être vu que les relations d'intérêts approfondies dans l'étude (notamment le biais de l'enseignement et celui du vaccin) demeurent significatives à 0.1% même après correction.

4.5 Groupe de référence : variables omises

Toutes les variables incluses dans les régressions sont des variables binaires. Ceci présente l'avantage de transformer les coefficients en effets marginaux facilement interprétables, mais implique que tous les résultats ont à être interprétés en relation au groupe de référence défini par les variables omises.

La première régression qui est présentée omet le quatrième groupe d'âge, soit les 50-59 ans. La **Figure 3.4** décomposant le risque par groupe d'âge suggère que cette tranche d'âge suit la tendance générale, et est donc un contrefactuel crédible. La troisième tranche de revenu (40-60k), le niveau d'éducation collégial, le ménage à 2 personnes et la région de Montréal sont les autres variables exclues de la régression, et correspondent donc au groupe de référence auquel se doivent d'être comparés chacune des variables.

Les principaux secteurs que ce mémoire désire analyser sont ceux de l'enseignement (6) et de la santé (7), il faut donc isoler le contrefactuel s'y rattachant de la manière la plus plausible possible parmi les autres choix. Les secteurs 8, 9 et 10 ont d'emblée été éliminés car ils ne correspondent pas au marché du travail. Les secteurs 1 et 5 ont ensuite été éliminés parce qu'ils réfèrent à un travail de nature physique qui semble dissocié des milieux de l'enseignement et de la santé. Afin de choisir parmi les 3 secteurs restants (2, 3 et 4), un critère de présence au travail a été appliqué. Une caractéristique commune aux secteurs de l'enseignement et de la santé est un fort taux de travail en présentiel durant la pandémie (respectivement 66 et 73% sur notre échantillon). Cette réalité était absente des secteurs 2 et 3 (respectivement 37 et 49% en présentiel), ce qui a poussé notre choix de contrefactuel sur le secteur 4, principalement composé d'emplois au service à la clientèle. Ce dernier avait un taux de travail en présentiel de 69%. Il semble judicieux d'utiliser ce critère puisque l'étude porte sur les probabilités (perçues et réelles) d'être infecté par la COVID-19, et que le contact humain au travail peut être un vecteur important de contagion.

5. Résultats

Les résultats des régressions et ceux des tests d'égalité sont présentés à la table 5.1. Les deux premières colonnes réfèrent à la régression ayant la probabilité subjective comme variable dépendante, les deux suivantes réfèrent à celle ayant la probabilité objective (risque réel) comme variable dépendante et les deux dernières réfèrent aux résultats des tests de Wald. Le degré de significativité habituel de 5% est utilisé dans les résultats présentés plus bas. Les différents coefficients sont également illustrés à l'aide de 3 figures. Les résultats relatifs aux variables du sexe, de l'âge, du revenu et de l'éducation sont présentés à la **Figure 5.1**. Ceux relatifs à la composition du ménage, la région de résidence et le statut vaccinal sont regroupés à la **Figure 5.2** alors que la décomposition par secteur d'activité et résumée à la **Figure 5.3**. Dû au nombre important de coefficients, la présentation textuelle des résultats offre un survol des principales relations trouvées, mais n'est en aucun cas exhaustive. Rappelons ici que chaque coefficient présente une variation en points de pourcentage relative à une mesure du risque. Certaines de ses variations peuvent sembler d'ampleur infime, mais se doivent d'être comparées à une probabilité moyenne de 13.9% pour la perception du risque et de 3.94% pour le risque réel.

Sexe

Notons premièrement que le sexe n'affecte pas l'estimation de la probabilité d'attraper la COVID-19. Toutefois, la régression ayant comme variable dépendante la probabilité objective montre qu'il existe une différence significative au niveau du risque objectif que court les deux sexes. En effet, les femmes semblent faire face à un risque réel significativement plus faible que celui des hommes, et ce à l'échelle de 0.39 points de pourcentage. Une diminution significative de près d'un demi-point de pourcentage est une variation proportionnellement très importante, ici équivalente à une variation de près de 10% du risque moyen de l'échantillon. Le test de Wald confirme que ces coefficients sont statistiquement différents. Ceci indique que la différence dans la probabilité d'infection entre les hommes et les femmes qui a été identifiée n'est pas prise en compte par une différence équivalente au niveau des probabilités subjectives. Il semblerait ainsi que les femmes (ou les hommes) n'aient pas ajusté leur perception du risque à la baisse (à la hausse). Il est alors possible de parler de la présence d'un biais, ici capté par la variable indicatrice du sexe. L'interprétation du signe du biais dépend de la perspective adoptée : soit les femmes ont un biais positif ou les hommes un biais négatif.

Âge

Au niveau de l'estimation du risque, chaque groupe d'âge entraîne sa propre variation significative lorsque comparé aux 50 à 59 ans. La tendance ressortant des coefficients est que les répondants plus jeunes ont une perception plus élevée de leur risque alors que les plus âgés en ont une perception plus faible. Par exemple, les 18 à 29 ans estiment que leur risque d'attraper la COVID-19 dans la semaine à venir est 4.25 points de pourcentage au-dessus de celui des 50 à 59 ans, alors que les plus de 70 ans estiment que leur risque est inférieur par 4.73 points de pourcentage au même groupe de référence. La comparaison avec les coefficients relatifs au risque réel témoigne d'une tendance très similaire. Il peut être vu que lorsque comparés aux 50 à 59 ans, les répondants de 18-29 ans courent un risque réel plus élevé par 2.85 points de pourcentage. Au contraire, lorsque comparés au même groupe de référence, les répondants de 70 ans et plus voient une diminution de leur risque de l'échelle de 0.96 point de pourcentage. Les 5 tests d'égalité de coefficients sont significatifs, indiquant qu'appartenir à chacun des groupes d'âge entraîne un biais. Les biais relatifs à l'âge sont substantiels, et bien qu'ils ne soient pas approfondis ici ce mémoire reconnaît la valeur que pourrait avoir une analyse relative plus complète.

Revenu

En termes de revenu, les ménages ayant les plus hauts revenus (>100k) et ceux dans la tranche de 20-40k perçoivent un risque significativement plus faible que le groupe de référence de 40-60k (-1.04 et -0.82 points de pourcentage respectivement). La décomposition par tranches de revenu n'offre pas de variation significative du risque réel. Les tests de Wald ne sont significatifs que pour les ménages rapportant <20k par année et ceux rapportant entre 20k et 40k. Dans les deux cas, le biais prend la forme d'une sous-estimation du risque.

Éducation

Seul le niveau d'éducation universitaire influe sur le risque perçu, et ce par une hausse de 0.66 point de pourcentage au-dessus des répondants ayant une éducation collégiale. Similairement pour le risque réel, les niveaux d'éducation primaire et secondaire n'offrent pas de variations significatives lorsque comparés au collégial, alors que les gens ayant un diplôme universitaire voient leur risque réel augmenter significativement de 0.55 point de pourcentage. Aucun test de Wald n'est significatif pour la décomposition par niveau d'éducation, suggérant que ces variables ne recensent pas de biais.

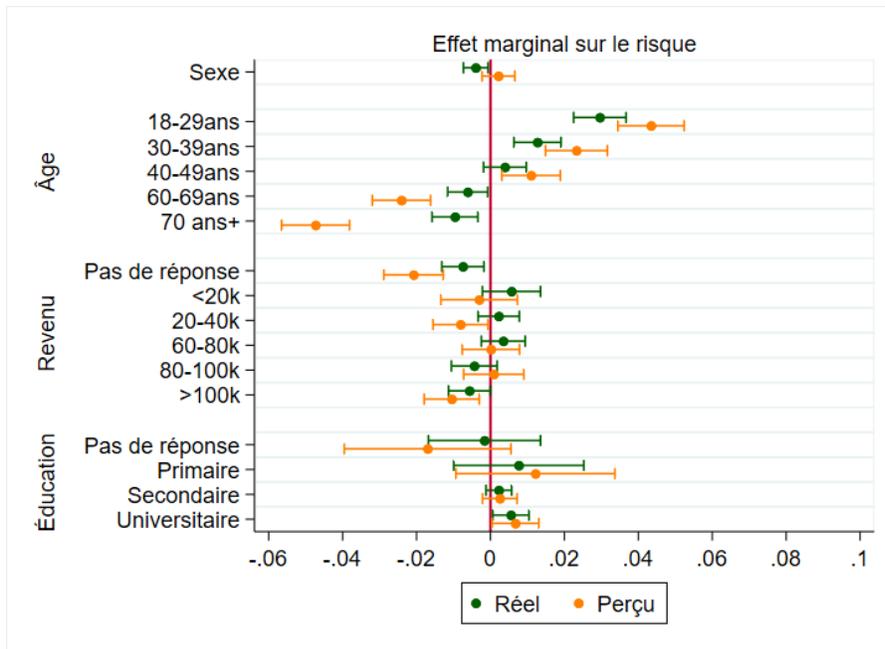


Figure 5.1 : Comparaison de l'effet marginal des variables sociodémographiques sur le risque réel et perçu

Composition du ménage

Les ménages ayant au moins un enfant (moins de 18 ans) vivant sous leur toit voient leur probabilité subjective augmentée de 3.8 points de pourcentage. L'augmentation du risque réel n'est toutefois que de 2.4 points de pourcentage. Le test permet d'affirmer que ces deux coefficients sont différents. Il est alors possible d'isoler un biais positif, prenant la forme d'une surestimation du risque qu'engendre la présence d'un enfant dans le ménage. Un biais de la sorte peut mener à des conséquences importantes car le parent agit suivant son estimation subjective (biaisée) du risque et non selon le risque réel.

Les ménages composés de 4 personnes ont une estimation du risque plus faible par 0.93 point de pourcentage à celle des ménages de deux personnes. Avoir un ménage de 4 personnes semble également diminuer le risque réel, et ce à l'échelle de 0.96 point de pourcentage. Le test de Wald confirmant l'égalité de ces coefficients.

Région de résidence

Lorsque comparés aux résidents de la grande région de Montréal, aucune différence significative n'existe pour les répondants de Québec au niveau de leur perception du risque, alors que les personnes habitant dans les autres régions de la province estiment avoir un risque d'infection plus élevé par 0.83 point de

pourcentage. Cette différence de perception ne semble toutefois pas être fondée, puisque la décomposition du risque réel ne résulte en aucune différence au niveau de la région de résidence. Le test de Wald permet encore une fois d'isoler un biais de perception. Si nous assumons la rationalité des attentes des habitants de Montréal, il pourrait être dit que les personnes habitant dans les autres régions de la province (excluant la ville de Québec) entretiennent un biais positif. Autrement, la différence de coefficients serait expliquée par une sous-estimation du risque par les habitants de Montréal.

Vaccin

Il est particulièrement intéressant de voir que les gens ayant reçu au moins une dose de vaccin estiment leur risque d'attraper la COVID-19 à 7.99 points de pourcentage au-dessus de l'estimation des non-vaccinés. Le risque réel d'infection n'est quant à lui pas affecté par le statut vaccinal. Il peut être conclu que ces coefficients sont significativement différents. Nous nous retrouvons alors ici en présence d'un biais positif très important. Il devrait être précisé que ces résultats ne permettent pas encore de conclure sur quel groupe parmi les vaccinés et non-vaccinés est biaisé, il peut seulement être dit qu'il existe un biais d'envergure considérable chez au moins l'un d'entre eux.

Il peut sembler surprenant que le vaccin ne diminue pas le risque réel d'être infecté par la COVID-19, mais rappelons qu'en début 2022, moment de la collecte de données, le variant de COVID-19 qui était le plus prévalent au Québec était Omicron. En effet, l'INSPQ a publié un communiqué le 21 décembre 2021 indiquant que seulement trois semaines après le premier cas d'Omicron recensé au Québec, le variant avait déjà atteint une prévalence de 80%¹⁴. Les caractéristiques de ce variant font en sorte que le vaccin protège beaucoup moins contre son infection, mais prévient l'apparition de symptômes graves (Abu-Raddad et al., 2022). Il est autrement possible qu'un effet rebond ait fait en sorte que les personnes vaccinées adoptent des comportements plus risqués, suivant un mécanisme semblable à celui identifié par Luckman et al. (2021), augmentant ainsi leur probabilité objective.

Il est également intéressant de se pencher sur la situation des 745 personnes ayant omis de répondre aux questions sur le statut vaccinal. Il est obtenu que ce groupe présente un risque réel et subjectif non-différent de celui des non-vaccinés. Ce résultat suggère que les personnes ayant omis de fournir leur statut vaccinal sont sans doute majoritairement non-vaccinées. L'analyse du biais associé au statut vaccinal est approfondie à la section 5.3.2.

