

Université de Montréal

**Analyse de sensibilité de l'effet d'un programme de
prévention avec randomisation**

**Application de trois techniques d'appariement pour balancer les
groupes contrôle et expérimental: distance de Mahalanobis,
score de propension et algorithme génétique**

par

François Maurice

Département de sociologie

Faculté des arts et des sciences

Mémoire présenté à la Faculté des arts et des sciences

en vue de l'obtention du grade de

Maître ès sciences (M. Sc.)

en Sociologie

Mars, 2014

© François Maurice, 2014

Université de Montréal
Faculté des arts et des sciences

Ce mémoire intitulé :

**Analyse de sensibilité de l'effet d'un programme de
prévention avec randomisation**

Application de trois techniques d'appariement pour balancer les groupes contrôle et expérimental: distance de Mahalanobis, score de propension et algorithme génétique

Présenté par :
François Maurice

a été évalué par un jury composé des personnes suivantes :

Claire Durand, président-rapporteur
Éric Lacourse, directeur de recherche
Marc Henry, membre du jury

Résumé

L'estimation sans biais de l'effet causal d'une intervention nécessite la comparaison de deux groupes homogènes. Il est rare qu'une étude observationnelle dispose de groupes comparables et même une étude expérimentale peut se retrouver avec des groupes non comparables. Les chercheurs ont alors recours à des techniques de correction afin de rendre les deux groupes aussi semblables que possible. Le problème consiste alors à choisir la méthode de correction appropriée. En ce qui nous concerne, nous limiterons nos recherches à une famille de méthodes dites d'appariement. Il est reconnu que ce qui importe lors d'un appariement est l'équilibre des deux groupes sur les caractéristiques retenues. Autrement dit, il faut que les variables soient distribuées de façon similaire dans les deux groupes. Avant même de considérer la distribution des variables entre les deux groupes, il est nécessaire de savoir si les données en question permettent une inférence causale. Afin de présenter le problème de façon rigoureuse, le modèle causal contrefactuel sera exposé. Par la suite, les propriétés formelles de trois méthodes d'appariement seront présentées. Ces méthodes sont l'appariement par la distance de Mahalanobis, de l'appariement par le score de propension et de l'appariement génétique. Le choix de la technique d'appariement appropriée reposera sur quatre critères empiriques dont le plus important est la différence des moyennes standardisées. Les résultats obtenus à l'aide des données de l'Enquête longitudinale et expérimentale de Montréal (ÉLEM) indiquent que des trois techniques d'appariement, l'appariement génétique est celui qui équilibre mieux les variables entre les groupes sur tous les critères retenus. L'estimation de l'effet de l'intervention varie sensiblement d'une technique à l'autre, bien que dans tous les cas cet effet est non significatif. Ainsi, le choix d'une technique d'appariement influence l'estimation de l'effet d'une intervention. Il est donc impérieux de choisir la technique qui permet d'obtenir un équilibre optimal des variables selon les données à la disposition du chercheur.

Mots-clés : analyse causale, effet causal, modèle causal contrefactuel, graphe causal, graphe orienté acyclique, appariement, distance de Mahalanobis, score de propension, algorithme génétique, multiniveaux, effet d'une intervention, données longitudinales

Abstract

The unbiased estimate of the causal effect of an intervention requires the comparison of two homogeneous groups. It is rare that an observational study has comparable groups and even an experiment may end up with non-comparable groups. The researchers then used correction techniques to make the two groups as similar as possible. The problem then is to choose the appropriate correction method. In our case, we will restrict our research to a family of so-called matching methods. It is recognized that what matters in a match is the balance between the two groups on selected characteristics. In other words, it is necessary that the variables are distributed similarly in both groups. Even before considering the distribution of variables between the two groups, it is necessary to know whether the data in question allow for causal inference. To present the problem rigorously, the counterfactual causal model will be exposed. Thereafter, the formal properties of three matching methods will be presented. Those methods are the Mahalanobis matching, the propensity score matching and genetic matching. The choice of the appropriate matching technique is based on four empirical criteria which the most important is the standardized mean difference. Results obtained using data from the Montréal Longitudinal and Experimental Study indicate that of the three matching techniques, genetic matching is the one that better balance the variables between groups on all criteria. The estimate of the effect of intervention varies substantially from one technique to another, although in all cases this effect is non significant. Thus, the selection of a matching technique influences the estimation of the effect of an intervention. Therefore, it is imperative to choose the technique that provides an optimal balance of the variables based on data available to the researcher.

Keywords : causal analysis, causal effect, counterfactual causal model, causal graph, directed acyclic graph, matching, Mahalanobis distance, propensity score, genetic algorithm, multilevel, effect of an intervention, longitudinal data

Table des matières

1	Introduction.....	1
2	L'analyse causale en sociologie.....	4
3	Le modèle causal contrefactuel.....	8
3.1	Définitions et postulats d'identification de l'effet causal.....	10
4	L'appariement pour équilibrer les variables.....	24
4.1	Caractéristiques générales des méthodes d'appariement multivariées.....	27
4.2	L'appariement par la distance de Mahalanobis.....	29
4.3	L'appariement par le score de propension.....	31
4.4	L'appariement génétique.....	36
4.5	Une évaluation des méthodes d'appariement par simulation.....	37
5	Application de l'analyse causale.....	40
5.1	L'Étude longitudinale et expérimentale de Montréal (ÉLEM).....	40
5.1.1	Présentation de la composante expérimentale.....	41
5.1.1.1	Sélection des cas.....	41
5.1.1.2	Le programme d'intervention auprès des enfants et des parents.....	44
5.2	Effet causal d'une intervention psychosociale sur le comportement agressif.....	45
5