¹⁴ <https://www.inspq.qc.ca/nouvelles/variant-omicron-est-desormais-dominant-quebec>

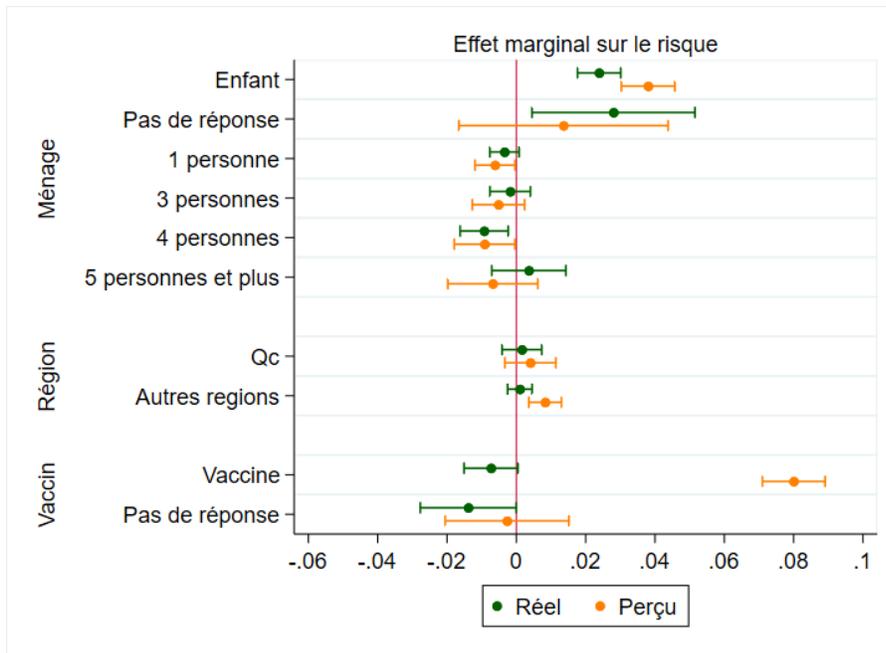


Figure 5.2 : Comparaison de l'effet marginal de variables sociodémographiques sur le risque réel et perçu

Secteurs

Pour ce qui en est des secteurs d'emploi, les secteurs 2, 5 et celui regroupant les étudiants n'ont pas de variation significative du risque perçu lorsque comparés aux travailleurs du secteur 4. Au contraire, on remarque que travailler dans les secteurs 1 et 3 augmente la perception du risque par 1.4 et 1.32 points de pourcentage. Au niveau du risque réel, il n'y a que le secteur 1 qui présente un risque significativement plus élevé, et ce par 1.31 points de pourcentage. Le test relatif au premier secteur conclue en faveur de l'égalité des coefficients, suggérant que les travailleurs de ce secteur ont rationnellement anticipé l'augmentation de leur risque relativement au secteur de référence.

Les personnes retraités et sans emploi voient leur situation affecter leur perception du risque de manières similaires, avec des baisses respectives de 2.18 et 2.14 points de pourcentage. De ces deux groupes, il n'y a toutefois que les personnes sans emploi qui ont une réelle diminution du risque (-1.23 points de pourcentage). Les tests de Wald permettent ici de confirmer la significativité de la différence entre les deux coefficients des retraités et également ceux des sans-emplois. Nous nous retrouvons alors ici en présence de nouveaux biais. Ceux-ci peuvent être vus comme négatifs si il est assumé qu'ils proviennent des groupes de traitement (les retraités et les sans-emplois) ou comme positif si il est estimé qu'il provient plutôt du groupe de référence lui-même biaisé.

Les travailleurs de la santé voient leur probabilité subjective augmentée de 3.74 points de pourcentage alors que leur risque réel n'est pas différent de celui de notre groupe de référence. Le test d'égalité des coefficients pour le secteur de la santé permet de confirmer la présence d'inégalité. Ceci laisse présager l'existence d'un biais positif considérable. Il est intéressant de voir que les employés du secteur de la santé n'étaient pas plus à risque de contracter la COVID-19 que les employés du secteur des services à la clientèle (du moins sur la période d'étude, en 2022).

Le secteur de l'enseignement présente une conclusion similaire à celui de la santé, quoique la différence soit d'envergure beaucoup plus marquée. En effet, malgré un risque non-différent du secteur 4, les enseignants perçoivent leur risque d'infection comme étant plus élevé par 7.06 points de pourcentage. L'envergure considérable de ce biais fait en sorte que son analyse se doit d'être approfondie à la section 5.3.1.

Il ressort de ce survol des résultats plusieurs biais, certains mineurs et d'autres importants. Le fait que la plupart de ces biais soient positifs et significatifs peut relever du fait que les gens ont tendance à surestimer la probabilité d'événements à faible occurrence. Il est également possible que la peur d'être infectés par la COVID-19 ait influencé la perception du risque des répondants à la hausse. Par soucis de concision, il n'est possible d'approfondir chacun des biais identifiés. La section d'analyse subséquente se concentre donc sur les biais qui ont été ciblés comme étant les plus intéressants, soit ceux liés au secteur de l'enseignement et au vaccin.

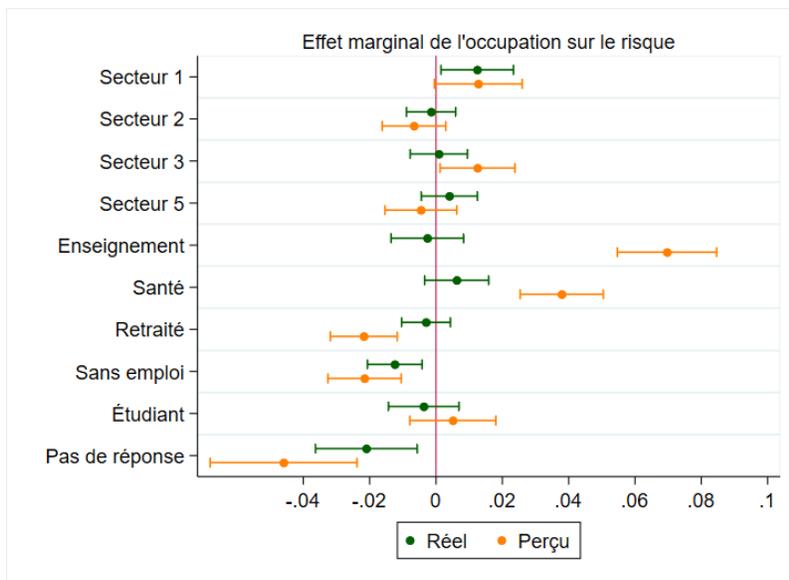


Figure 5.3 : Comparaison de l'effet marginal de l'occupation principale sur le risque réel et perçu

Table 5.1 : Résultats des régressions sur la probabilité subjective et objective (en points de pourcentage)

Variable	(1) Probabilité subjective ($\bar{P} = 17.6\%$)		(2) Probabilité objective ($\bar{Y} = 3.94\%$)		(3) Test de Wald ($\beta_{j,s} - \beta_{j,o} = 0$)	
	β_s	SE (robuste)	β_o	SE (robuste)	χ^2	$P > \chi^2$
Sexe						
Femme	0.202	0.227	-0.394*	0.17	11.156	<0.001***
Âge						
18-29 ans	4.250***	0.44	2.850***	0.344	16.817	<0.001***
30-39 ans	2.306***	0.427	1.260***	0.325	9.748	0.001**
40-49 ans	1.094**	0.404	0.391	0.295	4.713	0.029*
60-69 ans	-2.402***	0.402	-0.617*	0.276	29.381	<0.001***
70 ans et plus	-4.727***	0.468	-0.963**	0.316	97.290	<0.001***
Revenu						
Pas de réponse	-2.086***	0.41	-0.755**	0.291	16.077	<0.001***
<20k	-0.368	0.527	0.520	0.398	4.652	0.031*
20-40k	-0.822*	0.379	0.211	0.284	12.128	<0.001***
60-80k	0.007	0.396	0.350	0.302	1.239	0.266
80-100k	0.086	0.415	-0.431	0.312	2.481	0.115
>100k	-1.043**	0.38	-0.559	0.288	2.609	0.106
Éducation						
Pas de réponse	-1.688	1.15	-0.155	0.774	2.692	0.101
Primaire	1.239	1.098	0.777	0.898	0.329	0.566
Secondaire	0.246	0.239	0.228	0.176	0.009	0.924
Universitaire	0.663*	0.323	0.554*	0.25	0.188	0.665
Ménage						
Enfant	3.801***	0.393	2.403***	0.318	21.997	<0.001***
Pas de réponse	1.378	1.542	2.814*	1.20	1.506	0.220
1 personne	-0.617*	0.295	-0.329	0.215	1.506	0.220
3 personnes	-0.501	0.385	-0.185	0.297	1.112	0.292
4 personnes	-0.926*	0.445	-0.955**	0.352	0.007	0.933
5 personnes +	-0.679	0.66	0.322	0.542	4.054	0.044*
Région						
Québec	0.419	0.375	0.163	0.292	0.786	0.375
Autres régions	0.834**	0.24	0.106	0.18	14.755	<0.001***
Vaccin						

Vacciné	7.993***	0.462	-0.737	0.396	914.301	<0.001***
Pas de réponse	-0.251	0.91	-1.382	0.707	3.079	0.079
Secteur						
Secteur 1	1.404*	0.668	1.306*	0.554	0.040	0.841
Secteur 2	-0.750	0.474	-0.195	0.368	2.306	0.129
Secteur 3	1.317*	0.564	0.058	0.433	8.048	0.005**
Secteur 5	-0.468	0.541	0.399	0.424	4.434	0.035*
Secteur 6	7.055***	0.74	-0.234	0.542	137.469	<0.001***
Secteur 7	3.744***	0.622	0.517	0.48	42.237	<0.001***
Secteur 8	-2.178***	0.506	-0.315	0.369	21.644	<0.001***
Secteur 9	-2.138***	0.553	-1.231**	0.414	4.355	0.037*
Secteur 10	0.601	0.677	-0.092	0.553	1.842	0.175
Pas de réponse	-4.574***	1.125	-2.090**	0.779	7.983	0.005**
Constantes hebdo						
Semaine 1	6.573***	0.671	3.709***	0.518	29.892	<0.001***
Semaine 2	4.807***	0.65	2.159***	0.474	25.616	<0.001***
Semaine 3	2.555***	0.638	2.003***	0.47	1.211	0.271
Semaine 4	1.352*	0.62	0.838	0.44	1.055	0.304
Semaine 5	1.652**	0.62	1.210**	0.44	0.787	0.375
Semaine 6	-0.811	0.592	0.021	0.406	2.983	0.084
Semaine 7	-0.552	0.598	0.107	0.409	1.831	0.176
Semaine 8	-2.051**	0.589	0.199	0.407	22.541	<0.001***
Semaine 9	-1.356*	0.591	-0.295	0.395	4.781	0.029*
Semaine 10	-1.605**	0.594	0.049	0.411	11.804	<0.001***
Semaine 11	1.861**	0.631	0.923*	0.455	3.461	0.063
Semaine 12	6.405***	0.662	2.146***	0.498	65.040	<0.001***
Semaine 13	8.090***	0.688	3.772***	0.541	64.241	<0.001***
Semaine 14	7.098***	0.674	3.153***	0.514	54.745	<0.001***
Semaine 15	5.074***	0.668	3.081***	0.508	14.644	<0.001***
Semaine 16	2.996***	0.639	1.773***	0.467	5.821	0.016*
Semaine 17	1.684**	0.631	1.283**	0.458	0.650	0.420
Constante	7.110***	0.824	2.488***	0.634		
Observations	54 155		54 155			
χ^2	3250.59		833.71			

*p<0.05, **p<0.01, ***p<0.001

5.1 Comparaison entre MCO et probit

Avant de progresser dans l'analyse, il convient de justifier l'usage du modèle de régression linéaire appliqué à la variable dépendante limitée du risque objectif (Y_i). À cette fin, la table 5.2 est présentée, où peuvent être lus les résultats de deux régressions. Dans la première colonne sont présentés les effets marginaux d'une régression probit sur le vecteur de variables explicatives utilisé et présenté plus haut. Plus spécifiquement, la régression estimée est de la forme suivante, où Φ réfère à la cumulée de la loi normale :

$$Pr(Y_i = 1 | X_i) = \Phi\left(\frac{X_i\beta_o}{\sigma}\right) \quad (14)$$

Définissons deux réalisations du vecteur X_i . La première, dénotée X_{1i} , a la caractéristique $x_{ij} = 1$ et maintient toutes les autres variables à une valeur fixe. Le deuxième (X_{0i}) a la caractéristique $x_{ij} = 0$, maintenant toujours les autres variables à une valeur fixe. Les effets marginaux (EM) pour variables explicatives discrètes de la régression probit ont alors été estimés suivant cette forme :

$$EM_j = \Phi\left(\frac{X_{1i}\beta_o}{\sigma}\right) - \Phi\left(\frac{X_{0i}\beta_o}{\sigma}\right) \quad (15)$$

Ces effets marginaux sont présentés à la première (1) colonne de la table 5.2. Ceux-ci ont à être comparés aux coefficients obtenus d'une estimation linéaire, ici calculée par MCO. Rappelons que puisque que le vecteur de variables explicatives est identique entre les régressions sur les deux mesures de risque, les coefficients de la régression sur la probabilité objective estimés par MCO sont équivalents à ceux obtenus par la méthode des *Feasible Generalized Least Squares* employée par le modèle de *Seamingly Unrelated Regression*.

À travers la juxtaposition des résultats, il peut être vu que la plupart des coefficients sont légèrement différents d'une colonne à l'autre. Les changements sont de petites amplitudes, parfois à la hausse et parfois à la baisse. Il en revient au lecteur d'accepter le coût qu'entraîne l'estimation linéaire (une légère perte de précision) en vue de ses bénéfices (l'opportunité de tester statistiquement la différence des coefficients).

Table 5.2 : Comparaison entre estimations Probit et MCO (en points de pourcentage)

	(1)		(2)	
	Probit		MCO	
Variable	Effets marginaux	SE	β_0	SE (robuste)
Sexe				
Femme	-0.481**	0.174	-0.394*	0.193
Âge				
18-29 ans	2.759***	0.404	2.850***	0.401
30-39 ans	1.432***	0.354	1.260***	0.352
40-49 ans	0.679*	0.325	0.391	0.318
60-69 ans	-0.723*	0.323	-0.617*	0.294
70 ans et plus	-1.366***	0.353	-0.963**	0.336
Revenu				
Pas de réponse	-0.894**	0.308	-0.755*	0.341
<20k	0.332	0.454	0.520	0.518
20-40k	0.333	0.333	0.211	0.334
60-80k	0.287	0.312	0.350	0.338
80-100k	-0.216	0.299	-0.431	0.347
>100k	-0.602*	0.264	-0.559	0.325
Éducation				
Pas de réponse	-0.398	1.267	-0.155	1.583
Primaire	1.517	1.178	0.777	1.056
Secondaire	0.054	0.231	0.228	0.242
Universitaire	0.399*	0.203	0.554*	0.216
Ménage				
Enfant	1.724***	0.307	2.403***	0.403
Pas de réponse	0.775	1.346	2.814	2.146
1 personne	-0.372	0.232	-0.329	0.234
3 personnes	-0.097	0.271	-0.185	0.364
4 personnes	-0.299	0.293	-0.955*	0.425
5 personnes +	0.752	0.46	0.322	0.64
Région				
Québec	0.298	0.303	0.163	0.348
Autres régions	0.099	0.179	0.106	0.198
Vaccin				
Vacciné	-0.688	0.371	-0.737	0.466
Pas de réponse	-1.478**	0.563	-1.382	0.921
Secteur				

Secteur 1	1.289*	0.54	1.306*	0.646
Secteur 2	-0.144	0.331	-0.195	0.419
Secteur 3	0.037	0.382	0.058	0.478
Secteur 5	0.252	0.413	0.399	0.507
Secteur 6	-0.169	0.42	-0.234	0.542
Secteur 7	0.656	0.444	0.517	0.518
Secteur 8	-0.385	0.406	-0.315	0.418
Secteur 9	-0.968*	0.372	-1.231*	0.504
Secteur 10	-0.248	0.433	-0.092	0.730
Pas de réponse	-1.688*	0.811	-2.090*	0.942
Constantes hebdo				
Semaine 1	3.818***	0.813	3.709***	0.564
Semaine 2	2.046**	0.701	2.159***	0.517
Semaine 3	1.874**	0.689	2.003***	0.523
Semaine 4	0.729	0.609	0.838	0.467
Semaine 5	0.884	0.623	1.210*	0.489
Semaine 6	-0.508	0.513	0.021	0.44
Semaine 7	-0.403	0.522	0.107	0.432
Semaine 8	-0.413	0.523	0.199	0.449
Semaine 9	-0.834	0.487	-0.295	0.423
Semaine 10	-0.315	0.529	0.049	0.429
Semaine 11	1.330*	0.652	0.923*	0.46
Semaine 12	2.939***	0.758	2.146***	0.505
Semaine 13	4.782***	0.871	3.772***	0.557
Semaine 14	3.728***	0.809	3.153***	0.554
Semaine 15	3.340***	0.782	3.081***	0.567
Semaine 16	1.825**	0.686	1.773***	0.508
Semaine 17	1.489*	0.662	1.283**	0.484
Constante			2.488***	0.711
Observations	54 155		54 155	
χ^2	791.05			

*p<0.05, **p<0.01, ***p<0.001

5.2 Analyse de sensibilité

Cette section est proposée en réponse au problème de l'incertitude épistémique. Afin d'évaluer la robustesse des résultats de ce mémoire, une régression estimant la relation entre la probabilité subjective (P_i) et le vecteur de variables explicatives (X_i) est calculée, mais cette fois-ci en excluant de la population à l'étude les répondants ayant estimé leur risque comme étant égal à 50%. Ces nouveaux coefficients

peuvent ensuite être comparés à ceux de la régression sur l'échantillon entier. La colonne (1) de la table 5.3 réfère donc à l'estimation de l'équation (16) alors que la colonne (2) compile les résultats de l'équation (17). Les deux régressions sont estimées par la méthode des MCO avec écart-types robustes.

$$P_i = X_i\beta_s + \varepsilon_{i,s} \forall i \quad (16)$$

$$P_i = X_i\beta_s + \varepsilon_{i,s} \forall i \text{ tq } P_i \neq 0.5 \quad (17)$$

Rappelons premièrement que retirer tous les répondants ayant fourni une estimation du risque à 50% est sans doute une surcorrection du problème engendré par l'incertitude épistémique. Il convient alors de nuancer les coefficients obtenus ici avec ceux présentés précédemment sur l'échantillon entier. Ceci étant dit, quelques résultats intéressants ressortent de la table 5.3. Afin de maintenir concise cette analyse, seuls les résultats relatifs aux principaux biais sont abordés. Il peut premièrement être vu que le coefficient relatif à l'effet du sexe sur la perception du risque devient significatif. Ceci est probablement causé par le fait que la probabilité 50 est employée par plus de femmes que d'homme dans notre échantillon (57% des probabilités 50 recensées ont été fournies par des femmes). Le coefficient corrigé est significatif et négatif, suggérant que les femmes ont mieux incorporé leur diminution du risque réel dans leur estimation du risque que ne le laissait entendre les résultats précédents.

Quant aux trois biais principaux identifiés par ce mémoire (celui de l'âge, du secteur de l'enseignement et du vaccin), il peut être vu qu'ils sont affectés différemment. Il est premièrement intéressant de voir qu'en restreignant l'échantillon, le biais de l'âge s'accroît pour les deux plus jeunes groupes alors qu'il diminue pour les deux plus âgés. En testant la différence entre ces nouveaux coefficients relatifs à l'âge (β_s) et ceux sur le risque réel (β_o), il est obtenu que chacune des différences demeure significative à 1%. Ceci suggère que le biais réel qu'entretenaient les plus jeunes répondants est d'envergure supérieure à celui estimé par les régressions sur l'échantillon entier, alors que celui des répondants plus âgés serait plus faible. Cette nouvelle dynamique propre à l'échantillon tronqué contribue à justifier l'importance d'une analyse future plus complète relative au biais de l'âge.

L'effet sur la perception du risque du secteur de l'enseignement diminue quant à lui de 1.59 points de pourcentage en restreignant l'échantillon, alors que celui du vaccin diminue de 2.72 points de pourcentage. Les deux tests de Wald estimant l'hypothèse d'égalité des coefficients ($\beta_s - \beta_o$) permettent de confirmer au degré de significativité de 0.1% que les biais demeurent différents de 0.

Il est possible d'effectuer un exercice similaire, cette fois-ci en retirant de l'échantillon les répondants ayant la COVID-19 au moment de répondre. Ceux-ci ayant été codés comme ayant une perception du risque à 100%, il est possible qu'ils viennent biaiser à la hausse les résultats. Il est obtenu de cette nouvelle régression que les biais principaux persistent. Chacun des groupes d'âge demeurent significatifs, quoique les coefficients soient d'amplitude inférieure à ceux de la régression sur l'échantillon entier. Il est intéressant de voir que le coefficient relatif au statut vaccinal augmente à 9.12 points de pourcentage (p-value <.001) en restreignant l'échantillon. Quant au coefficient relatif au secteur de l'enseignement, il demeure presque inchangé, passant de 7.06 à 7.09 points de pourcentage (p-value <0.001). Ces différences confirment que les principaux résultats de ce mémoire sont robustes au codage du risque subjectif des répondants infectés.

Table 5.3 : Résultats des régressions sur la probabilité subjective (en points de pourcentage)

Variable	(1) Échantillon entier ($\bar{P} = 17.6\%$)		(2) Échantillon tronqué ($P_i \neq 0.5$) ($\bar{P} = 13.2\%$)	
	β_s	SE (robuste)	β_s	SE (robuste)
Sexe				
Femme	0.202	0.227	-0.655**	0.23
Âge				
18-29 ans	4.250***	0.44	6.064***	0.459
30-39 ans	2.306***	0.427	2.895***	0.443
40-49 ans	1.094**	0.404	1.211**	0.415
60-69 ans	-2.402***	0.402	-1.694***	0.394
70 ans et plus	-4.727***	0.468	-3.218***	0.45
Revenu				
Pas de réponse	-2.086***	0.41	-2.240***	0.405
<20k	-0.368	0.527	-0.427	0.534
20-40k	-0.822*	0.379	-0.704	0.381
60-80k	0.007	0.396	0.037	0.404
80-100k	0.086	0.415	-0.457	0.428
>100k	-1.043**	0.38	-0.912*	0.391
Éducation				
Pas de réponse	-1.688	1.15	-2.046	1.102
Primaire	1.239	1.098	0.090	1.127
Secondaire	0.246	0.239	-0.403	0.241
Universitaire	0.663*	0.323	1.130*	0.33
Ménage				

Enfant	3.801***	0.393	3.519***	0.416
Pas de réponse	1.378	1.542	1.715	1.557
1 personne	-0.617*	0.295	-0.098	0.293
3 personnes	-0.501	0.385	-0.169	0.4
4 personnes	-0.926*	0.445	-0.560	0.471
5 personnes +	-0.679	0.66	0.165	0.705
Région				
Québec	0.419	0.375	0.598	0.387
Autres régions	0.834**	0.24	0.668**	0.244
Vaccin				
Vacciné	7.993***	0.462	5.278***	0.461
Pas de réponse	-0.251	0.91	-1.347	0.244
Secteur				
Secteur 1	1.404*	0.668	1.609*	0.7
Secteur 2	-0.750	0.474	-0.973*	0.486
Secteur 3	1.317*	0.564	0.656	0.591
Secteur 5	-0.468	0.541	-0.549	0.554
Secteur 6	7.055***	0.74	5.461***	0.83
Secteur 7	3.744***	0.622	2.734***	0.671
Secteur 8	-2.178***	0.506	-1.190*	0.503
Secteur 9	-2.138***	0.553	-2.168***	0.561
Secteur 10	0.601	0.677	0.585	0.708
Pas de réponse	-4.574***	1.125	-4.832***	1.032
Constantes hebdo				
Semaine 1	6.573***	0.671	6.686***	0.695
Semaine 2	4.807***	0.65	5.001***	0.66
Semaine 3	2.555***	0.638	2.882***	0.638
Semaine 4	1.352*	0.62	1.747**	0.612
Semaine 5	1.652**	0.62	2.217***	0.61
Semaine 6	-0.811	0.592	-0.378	0.568
Semaine 7	-0.552	0.598	0.318	0.577
Semaine 8	-2.051**	0.589	-0.913	0.564
Semaine 9	-1.356*	0.591	-0.611	0.567
Semaine 10	-1.605**	0.594	-0.662	0.574
Semaine 11	1.861**	0.631	2.219***	0.625
Semaine 12	6.405***	0.662	5.501***	0.685
Semaine 13	8.090***	0.688	7.752***	0.724
Semaine 14	7.098***	0.674	6.386***	0.704
Semaine 15	5.074***	0.668	5.082***	0.682

Semaine 16	2.996***	0.639	2.711***	0.638
Semaine 17	1.684**	0.631	2.003**	0.622
Constante	7.110***	0.824	4.797***	0.826
Observations	54 155		47 560	

*p<0.05, **p<0.01, ***p<0.001

5.3 Analyse des biais principaux

5.3.1 Biais associé au secteur de l'enseignement

Une manière d'approfondir notre étude du biais du secteur de l'enseignement est de profiter de l'aspect temporel des données. Nous avons jusqu'à maintenant utilisé ces dernières de manière transversale, mais il est également possible d'obtenir l'effet marginal capté par ce secteur sur le risque perçu et réel de manière hebdomadaire. Cette décomposition temporelle permet de suivre l'évolution de chacune des mesures de risque et du biais total.

Rappelons que la collecte de données s'est échelonnée de manière hebdomadaire du 13 janvier au 17 mai. Au Québec, il y a eu réouverture des écoles primaires et secondaires le 17 janvier 2022. Le port du masque y était obligatoire. Les cégeps et les universités ont quant à eux dû assurer une réouverture avant le 31 janvier. L'échantillon à l'étude commence par une collecte effectuée du 13 au 18 janvier. Il est donc possible que certains des retours en classe de cégeps et d'universités aient été capturés par la deuxième et troisième semaine de l'échantillon, mais de manière générale l'enseignement sur la période de référence s'est accompli en présentiel.

Débutons ce survol temporel par une simple comparaison de la perception du risque hebdomadaire entre les enseignants et l'échantillon entier. Cette dernière est résumée à la **Figure 5.6**. On y voit qu'à chacune des semaines, la perception du risque est supérieure pour les employés du secteur de l'enseignement que pour l'échantillon représentatif. La tendance générale de la courbe des enseignants semble suivre celle de l'échantillon entier, mais avec des hausses menant à des sommets particulièrement marqués. La seule dynamique propre à la perception du risque des enseignants qui ressort est la hausse isolée du risque perçu à la semaine 8. Cette anomalie est sans doute expliquée par le fait qu'elle corresponde à l'estimation subjective collectée entre le 3 et 8 mars 2022. Les masques en classe pour les écoles primaires et secondaires de la province de Québec ont été retirés le 7 mars¹⁵. Il n'est donc pas surprenant qu'à l'approche de cette date, la perception du risque des enseignants aient fortement augmentée. Cette

¹⁵ Compilation Spéciale de l'INSPQ (2022, 5 octobre). « Ligne du temps COVID-19 au Québec », INSPQ. Récupéré le 20 février 2023 de <https://www.inspq.qc.ca/COVID-19/donnees/ligne-du-temps>

hausse n'a toutefois été que momentanée, la mesure retombant à sa valeur précédente la semaine suivante.

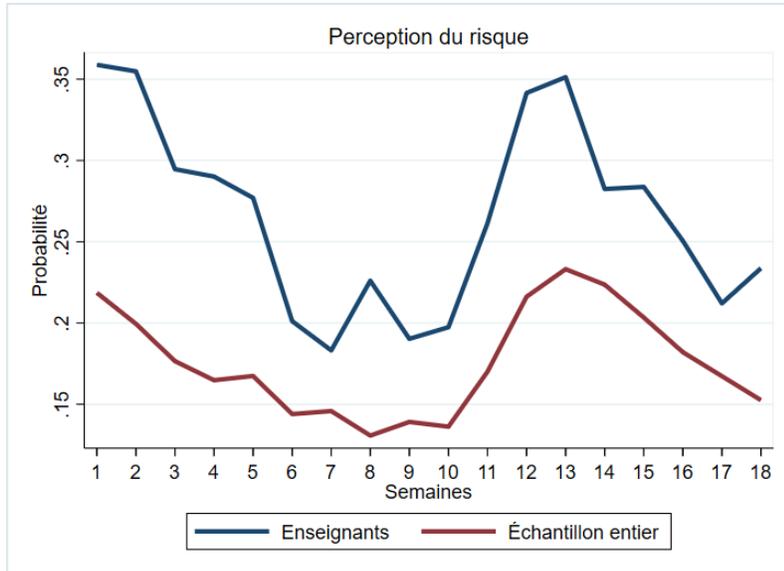


Figure 5.4 : Comparaison de l'évolution du risque perçu entre les employés du secteur de l'éducation et l'échantillon entier

Il vient maintenant temps de s'intéresser à la décomposition temporelle de l'effet marginal de travailler dans le secteur de l'enseignement sur les mesures de risque. Rappelons que la variable de risque réel provient d'une question demandant si le répondant a été infecté par la COVID-19 dans la semaine précédente, alors que la question sur le risque perçu est posée relativement à la semaine à venir. Afin d'obtenir une mesure adéquate du biais, il convient de faire concorder temporellement ces données. À cette fin, la mesure du risque objectif d'une semaine t sera comparée à la mesure du risque subjectif collectée à la semaine $t-1$. Nous perdons donc l'équivalent d'une semaine de données dans le processus. 17 régressions seront estimées ici, une pour chacune des semaines restantes. Le vecteur de variables explicatives est équivalent à celui utilisé depuis le début du mémoire, la seule différence étant que les constantes relatives à chaque semaine ont été retirées. Le rôle de ces dernières étant de compenser pour l'évolution temporelle de la pandémie, elles deviennent inutiles dans un contexte de régressions hebdomadaires. Voici alors les trois équations qui sont estimées et présentées visuellement plus bas. La variable DNP_t correspond à une estimation de la différence nette de probabilités (effet marginal sur le

risque subjectif moins l'effet marginal sur le risque objectif) qu'à le secteur de l'enseignement sur le biais des enseignants.

$$P_{i,t-1} = X_{i,t-1}\beta_{s,t-1} + \varepsilon_{i,s,t-1} \quad (18)$$

$$Y_{i,t} = X_{i,t}\beta_{o,t} + \varepsilon_{i,o,t} \quad (19)$$

$$DNP_t = \beta_{s,t-1} - \beta_{o,t} \quad (20)$$

Trois séries d'intérêt sont obtenues de cette décomposition temporelle. Premièrement, il y a la valeur que prend le coefficient relatif à l'enseignement à chacune des semaines lorsque régressé sur la probabilité subjective (équation 18). Ce dernier peut être interprété comme étant l'effet marginal capté par le fait de travailler dans le secteur de l'enseignement sur la perception du risque à la semaine t. Ensuite, une mesure similaire est obtenue pour l'effet marginal sur le risque objectif d'attraper la COVID-19 (équation 19). Finalement, une différence est calculée entre les deux coefficients, de manière à obtenir une estimation hebdomadaire de l'impact net qu'a le secteur de l'enseignement sur le biais de perception du répondant (équation 20).

La **Figure 5.7** illustre l'effet marginal du secteur de l'enseignement sur la probabilité subjective d'attraper la COVID-19. Chacune des lignes correspond à une semaine de collecte de données. On y voit que cet effet suit l'évolution des cas de COVID-19 sur la période, avec un sommet dans les premières semaines suivi d'une vague plafonnant à la 13^{ème} semaine. Il est également possible d'y voir que pour la majorité des semaines, les employés du secteur de l'enseignement avaient une perception du risque significativement supérieure au groupe de référence. La hausse de la semaine 8 abordée précédemment est également bien visible sur la figure. Seul le 14^{ème} point semble dévier de la tendance générale. Ce point correspond à la collecte effectuée durant le congé de Pâques, vers la fin de la session universitaire. Il est possible que ces événements aient contribué à la faible perception du risque qui y est recensée.

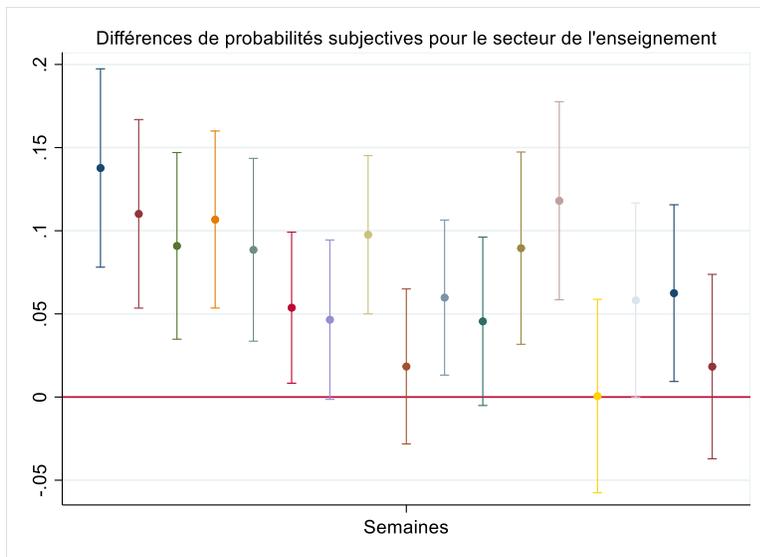


Figure 5.5 : Effet marginal capté par le secteur de l'enseignement sur la perception du risque hebdomadaire

La **Figure 5.8** illustre l'effet marginal du secteur de l'enseignement, mais cette fois-ci sur le risque réel que courent ses employés d'attraper la COVID-19. On y voit que le risque n'est significativement supérieur à 0 à aucune semaine, suggérant que les employés du secteur de l'enseignement n'ont en 2022 jamais fait face à plus de chance d'attraper la COVID-19 que les employés du secteur 4. Quant à l'estimation de la semaine 9 (8^{ème} ligne sur le graphique) qui semble diverger de la tendance, celle-ci correspond à la collecte de données effectuée entre le 10 et 15 mars. Il n'est pas surprenant d'obtenir la plus faible mesure du risque objectif entre ces dates puisqu'elles correspondent à la semaine suivant la semaine de relâche, durant laquelle la contamination au travail des enseignants a dû être à son minimum.

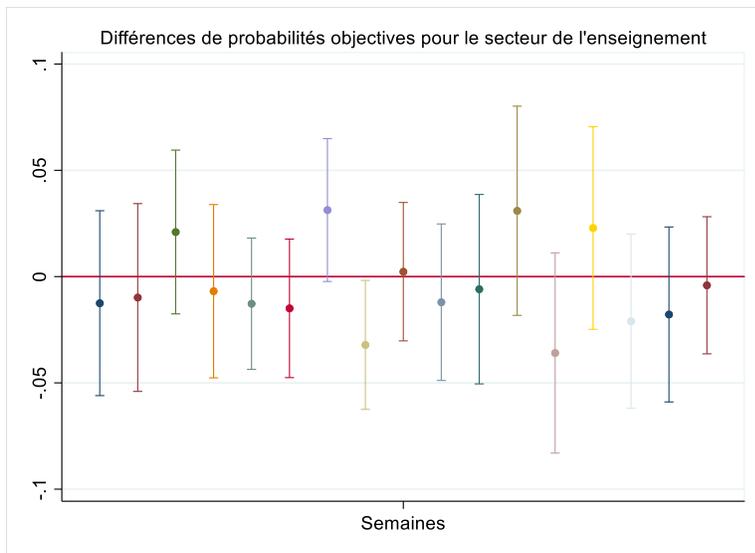


Figure 5.6 : Effet marginal capté par le secteur de l'enseignement sur le risque réel hebdomadaire

La figure suivante (5.9) illustre la différence hebdomadaire entre les deux coefficients illustrés plus haut. Ceci représente une estimation de l'effet net qu'a le secteur de l'éducation sur le biais de perception du répondant. On y voit que pour la majorité des semaines, le biais est significativement supérieur à celui du secteur de référence (4). Le biais crée par le secteur de l'enseignement varie entre 0 et 16%. Il semblerait également que, à l'exception des semaines 8 et 14, l'évolution du biais suit l'évolution des cas de COVID-19. En effet, on retrouve une pente descendant d'un sommet dans les premières semaines, suivie d'une vague entourant la treizième semaine. Ceci semble indiquer que plus il y a de cas de COVID-19 recensés à l'échelle sociétale, plus l'estimation du risque des enseignants est biaisée à la hausse.

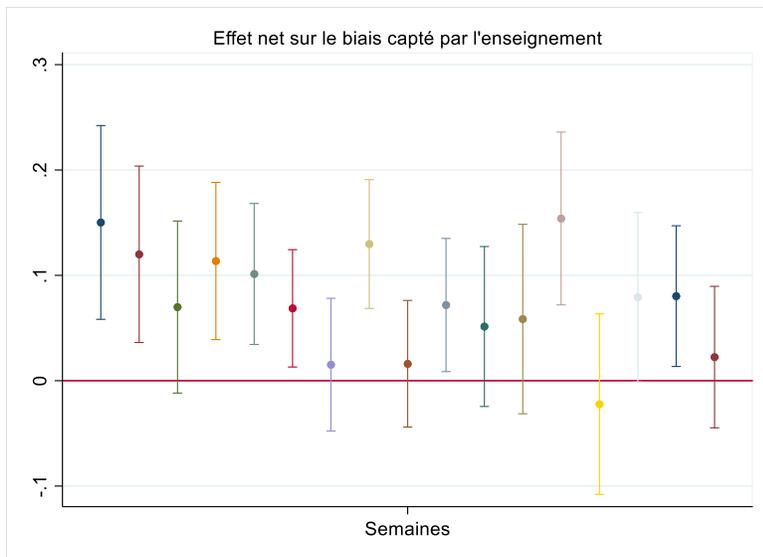


Figure 5.7 : Estimation du biais de perception hebdomadaire capté par le secteur de l'enseignement

La tendance générale du biais du secteur de l'enseignement semble indiquer que ses employés percevaient un risque beaucoup plus important que ceux des autres secteurs de l'économie, sans toutefois que ce risque ait été motivé par un plus haut risque réel d'être infecté par la COVID-19. Il est possible que les enseignants aient fortement surréagis au risque que présente le contact humain en classe. Les estimations erronées de ce groupe en ont alors fait l'un des plus biaisés de l'échantillon.

5.3.2 Biais associé au statut vaccinal

L'autre caractéristique identifiée par cette analyse comme étant responsable d'un fort biais est le statut vaccinal. La **Figure 5.10** illustre la probabilité subjective hebdomadaire moyenne des populations de vaccinés et non-vaccinés. Il y est vu qu'à l'exception de la semaine 5 de collecte, la perception moyenne du risque des vaccinés est supérieure à celle des non-vaccinés. La différence entre les deux courbes semble suivre une tendance assez similaire d'une semaine à l'autre. Le fait que la courbe de perception du risque des vaccinés paraisse beaucoup plus lisse peut être la conséquence du nombre plus important de répondants tombant dans cette catégorie.

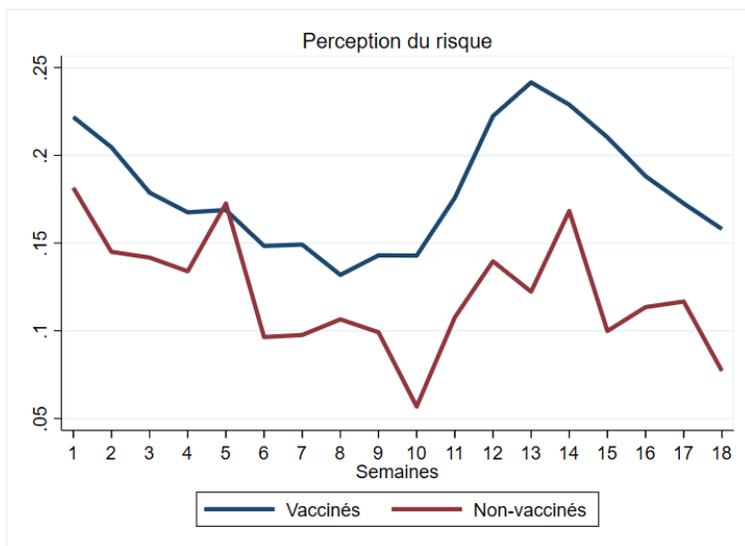


Figure 5.8 : Comparaison de probabilité subjective entre les vaccinés et non-vaccinés

En effectuant la même décomposition temporelle que pour les enseignants, on obtient les trois figures suivantes. Elles illustrent premièrement l'effet marginal d'être vacciné sur la perception du risque (**Figure 5.11**) et sur le risque réel (**Figure 5.12**) à la semaine t . Il peut y être vu que le fait d'être vacciné augmente la perception du risque au-dessus du groupe de référence (les non-vaccinés) à presque chacune des semaines, avec une hausse particulièrement marquée lors de la vague entourant la semaine 13. Quant à l'effet marginal du vaccin sur le risque objectif, il est plus faible (voir négatif) dans les premières semaines, et oscille autour de 0 par la suite. Cette différence temporelle pourrait être le résultat de la prévalence croissante du variant Omicron, ou encore la conséquence de la distribution de nouvelles doses de vaccin durant la collecte de données. Le taux de l'échantillon représentatif étant passé de 2 à 3 doses de vaccin augmente de près de 20 points de pourcentage dans les 5 premières semaines de l'étude. La tendance générale de la **Figure 5.12** suggère tout de même que les personnes vaccinés et non-vaccinés faisaient face à un risque d'infection similaire, surtout lors des 9 dernières semaines de la collecte de données.

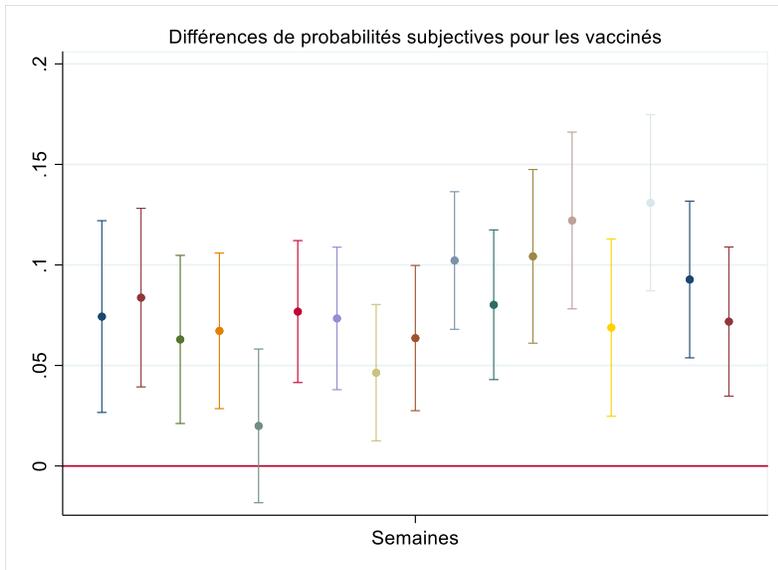


Figure 5.9 : Effet marginal capté par le fait d’être vacciné sur la perception du risque hebdomadaire

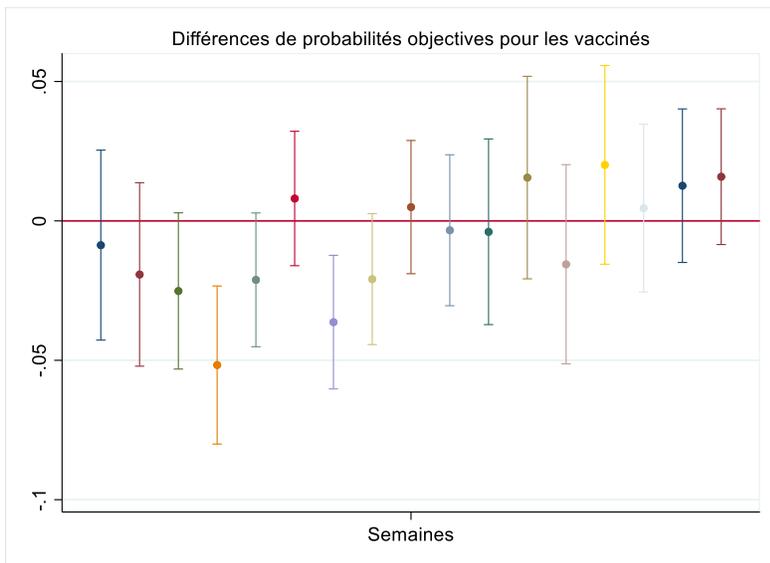


Figure 5.10 : Effet marginal capté par le fait d’être vacciné sur le risque objectif hebdomadaire

L’effet net du vaccin sur le biais total peut être vu à la **Figure 5.13**. Il ne semble pas suivre une tendance déterminée. Il semble au contraire être relativement constant dans le temps, et significatif à presque chacune des semaines. Rappelons que l’effet net du secteur de l’enseignement sur le biais suivait

l'évolution de la pandémie. Ceci pourrait suggérer que les enseignants ont surestimé le risque que leur présentait une vague de cas. Il est intéressant de noter que l'effet net du vaccin sur le biais de la population des vaccinés ne semble pas suivre aussi fidèlement cette tendance. Ceci pourrait indiquer qu'il existe une différence sous-jacente entre personnes vaccinées et non-vaccinées dans la formation des attentes. Cette différence aurait été plus ou moins constante au fil des semaines de collecte de données.

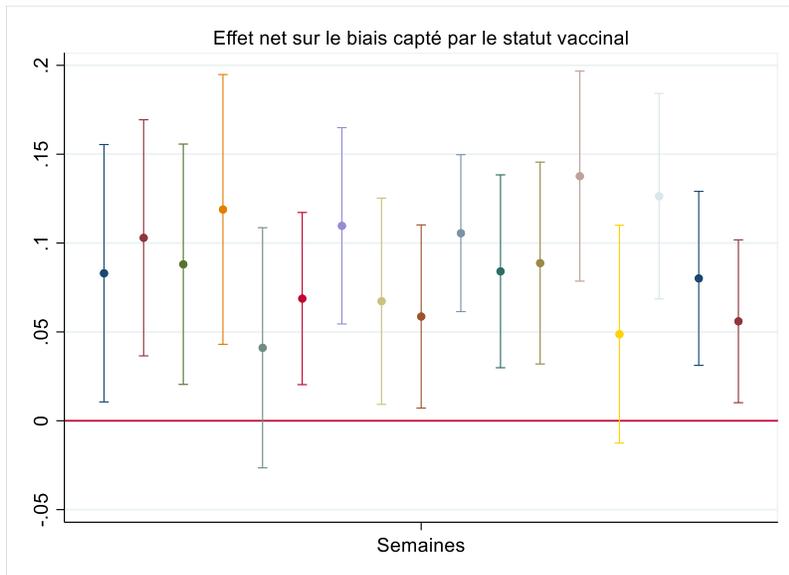


Figure 5.11 : Biais de perception hebdomadaire capté par le fait d'être vacciné

Le biais présenté ici est d'envergure considérable. L'effet marginal du vaccin sur le biais varie d'une semaine à l'autre d'un minimum de 4.11 (semaine 5) à 13.77 points de pourcentage (semaine 13). Afin de mettre ce résultat en perspective, rappelons que le biais moyen sur l'échantillon n'étant pas infecté au moment de répondre est de $(\bar{P} - \bar{Y} = 13.9\% - 3.94\%)$ est de 9.96%. On voit donc qu'une proportion énorme du biais total de l'individu moyen est captée par la variable indicatrice du vaccin. Il est conséquemment particulièrement intéressant de creuser cette relation car elle semble être responsable d'une déviation très importante de l'hypothèse de rationalité dans la formation des attentes.

5.3.3 Sources d'information

La première question qui doit être approfondie se doit de déterminer si le vaccin est responsable de ce biais, où s'il n'est pas plutôt une variable captant une différence non-mesurée entre les vaccinés et non-vaccinés. Il est très possible que ces deux sous-populations agissent de manière hétérogène d'une

manière qui n'est pas mesurée par les autres variables à l'étude. La variable du vaccin ne serait alors qu'un proxy capturant une différence plus fondamentale entre individus.

Considérons premièrement le fait que le biais total induit par le vaccin est largement causé par des différences en termes de perception du risque. Une explication possible serait alors qu'il existe une hétérogénéité entre les vaccinés et non-vaccinés quant à la formation de l'estimation du risque. Cette différence pourrait par exemple être relative à l'intérêt qui est porté à la recherche d'information sur l'évolution de la pandémie. Il serait plausible qu'une différence importante entre les deux groupes soit en lien avec la quantité (et la qualité) de l'information médiatique consommée. Rappelons ici les résultats de Leigh et al. (2022), qui trouvent que les personnes moins éduquées (identifiées dans notre échantillon comme étant proportionnellement fortement non-vaccinées) consomment moins de sources d'information distinctes en lien avec la COVID-19 et sont plus susceptibles de se fier à commentaires d'amis et famille sur les réseaux sociaux. À l'inverse, les personnes vaccinées consomment plusieurs sources médiatiques jugées comme crédibles. Des raccourcis cognitifs tel que l'heuristique de disponibilité (*availability heuristic*) nous rappellent que la perception du risque est susceptible d'être biaisée par la facilité avec laquelle de l'information pertinente à son estimation peut être ramenée en mémoire¹⁶. Ainsi, il ne serait pas surprenant qu'une population suivant de manière journalière l'évolution de la COVID-19 ait des perceptions du risque biaisées à la hausse. Ajoutant à cela le discours incessant et parfois alarmiste qui a entouré la pandémie, il serait très concevable que la consommation médiatique ait entraîné un biais considérable.

Afin de tester cette hypothèse, de nouvelles données ont été introduites à notre analyse. Ces dernières ont été obtenues sur le site de l'institut national de santé publique du Québec (INSPQ). Trois séries de données ont été ajoutées : une sur le nombre de nouveaux cas confirmés de COVID-19 (tel qu'estimé par résultats de tests PCR), une sur le nombre total de nouvelles hospitalisations liées à la COVID-19 et une sur le nombre de nouveaux décès liés à la COVID-19. Afin de palier au fait que l'usage de tests PCR ait été restreint aux individus prioritaires en 2022, une autre série sur le nombre de nouveaux cas de COVID-19 a été introduites, cette fois-ci en provenance du CIRANO. Ce dernier a fourni en 2022 une estimation hebdomadaire du nombre de personnes infectées par la COVID-19 au Québec. Cette estimation s'est trouvée à être beaucoup plus proche du nombre de cas réel que celle des données PCR, et a ainsi été publiée à maintes reprises dans les médias. Pour chacune de ces séries, une moyenne hebdomadaire a

¹⁶ Kassiani Nikolopoulou (2022, 7 décembre). « The Availability Heuristic », Scribbr. Récupéré le 7 décembre 2022 de <https://www.scribbr.com/research-bias/availability-heuristic/>

été calculée, et ce de manière à faire concorder les moyennes mobiles avec les dates de collectes de notre base de données initiale.

Afin de tester le plus formellement possible cette hypothèse, de nouveaux contrôles sont introduits à la liste de variables explicatives (X_i) présentée précédemment. Sont désormais incluses toutes les variables relatives au réseau de l'individu. Plus précisément, on se trouve à contrôler pour le nombre de personnes dans le réseau du répondant :

- ayant reçu un test positif de COVID-19 dans la dernière semaine
- ayant été en confinement dans la dernière semaine
- étant médecin
- n'étant pas vacciné
- résidant dans une résidence pour personnes âgées (RPA, CHSLD, RI-RTF)

L'exercice consiste alors à tenter de définir à quel point les variations des probabilités subjectives peuvent être expliquées par les variations présentes dans les différentes séries suivantes représentant des sources d'information potentielles :

- le nombre de cas de COVID-19 (estimations selon les données PCR) (PCR_t)
- le nombre de cas de COVID-19 (estimations selon les données du CIRANO) (CIR_t)
- le nombre de décès liés à la COVID-19 (D_t)
- le nombre d'hospitalisation liés à la COVID-19 (H_t)
- le nombre de personnes en isolement dans le réseau du répondant ($I_{i,t}$)
- le nombre de personnes ayant la COVID-19 dans le réseau du répondant ($C_{i,t}$)

L'évolution de chacune de ces six séries sur les 18 semaines à l'étude est présentée graphiquement à la **Figure 5.14**. Afin de rendre la comparaison visuelle possible, elles ont été normalisées à 1 à la première semaine. Posons ici l'hypothèse que la courbe estimée par le CIRANO est celle reflétant de la manière la plus précise l'évolution de la pandémie. En comparaison, suivre l'évolution des décès semble peu pertinent car la courbe sous-estime fortement la tendance et est à retard sur les autres (avec la vague plafonnant à la semaine 15 plutôt que 13). Les hospitalisations relatent une histoire similaire, avec un sommet retardé d'une semaine et une sous-estimation des cas. Le graphique illustre également la pertinence décroissante de l'estimation par tests PCR, qui sous-estime fortement la vague de la semaine 13. Ainsi, le nombre de cas de COVID-19 et le nombre de personnes en isolement dans le réseau semblent être les deux autres meilleures sources d'information. Il devrait toutefois être souligné que l'usage du

nombre de cas confirmés de COVID-19 dans le réseau est une mesure utile lorsqu'agrégée, mais qui risque au niveau individuel d'être de trop faible amplitude pour être fiable sur une base hebdomadaire.

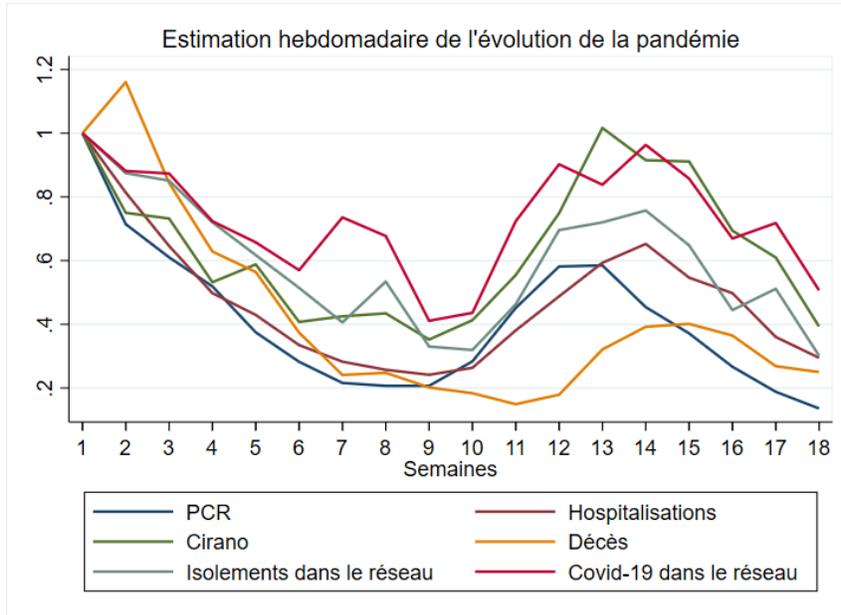


Figure 5.12 : Estimation hebdomadaire de l'évolution de la pandémie selon différentes sources d'information

D'un point de vue méthodologique, ce test est effectué en régressant un vecteur de variable explicative sur la probabilité subjective. Ce vecteur est composé des 6 variables correspondant aux différentes sources d'information, en plus des autres variables sociodémographiques de X_i précédemment utilisées (en plus des contrôles de réseau). Cette régression est effectuée à deux reprises, une fois sur la population des vaccinés (21) et l'autre fois sur la population des non-vaccinés (22).

$$P_{i,t} = \beta_s X_{i,t} + \beta_{s,PCR} PCR_t + \beta_{s,CIR} CIR_t + \beta_{s,D} D_t + \beta_{s,H} H_t + \beta_{s,I} I_{i,t} + \beta_{s,C} C_{i,t} + \varepsilon_{i,s} \forall i tq x_{i,vacc} = 1 \quad (21)$$

$$P_{i,t} = \beta_s X_{i,t} + \beta_{s,PCR} PCR_t + \beta_{s,CIR} CIR_t + \beta_{s,D} D_t + \beta_{s,H} H_t + \beta_{s,I} I_{i,t} + \beta_{s,C} C_{i,t} + \varepsilon_{i,s} \forall i tq x_{i,vacc} = 0 \quad (22)$$

Les valeurs des coefficients d'intérêt sont présentées à la table 5.4. Tous les coefficients sont présentés en points de pourcentage. La significativité a été déterminée par écart-types robustes. Les nombres de cas estimés par le CIRANO et par PCR sont exprimés en milliers, alors que les nombres d'hospitalisations

et de décès sont exprimés en centaines afin d'augmenter l'envergure des coefficients et les rendre ainsi plus faciles à interpréter. La première colonne présente les résultats pour la population des vaccinés. Il peut y être vu que le coefficient relatif au nombre de cas estimé par le CIRANO est significatif. Ce dernier implique que chaque tranche de 1000 cas de COVID-19 supplémentaire augmente la perception du risque par 0.047 point de pourcentage. Le nombre de personne en isolement dans le réseau du répondant est également significatif, chacune de ces personnes étant liée à une augmentation de la perception du risque de 0.185 point de pourcentage. Ceci suggère que les estimations du risque des vaccinés fluctuent en fonction de ces deux variables. Chez les non-vaccinés, seul le nombre de personnes en isolement dans le réseau est une source d'information significative. Pour cette sous-population, chaque personne du réseau en isolement augmente la probabilité subjective de 0.988 point de pourcentage.

L'hypothèse qu'il est possible de déduire les sources d'information consultées par les répondants strictement en observant les variations de la probabilité subjective est audacieuse. Toutefois, le fait qu'on obtienne une différence marquée entre les populations vaccinées et non-vaccinées vient suggérer une différence sous-jacente entre ces groupes. En effet, il semblerait que la perception du risque des vaccinés fluctue en fonction d'une source d'information agrégeant l'évolution de la pandémie de COVID-19 au niveau sociétal, alors que l'estimation des non-vaccinés ne fluctue qu'en fonction de l'état de la pandémie dans leur réseau.

Table 5.4 : Coefficients des différentes sources d'information lorsque régressées sur la probabilité subjective (points de pourcentage)

Population	Vaccinés		Non-vaccinés	
Variable	Coefficient	SE (robuste)	Coefficient	SE (robuste)
Hospitalisations (x100)	-0.243	0.271	1.531	1.051
Décès (x100)	0.918	0.571	-0.918	2.403
Réseau				
COVID-19	-0.0014	0.013	0.031	0.06
Isolement	0.185*	0.085	0.988**	0.351
PCR (x1000)	-0.136	0.115	-0.491	0.471
Cirano (x1000)	0.047***	0.008	0.003	0.033
Observations	50 412		2998	

*p<0.05, **p<0.01, ***p<0.001

Il est aussi intéressant de remarquer que les anticipations des répondants (vaccinés et non-vaccinés) varient en fonction du nombre de personnes en isolement dans leur réseau plutôt que le nombre de personnes ayant une infection confirmée de COVID-19. Si le répondant assume que chaque personne qu'il connaît en isolement est en effet infectée de la COVID-19, il n'est pas surprenant qu'on observe une surestimation généralisée du risque.

6. Discussion

La pandémie de COVID-19 est un événement particulier duquel on se doit de tirer des leçons afin de se prémunir contre un prochain épisode. Le présent travail contribue à cet objectif en présentant une étude des perceptions de la probabilité d'infection, et en juxtaposant des données sur les risques subjectifs et objectifs. La comparaison des deux démontre qu'il existe un biais important relativement généralisé dans l'estimation du risque prenant la forme d'une surestimation. Ce résultat permet de rejeter l'hypothèse d'anticipations rationnelles du risque d'infection. Ceci implique que cette hypothèse n'est sans doute pas une modélisation adéquate des anticipations pour les modèles épidémiologiques et économiques. Trouver une modélisation alternative reflétant de manière plus réelle la formation des attentes pourrait être un ajout pertinent à ces modèles.

Il ressort de l'analyse quelques biais particulièrement substantiels : celui propre aux personnes travaillant dans le secteur de l'enseignement et celui capté par la variable indicatrice du statut vaccinal. La discussion qui suit se concentre premièrement sur le développement et la conclusion de chacun de ces biais. Ensuite, un parallèle est proposé entre les résultats de ce mémoire et certains éléments de la littérature scientifique propre à la COVID-19. Par la suite, la pertinence économique des résultats de l'étude est abordée, finalement suivie d'une présentation des limites empiriques de ce travail.

6.1 Discussion sur le biais associé au secteur de l'enseignement

Les données permettent d'affirmer qu'en 2022, les enseignants n'avaient pas une plus haute probabilité de contracter la COVID-19 que les employés de notre secteur contrefactuel (le secteur 4, principalement composé d'emplois de service à la clientèle). Ceci est resté réalité même durant les semaines 9 à 18 de l'échantillon, durant lesquelles il n'y avait plus de masque en classe pour les élèves du primaire et secondaire alors qu'il y en avait toujours pour les employés de notre secteur de référence. Tel que mentionné plus tôt, il est possible que la perception du risque ait été biaisée parce que le milieu scolaire encourage un contact humain journalier avec un grand nombre de personnes. Cette réalité serait exacerbée par le fait que les jeunes élèves peuvent être asymptomatiques et les étudiants peu précautionneux. Il est également possible que les mesures gouvernementales et les messages médiatiques aient eu un rôle à jouer. Les écoles ont en effet souvent été définies comme étant des vecteurs de propagation du virus¹⁷. Le fait qu'elles aient été la cible de plusieurs mesures restrictives (dont

¹⁷ Serebin, Jacob (2021, 14 septembre). « Les cas de COVID-19 continuent d'augmenter dans les écoles au Québec », *Le Devoir*. Récupéré le 5 avril 2023 de <https://www.ledevoir.com/societe/sante/632100/les-cas-de-covid-19-continuent-d-augmenter-dans-les-ecoles-au-quebec>

la fermeture complète) aurait aussi pu contribuer à influencer la perception du risque des enseignants. Peu importe la cause, la réalité est que ces derniers ont été victimes d'un biais très élevé.

Un biais positif important comme celui qui a pesé sur les enseignants peut avoir un impact très réel sur les décisions de ce groupe. Rappelons par exemple que le retour à l'école en présentiel en janvier 2022 n'a pas fait l'unanimité. En effet, ce dernier a mené à des manifestations de la part de certains enseignants, en plus d'avoir été contesté par des syndicats représentant le corps professoral¹⁸. Ces réticences représentent concrètement en quoi peut se transformer un biais de perception positif. Alors qu'un biais de perception individuel nuit à l'utilité personnelle, un biais partagé par toute une sous-population peut mener à des conséquences beaucoup plus réelles et généralisées, surtout si elle est organisée derrière un front commun (comme un syndicat). Ce dernier permettant d'unifier l'effort, des mesures estimées comme justes par un groupe à la perception biaisée ont beaucoup plus de chances d'être adoptées, et ce malgré qu'elles puissent avoir des retombées sous-optimales pour le groupe en question (enseignants) et les autres acteurs (élèves, parents, gouvernement). Il peut donc être particulièrement important pour une société de cibler ces groupes unifiés et biaisés afin de tenter de faire arrimer leur perception du risque avec le risque réel, de manière à aboutir à une compréhension généralisée et adéquate du risque posée par la situation. Seulement alors pourront être mises en place des mesures pertinentes et optimales.

Une limite importante de ce mémoire est que nous n'évaluons que la situation des enseignants. Afin d'avoir une vision complète de la situation scolaire, il est requis d'également étudier la situation des élèves. Peu est connu de la situation des élèves du primaire et du secondaire puisqu'aucun mineur ne fait partie de la base de données. Tout de même, certains articles mettent en lumière des faits intéressants. Par exemple, des données du Ministère de l'éducation montrent que l'anxiété face à la situation scolaire en temps de pandémie n'est pas réservée aux enseignants, les étudiants en sont également victimes¹⁹. En effet, alors qu'en date du 28 janvier plus de 45% des absences à l'école étaient dues à la COVID-19, le 19 avril la pandémie causait directement moins de 15% des absences, et ce malgré une hausse du nombre total des absences. La cause alors citée est l'anxiété croissante chez les élèves, particulièrement marquée

¹⁸ Leduc, Louise et Marie-Ève Morasse (2022, 18 janvier). « Une rentrée dans la joie et l'inquiétude », *La Presse*. Récupéré le 10 décembre 2022 de <https://www.lapresse.ca/actualites/education/2022-01-18/retour-en-classe/une-rentree-dans-la-joie-et-l-inquietude.php>

¹⁹ Dion-viens, Daphnée (2022, 21 avril). « Hausse de 76% des absences d'élèves non liées à la COVID », *Le journal de Québec*. Récupéré le 13 janvier 2023 de https://www.journaldequebec.com/2022/04/21/hausse-de-76-des-absences-deleves-non-liees-a-la-covid?fbclid=IwAR2Uxs_MNgknRN8EEV3BYH2ZmFqEYO7wmaBy4iN6G9UWSKkXyrkbCacMh0

chez les filles. Il convient de rappeler ici les résultats de Leigh et al. (2022), qui trouvaient déjà au printemps 2020 que l'état de santé (mentale, physique, spirituelle et sociale) des jeunes et des femmes était particulièrement susceptible au stress engendré par la pandémie. Il est alors possible que deux années de pandémie plus tard, l'anxiété en lien avec la pandémie soit responsable d'une part croissante des absences chez les élèves. Ces données du Ministère de l'éducation peuvent suggérer une perception élevée du risque, même chez les écoliers. Éclaircir cette relation pourrait être une étape considérable dans la résolution du problème d'absentéisme particulièrement marqué en temps de pandémie.

Puisque n'est pas observée la situation des élèves dans ce travail, il est difficile d'évaluer la pertinence relative aux différentes décisions gouvernementales qui ont été prises concernant les écoles. Malgré les résultats qui suggèrent que les enseignants n'étaient pas plus à risque, il est fort possible que les écoles soient demeurées, en 2022, les vecteurs d'une propagation inter-étudiants du virus. Sans une étude plus approfondie de la situation des élèves, on ne peut se prononcer sur les mesures relatives au milieu scolaire. Tout ce que suggère ce mémoire est que, du moins du point de vue de la santé des enseignants, le retour en classe a été bien moins dangereux qu'anticipé.

6.2 Discussion sur le biais associé au statut vaccinal

Le biais le plus significatif identifié par ce mémoire est celui capté par la variable indicatrice du statut vaccinal. Plusieurs pistes sont à considérer dans l'évaluation de ce biais.

Premièrement, la différence marquée entre les estimations des vaccinés et non-vaccinés pourrait révéler que le risque d'infection de la COVID-19 est principalement présent à travers les contacts de l'individu avec son réseau. Les vaccinés auraient alors tenté d'incorporer des données agrégées sur la prévalence du virus à l'échelle sociétale alors que la prévalence du virus à l'échelle du réseau serait celle responsable du risque à considérer. En incorporant les données médiatiques, les vaccinés auraient fournis une perception du risque qui se serait rapprochée du risque d'une personne ayant la société comme réseau. Il ne serait alors pas surprenant d'isoler un biais positif considérable chez ces répondants. Cette explication potentielle peut correspondre au non-respect de la deuxième implication de l'hypothèse d'anticipations rationnelles de Muth (1961). Cette dernière affirme que la manière dont les anticipations rationnelles sont formées dépend spécifiquement de la structure théorique du système décrivant la situation économique (ou épidémiologique) en question. Il est alors possible que les vaccinés, en considérant l'état de la pandémie à l'échelle sociétale, aient pris en compte de l'information extérieure à la structure du problème les concernant, déviant du cadre structurant les anticipations comme rationnelles.

Il faut également aborder la présence potentielle d'un effet rebond. Les personnes vaccinées, de par la protection que leur procure leur statut vaccinal, pourraient s'être senties justifiées à réduire leurs comportements préventifs complémentaires, suivant le mécanisme identifié par le travail de Luckman et al. (2021). Cet effet rebond aurait pu avoir un effet positif marqué sur la probabilité objective d'infection, expliquant pourquoi nous obtenons une probabilité objective similaire entre vaccinés et non-vaccinés. Un autre point à considérer est que la décision d'une personne de ne pas se faire vacciner peut émaner d'une perception initiale plus faible du risque que présente la COVID-19. Il pourrait alors y avoir un biais de sélection relatif au statut vaccinal pouvant expliquer le biais identifié.

Concrètement, le fait que nous obtenons un biais positif si important pour le vaccin suggère que la prise de décision de ce groupe fut biaisée. Ce point est d'autant plus important que les vaccinés représentent près de 93% de l'échantillon. Un groupe de cette envergure a nécessairement une influence significative sur la prise de décisions et les mesures mises en place par le gouvernement. Il est alors probable que ce dernier est opté pour des mesures de manière à satisfaire une population qui surestimait grandement le risque que lui posait la pandémie en 2022, rappelant une des conséquences des biais de perception identifiées par Ropeik (2012). Il est également généralement accepté qu'une surestimation importante du risque peut mener à la présence de restrictions excessives (Viscusi, 1985). Afin de plus formellement aborder la relation délicate entre gouvernement, population et virus, une prochaine section (6.4) propose un cadre d'analyse plus structuré, à travers lequel sera d'un même coup abordé la pertinence d'une étude économique sur les biais de perception en temps de pandémie. Avant d'aborder cet important point de discussion, posons un parallèle entre les résultats de ce travail et ceux présentés dans la littérature.

6.3 Mise en perspective des résultats avec la littérature

Il est intéressant de comparer les résultats de ce mémoire avec ceux présentés dans la littérature scientifique relative à la COVID-19 premièrement afin d'estimer la validité de ce travail. De plus, la plupart des études abordant le sujet utilise des données datant de 2020 alors que ce travail a la chance de profiter de données de 2022. Un bénéfice additionnel à cette comparaison est alors de voir si les prévisions de 2020 ont tenu route durant la pandémie, ou au contraire si les comportements observés aujourd'hui divergent de la tendance précédemment observée.

La premier parallèle posé concerne la différence entre les sexes identifiée par Leigh et al. (2022). Nos résultats suggèrent que les femmes avaient en moyenne une perception du risque non-différente des hommes, et ce malgré un risque réel significativement inférieur, ce qui pourrait relever de l'adoption de plus de mesures préventives par ce sexe que par son homologue. Nos résultats de 2022 concordent alors

avec les prévisions obtenues en 2020, permettant de conclure que la pandémie a asymétriquement affecté les deux sexes à travers leur perception du risque.

Un autre point d'analyse intéressant de l'étude de Leigh et al. (2022) est qu'il y est démontré que les personnes moins éduquées considèrent moins de sources d'information et sont plus susceptibles de se fier aux commentaires de membres de la famille ou d'amis sur les réseaux sociaux dans leur recherche d'information sur la pandémie. Les résultats présentés à la section 5.3.3 de notre travail concluent que les personnes non-vaccinées ont une perception du risque qui ne varie qu'en fonction du nombre de personnes dans leur entourage étant en isolement, contrairement à la perception des vaccinés qui fluctuent également selon le nombre de cas de COVID-19 total agrégé par le CIRANO. Puisqu'il est su que les non-vaccinés sont plus fortement représentés dans les plus faibles niveaux d'éducation, il est possible de dire que les résultats de notre travail convergent vers ceux de Leigh et al. (2022), confirmant que les personnes moins éduquées, encore en 2022, étaient probablement plus susceptibles de former leurs anticipations en se basant sur leur famille et leurs amis.

Ce parallèle peut également être ramené au travail de Bhuiya et al. (2021), qui détermine qu'un plus faible niveau d'éducation est relié à une plus faible compréhension du phénomène et du risque entourant la COVID-19. Le fait que ces personnes considèrent moins de sources d'information est en congruence avec une plus faible compréhension du phénomène. Il est intéressant de remarquer que le groupe ayant été identifié par Bhuiya et al. comme étant susceptible d'avoir une mauvaise compréhension du phénomène de la COVID-19 est celui parmi lequel on observe aujourd'hui le plus faible taux de vaccination.

Un parallèle intéressant peut être posé avec certains des résultats présentés dans le survol de la littérature de Joslyn, Sinatra et Morrow (2021). À travers l'étude de pairs, ces auteurs illustrent la vulnérabilité de l'interprétation humaine aux statistiques. Le fait qu'on obtienne une perception du risque largement surestimée dans ce mémoire, principalement marquée chez le groupe qui est identifié comme ayant une perception du risque fluctuant avec les données agrégées du CIRANO pourrait être en partie la conséquence d'une mauvaise interprétation des données transigées par les médias. Les quelques ponts qui ont été tirés entre les résultats du mémoire et la littérature existante relative à la COVID-19 permettent de conclure sur une convergence des tendances obtenues en début de pandémie et celles relevées en 2022.

Finalement, malgré une tendance relativement similaire entre l'évolution du risque subjectif et objectif, les niveaux des deux courbes diffèrent largement, la différence prenant la forme d'une surestimation du

risque d'infection. En ce sens, nos résultats convergent vers le modèle de Prelec (1998), qui prévoit une surpondération des risques à faible occurrence, et vers ceux de Combs (1978), qui trouvent que les répondants sont généralement assez mal calibrés face au risque.

6.4 Discussion sur la pertinence économique

Afin de mieux saisir la pertinence économique de ces résultats, considérons un cadre théorique simple du comportement des principaux acteurs d'une société, et du problème que leur occasionne une pandémie.

6.4.1 Modélisation conceptuelle

Commençons par élaborer une version simplifiée du problème que pose une pandémie pour l'allocation des ressources. Considérons deux acteurs, une population de consommateurs homogènes et un gouvernement. Le problème du consommateur a été abordé plus tôt dans l'étude (section 2.1); il s'agit de prendre des décisions maximisant son utilité en présence de risque. Un consommateur pourrait ainsi vouloir minimiser la différence entre la perception du risque d'infection et le risque objectif ($\min_{p^*} (p^* - p)$), de manière à former des anticipations rationnelles. Ces anticipations peuvent ensuite être utilisées par l'agent dans sa prise de décision. Minimiser la différence lui permet de s'assurer autant de libertés que possible tout en maintenant un niveau de risque qu'il juge adéquat. Un biais de perception ($p^* \neq p$) vient alors nuire à l'agent, l'empêchant d'estimer avec précision le risque auquel il fait face. Ceci fera en sorte que sa prise de décision risque d'avoir un penchant de surprotection (si le biais est positif, $p^* > p$) ou de sous-protection (si le biais est négatif, $p^* < p$).

Il devrait toutefois être souligné que d'un point de vue sociétal, la maximisation de l'utilité personnelle de chaque agent peut elle-même mener à une solution agrégée sous-optimale. En effet, il est raisonnable de penser qu'avec une pandémie vient l'ajout d'une nouvelle externalité négative, la contagion (Krueger et al., 2020; Fernández-Villaverde et Jones, 2020). Le concept d'externalité réfère à la conséquence d'une action dont le coût n'est pas internalisé par l'agent. Ce dernier, en pesant les bénéfices et les coûts d'une telle action, se retrouve donc à sous-estimer ses coûts sociaux. Il s'ensuit que l'action sera surconsommée par l'agent (au-delà du niveau socialement optimal). Dans notre contexte, l'externalité réfère simplement au fait que le risque d'infection p_i varie non seulement en fonction des actions de préventions e de l'agent i , mais également en fonction des retombées des actions de l'agent j , dénotées r_j . Le risque prend alors la forme suivante : $p_i(e_i, r_j)$, où le coefficient de e_i est négatif alors que celui de r_j est positif (puisque l'externalité est négative). L'agent i n'a pas de contrôle sur la réalisation de r_j . Il est parallèlement

peu probable qu'un agent j maximisant son utilité personnelle cherche à minimiser r_j , puisque ceci lui entraînerait une perte d'utilité alors que les bénéfices seraient perçus par l'agent i .

Concrètement, l'externalité de contagion implique que dans sa prise de décision, l'agent considère le risque auquel il s'expose mais omet de prendre en compte le risque qu'il engendre pour autrui. Par exemple, il pourrait choisir de porter un masque alors qu'il s'expose à un risque mais omettre de le porter alors qu'il présente un risque pour les autres. Laisse à lui-même, cet agent aura donc tout intérêt à minimiser son biais personnel, mais maintiendra un niveau de risque trop élevé pour la société. L'existence de cette externalité fait alors en sorte que le comportement de l'agent se doit d'être corrigé pour qu'une solution socialement optimale soit atteinte.

Ici peut intervenir le second acteur de notre modèle simplifié, le gouvernement. La majorité des objectifs de la population découle de besoins physiques à combler (logement, nourriture, etc.). Un des rôles du gouvernement devient donc d'assurer le maintien d'une certaine liberté et prospérité économique permettant l'assouvissement de ces besoins par l'initiative personnelle. Ce que vient ajouter la pandémie à cette situation est une priorisation croissante par la population du besoin de sécurité. La responsabilité du gouvernement à cet égard implique notamment la correction des externalités. L'externalité de contagion complexifie alors beaucoup la tâche du gouvernement, puisqu'il se doit de satisfaire une population qui, laissée à elle-même, aboutirait à une allocation sous-optimale (n'internalisant pas le coût social de la contagion, la population échoue à discerner l'allocation qui est la meilleure pour elle). L'action du gouvernement devient d'autant plus ardue que la mise en place de mesures protectrices se fait souvent au détriment de son objectif parallèle d'encourager l'activité économique (Acemoglu et al., 2021).

Le gouvernement a toutefois plusieurs stratégies pouvant être employées afin de limiter la contagion. Certaines options (plutôt structurelles) pourraient être, par exemple, d'assurer un approvisionnement en vaccins, en masques et en lits d'hôpitaux. Afin d'assurer un impact significatif, le gouvernement judicieux combinera ces stratégies d'approvisionnement avec d'autres stratégies, de nature rectificative, ayant comme objectif de faire internaliser le coût de la contagion à la population. Ces dernières pourraient par exemple prendre la forme d'éducation sur la vaccination ou de mandat obligeant le port du masque. Le travail de Fernández-Villaverde et Jones (2020) confirme l'importance des interventions non-pharmaceutiques qu'entreprend le gouvernement dans le contrôle de la pandémie (les exemples des masques, des limites de contacts humains, des quarantaines et de l'amélioration de la ventilation sont utilisés).

Deux types de solutions s'offrent au gouvernement afin de corriger l'externalité de contagion. Il peut mettre en place des mesures soit restrictives ou éducationnelles, l'objectif conjoint étant d'affecter le comportement (e) de sa population. Prenant l'exemple du port du masque, le gouvernement a l'option d'imposer son port (mesure restrictive) ou de convaincre la population de l'importance de son port (mesure éducationnelle). La mise en place de mesures restrictives correspond à une situation où le gouvernement détermine un seuil minimal x de prévention ($e_i \geq x$), qui devrait limiter l'importance des retombées r_i . Plusieurs exemples de mesures restrictives ont été employés durant la pandémie au Québec, les principaux étant sans doute le passeport vaccinal, le confinement, le port du masque obligatoire et la fermeture de la plupart des services intérieurs. La nature de ces mesures fait en sorte qu'elles risquent d'être sous-optimales car à la fois très coûteuses (pertes économiques) et très contraignantes pour l'agent (pertes d'utilité). Celles-ci demeurent tout de même une option viable pour le gouvernement car la menace de sanction qui les accompagne assure une certaine coopération de la part de l'agent.

Autrement, le gouvernement peut choisir d'intervenir directement sur la prise de décision de l'agent en venant influencer sa perception du risque (p^*). Il est maintenant généralement accepté que les changements volontaires de comportements (plutôt qu'imposés) ont joué un rôle très important dans le contrôle de la pandémie (Fernández-Villaverde et Jones, 2020). Ces auteurs avancent également que le travail d'éducation et de partage de l'information entrepris par les gouvernements a assurément influencé les croyances et donc les comportements de la population en lien avec la pandémie, un consensus également présent dans la revue de littérature de Joslyn, Morrow et Sinatra (2021). Afin que ce type d'intervention soit optimal, le gouvernement devrait venir créer un biais positif dans sa population équivalent au coût social engendré par l'externalité de contagion ($p_i^* = p_i + r_i$). Ceci devrait permettre l'atteinte d'une allocation socialement optimale, et ce à bien moindre coût que par mesures restrictives. Cette méthode a l'avantage de venir changer intrinsèquement le comportement de l'agent, ce qui devrait assurer qu'il agisse conséquemment même lorsqu'il n'est pas observé (restreint) par le principal.

6.4.2 Implications

Cette courte structuration permet de mieux comprendre le contexte dans lequel vient s'imbriquer les biais de perception qui ont été analysés plus tôt. Le présent mémoire permet d'ailleurs d'ajouter quelques éléments empiriques à ce cadre conceptuel. En effet, l'existence d'un biais positif ($p^* > p$) a clairement été identifiée par nos résultats. Tel que discuté plus haut, bien que l'individu tente de s'en débarrasser, ce biais peut être quelque chose de socialement désirable. Ce résultat pourrait suggérer que les mesures

éducationnelles à la disposition du gouvernement aient été une solution envisageable. Ce mémoire a l'avantage de directement quantifier ce biais positif. Toutefois, afin de juger de son amplitude, il convient de le comparer à l'externalité de contagion, pour laquelle aucune estimation n'est offerte ici. Il pourrait être intéressant qu'une étude subséquente soit dédiée à l'estimation de cette externalité.

Une autre considération importante est que la collecte de données s'est effectuée en 2022. Après deux ans à côtoyer la COVID-19, il est fort probable que la population s'y soit habituée et ait ajusté sa perception du risque en conséquence. Il aurait été intéressant d'avoir une collecte de données en 2020, lorsque la maladie n'était qu'à ses débuts. Il ne serait pas surprenant d'y avoir trouvé une estimation du risque beaucoup plus volatile que celle recensée ici. Ce point devrait être considéré avec attention si les résultats de ce travail se trouvent à être comparés à ceux d'une prochaine pandémie.

Similairement, nous évaluons le biais de perception deux ans après le premier cas de COVID-19 recensé au Québec, et suivant la mise en place et le retrait de plusieurs mesures restrictives. Essayer de déterminer quelles sont les causes ayant entraîné les biais de perception qu'on observe aujourd'hui ne peut être qu'un exercice de spéculation. Il serait injuste d'assumer que ces biais proviennent strictement d'efforts éducatifs gouvernementaux. Ceci implique que nous n'avons pas une estimation claire de l'impact que peut avoir une mesure non restrictive sur la calibration des anticipations.

Ce mémoire identifie également un problème supplémentaire laissé sous silence dans le cadre conceptuel. Ce dernier présupposait une population homogène, mais la décortication du biais présentée plus tôt nous apprend qu'au contraire, les biais semblent être assez inégalement répartis dans la population. Ceci pourrait avoir comme implication que les mesures de type éducatif ne fonctionnent pas sur certaines sous-populations, par exemple celle des non-vaccinés. Ceux-ci étant probablement moins connectés à la couverture médiatique susceptible d'influencer leur perception du risque (Leigh et al., 2022), la meilleure manière de les rejoindre et d'entraîner leur comportement pourrait être l'usage de mesures restrictives. Ainsi, malgré le coût supérieur engendré, il pourrait demeurer optimal pour le gouvernement de mettre en place un mélange de mesures restrictives et éducatives.

Il existe également une difficulté importante propre aux mesures visant à compenser l'externalité de contagion, soit que cette dernière est différente pour chaque personne. En effet, une multitude de facteurs tels que l'usage de mesures préventives, la présence d'enfants, la taille du cercle social ou encore la nature de l'emploi vient influencer la valeur de cette externalité. La situation ponctuelle de la pandémie vient également affecter sa valeur, puisqu'en approche d'un sommet de vague le risque de contagion

augmente considérablement. La nature changeante et personnelle de la valeur de l'externalité de contagion fait alors en sorte qu'elle devient difficile à assujettir adéquatement aux différents segments de la population. Le gouvernement a toutefois la chance d'agir de manière plutôt efficace sur les segments organisés (par exemples les enseignants) car ceux-ci se trouvent à être relativement unifiés. L'étude d'Acemoglu et al. (2021) arrive d'ailleurs à la conclusion que la mise en place de mesures restrictives ciblées (selon le groupe d'âge) est une manière efficace de réduire la mortalité tout en préservant un certain niveau d'activité économique (lorsque comparé à la mise en place de mesures uniformes). Une conclusion similaire peut être tirée de nos résultats, quoique notre analyse suggère qu'il serait intéressant que les efforts gouvernementaux soient ciblés plus précisément que simplement par groupe d'âge.

Il serait ainsi judicieux, lors d'un prochain épisode pandémique, de recenser assez rapidement les biais de perception de la population. Ceci permettrait au gouvernement d'identifier d'emblée les groupes qui risquent de militer pour l'adoption de mesures sous-optimales. Il pourrait ensuite venir réduire ces biais par l'entremise de mesures éducationnelles proactives ciblées, suivant l'intuition derrière les résultats de Keren (1987) qui suggère que la familiarisation avec un risque peut mener à une meilleure calibration des anticipations. Reprenant l'exemple des enseignants, avant de mettre en place une mesure restrictive telle que le retour en classe, qui risque de susciter des objections de la part des parties prenantes, le gouvernement pourrait tenter de corriger le biais de perception des enseignants en les éduquant sur le risque réel qu'ils encourent. Déconstruire un biais de perception par l'éducation sur le risque pourrait être une clé permettant de désamorcer beaucoup de la résistance aux mesures lors d'une prochaine pandémie.

Il convient finalement d'aborder une implication importante relative à la mise en place de mesures éducationnelles. Il a été dit plus tôt que celles-ci ont un coût bien moindre aux mesures restrictives, mais il ne faut pas assumer pour autant que leur coût soit nul. Il est important de corriger la perception du risque pour l'externalité de contagion, mais il est également important de ne pas sur-corriger, car ceci peut également mener à des conséquences indésirables. La surcorrection d'un biais négatif correspondrait à une situation où le gouvernement, désirant augmenter la perception du risque d'un certain segment de population, lui ferait percevoir un risque trop important, excédant le risque réel additionné de l'externalité de contagion ($p_i^* > p_i + r_i$). Le danger ici présent est que le stress occasionné par une perception excessive du risque peut venir nuire directement à l'état de santé de la population, suivant un mécanisme rappelant celui identifié par Ropeik (2012) et Han et al. (2021). Une illustration appropriée de cette situation est trouvée dans l'étude de Cohen et al. (1993), qui confirme l'existence

d'une dynamique liant le niveau de stress et la vulnérabilité à l'infection (au virus du rhume dans l'étude). Ce résultat entend que bien qu'un biais négatif présente un danger de contagion important pour la société, un biais positif exagéré est également nuisible, et ce parce qu'il risque d'augmenter la vulnérabilité de la population à l'infection. Il peut aussi plus généralement être affirmé qu'un biais positif exagéré vient nuire à l'optimalité de la prise de décision de l'agent, résultant nécessairement en une perte d'utilité. Ces arguments font en sorte qu'il convient de faire arrimer précisément l'estimation à la réalisation du risque.

6.5 Limites

Quelques limites empiriques sont importantes à considérer avant d'accepter les résultats obtenus par ce mémoire. Premièrement, la majeure partie de ce travail est centrée autour de l'analyse d'une variable de probabilité subjective. Il suffit de considérer les problèmes de l'incertitude épistémique ou de l'arrondissement mentionnés plus tôt pour comprendre la vulnérabilité de cette variable. Il pourrait être intéressant pour une prochaine étude de permettre aux répondants d'éliciter des étendues de probabilité plutôt que des valeurs absolues de manière à atténuer ces problèmes.

Une autre limite potentielle découle du fait que comptabiliser les cas confirmés de COVID-19 est une tâche difficile à accomplir. Nous avons choisi dans ce mémoire d'intégrer les autodiagnostic dans notre échantillon à cause du manque de disponibilité et de précision des tests rapides, mais il est évident que nous ouvrons ici la porte à une surestimation de la probabilité objective. Ce manque de précision, si significatif, risque toutefois de faire en sorte que les biais positifs soient sous-estimés par ce travail, apportant une certaine robustesse aux résultats. Une analyse visant à déterminer la sensibilité de nos résultats à l'inclusion des autodiagnostic serait un ajout intéressant à ce mémoire. Inversement, la nature asymptomatique de certaines des infections pourraient faire en sorte que notre mesure du risque objectif soit une sous-évaluation du risque réel. Ceci aurait influencé à la hausse nos résultats relatifs aux biais positifs (notamment enseignement et vaccin).

Abordons finalement la validité externe du mémoire. Il est important de considérer que chaque pays a affronté différemment la pandémie. Même au sein du Canada, les gouvernements provinciaux ont été très impliqués dans la gestion du virus, ce qui fait en sorte que la pandémie a évolué différemment dans chaque province. Il est fort probable que les mesures mises en place au Québec aient eu un impact non négligeable sur les résultats obtenus, à la fois au niveau du risque subjectif et du risque objectif. La surestimation présente dans nos résultats est sans doute indicatrice d'un phénomène généralisé, mais

une considération particulière devrait être portée aux différences entre états avant d'extrapoler les résultats.

6.6 Pistes futures

Des chercheurs de plusieurs disciplines ont tenté d'approfondir notre compréhension du phénomène de la COVID-19, se ralliant derrière l'objectif d'être mieux préparé en vue d'une éventuelle réitération. Le présent travail contribue à cet effort commun en s'intéressant à un sujet qui est resté, à notre connaissance, délaissé par la littérature économique, soit la calibration des anticipations. Les résultats du mémoire permettent de remettre en question la validité de l'hypothèse de rationalité dans la formation des attentes appliquée au contexte pandémique. Il serait pertinent qu'un prochain travail s'attarde à identifier une modélisation alternative reflétant adéquatement la forme que suivent les anticipations du risque d'infection. Cette dernière pourrait ensuite être substituée à l'hypothèse d'anticipations rationnelles présentes dans certains modèles épidémiologiques. L'exercice devrait résulter en des modèles prédisant plus réalistement l'évolution du virus.

Il serait intéressant qu'un travail subséquent soit dédié aux biais identifiés mais peu approfondis par ce mémoire, notamment celui de l'âge qui semble particulièrement saillant. Autrement, il serait intéressant d'étudier la nature du mécanisme ayant mené à la réalisation de ces biais. Comprendre leur origine pourrait être une étape cruciale permettant de faire arrimer la perception du risque avec le risque objectif. Il serait également intéressant d'étudier en quelle sorte de comportements ces biais se sont manifestés. Répondre à cette question permettrait de mieux cibler et prédire l'adhérence de la population aux mesures lors d'un prochain épisode.

Un mot se doit finalement d'être glissé sur l'hétérogénéité des biais entretenus par la population. À la lumière de nos résultats, il semble pertinent pour le gouvernement d'agir différemment sur différents segments de la population. Recenser les biais de chaque groupe semble un prérequis indispensable à la mise en place de mesures. Ceci permettrait la réalisation d'un effort éducatif ciblé répondant au biais de chacun, ce que ce mémoire reconnaît comme une forme intéressante d'intervention.

Bibliographie

- Aadland, David, David Finnoff et Kevin X. D. Huang (2011). « The dynamics of economic epidemiology equilibria », *University of Wyoming*. Récupéré le 5 avril 2023 de <https://mpra.ub.uni-muenchen.de/29299/>
- Abu-Raddad, Laith J., Hiam Chemaitelly, Houssein H. Ayoub et Sawsan AlMukdad (2022). « Effect of mRNA Vaccine Boosters against SARS-CoV-2 Omicron Infection in Qatar », *The New England Journal of Medicine*, vol. 386, p. 1804-1816.
- Acemoglu, Daron, Victor Chernozhukov, Ivan Werning et Michaels D. Whinston (2021). « Optimal targeted lockdowns in a multigroup SIR model », *American Economic Review: Insights*, vol. 3, no 4, p. 487-502.
- Benitez-Silva, Hugo, Moshe Buchinsky, Hiu Man Chan, Sofia Cheidvasser et John Rust (2004). « How large is the bias in self-reported disability? », *Journal of Applied Econometrics*, vol. 19, no 6, p. 649-670.
- Bhuiya, Tanzim, Richard Klares III, Madellena A Conte et Joseph Steven Cervia (2021). « Predictors of misperceptions, risk perceptions, and personal risk perceptions about COVID-19 by country, education and income », *Journal of Investigative Medicine*, vol. 69, no 8, p. 1473-1478.
- Boisclair, David, Roxanne Borgès Da Silva, Vincent Boucher, Nathalie De Marcellis-Warin, Pierre-Carl Michaud et Ingrid Peignier (2022). *Combien de personnes ont développés des symptômes ou contracté la COVID-19 au Québec*, Cahier scientifique no 2022s-03, Montréal, CIRANO.
- Borrelli, Belinda, Rashelle B. Hayes, Shira Dunsiger et Joseph L. Fava (2010). « Risk perception and smoking behavior in medically ill smokers: a prospective study », *Addiction*, vol. 105, no 6, p. 1100-1108.
- Brenner, Lyle A. (2003). « A random support model of the calibration of subjective probabilities », *Organizational Behavior and Human Decision Processes*, vol. 90, no 1, p. 87-110.
- Carroll, Christopher D. (2001). « The epidemiology of macroeconomic expectations », *National Bureau of Economic Research*, no w8695.
- Cohen, Sheldon, David A. Tyrrell et Andrew P. Smith (1993). « Negative life events, perceived stress, negative affect, and susceptibility to the common cold », *Journal of Personality and Social Psychology*, vol. 64, no 1, p. 131-140.
- Combs, Barbara, Baruch Fischhoff, Mark Layman, Sarah Lichtenstein et Paul Slovic (1978). « Judged frequency of lethal events », *Journal of Experimental Psychology: Human Learning and Memory*, vol. 4 no 6, p. 551-578.
- De Bruin, Wändi Bruine, Baruch Fischhoff, Susan G. Millstein et Bonnie L. Halpern-Felsher (2000). « Verbal and numerical expressions of probability: "It's a fifty-fifty chance" », *Organizational behavior and human decision processes*, vol. 81, no 1, p. 115-131.
- Eckhoudt, Louis, Christian Gollier et Harris Schlesinger (2004). « Optimal prevention », dans *Economic and financial decisions under uncertainty*, Princeton University Press, p. 153-159.
- Fernández-Villaverde, Jesus et Charles I. Jones (2020). « Macroeconomic outcomes and COVID-19: a progress report », *National Bureau of Economic Research*, no w28004.

- Ferrer, Rebecca A. et William M. Klein (2015). « Risk perceptions and health behavior », *Current opinion in psychology*, vol. 5, p. 85-89.
- Garfin, Dana Rose, Baruch Fischhoff, E. Alison Holman et Roxanne Cohen Silver (2021). « Risk perceptions and health behaviors as COVID-19 emerged in the United States: results from a probability-based nationally representative survey », *Journal of Experimental Psychology: Applied*, vol. 27, no 4, p. 584-598.
- Han, Qing, Bang Zheng, Maximilian Agostini, Jocelyn J. Bélanger, Ben Gützkow, Jannis Kreienkamp *et al.* (2021). « Associations of risk perception of COVID-19 with emotion and mental health during the pandemic », *Journal of Affective Disorders*, vol. 284, p. 247-255.
- Hurd, Michael D. (2009). « Subjective probabilities in household surveys », *Annual Review of Economics*, vol. 1, no 1, p. 543-562.
- Joslyn, Susan, Gale M. Sinatra et Daniel Morrowz (2021). « Risk perception, decision making, and risk communication in the time of COVID-19 », *Journal of Experimental Psychology: Applied*, vol. 27, no 4, p. 579-583.
- Karni, Edi (2005). « Savages' subjective expected utility model », *Johns Hopkins University*. Récupéré le 20 mars 2023 de <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=1e423a8313f33b21b3d3e9b857fdc ae9dfcfd98>
- Keren, Gideon (1987). « Facing uncertainty in the game of bridge: a calibration study », *Organizational Behavior and Human Decision Processes*, vol. 39, no 1, p. 98-114.
- Krueger, Dirk, Harald Uhlig et Taojun Xie (2020). « Macroeconomic dynamics and reallocation in an epidemic: evaluating the "swedish solution" », *National Bureau of Economic Research*, no w27047.
- Leigh, Jeanna Parsons, Rebecca Brundi-Mather, Andrea Soo, Emily Fitzgerald, Sara Mizen, Alexandra Dodds *et al.* (2022) « Public perceptions during the first wave of the COVID-19 pandemic in Canada: a demographic analysis of self-reported beliefs, behaviors, and information acquisition », *BMC Public Health*, vol. 22, no 699.
- Lichtenstein, Sarah, Baruch Fischhoff et Lawrence D. Phillips (1976). « Calibration of probabilities: The state of the art », *Theory and Decision Library*, vol. 16, p. 275-324.
- Lovell, Michael C. (1986). « Tests of the rational expectations hypothesis », *The American Economic Review*, vol. 76, no 1, p. 110-124.
- Manski, Charles F. (2004). « Measuring expectations », *Econometrica*, vol. 75, no 5, p. 1329-1376.
- Maski, Charles F. et Francesca Molinari (2010). « Rounding probabilistic expectations in surveys », *Journal of Business & Economic Statistics*, vol. 28, no 2, p. 219-231.
- Muth, John F. (1961). « Rational expectations and the theory of price movements », *Econometrica*, vol. 29, no 3, p. 315-335.
- Prelec, Drazen (1998). « The probability weighting function », *Econometrica*, vol. 66, no 3, p. 497-425.
- Ropeik, David (2012). « The perception gap: recognizing and managing the risks that arise when we get risk wrong », *Food and chemical toxicology*, vol. 50, no 5, p. 1222-1225.
- Savage, Leonard J. (1954). *The foundations of statistics*, New York, Willey.

Viscusi, William Kip (1985). « Are individuals bayesian decision makers? », *The American Economic Review*, vol. 75, no 2, p.381-385.

Von Neumann, John et Oskar Morgenstern (1947). *Theory of games and economic behavior*, 2^e éd., Princeton, Princeton University Press.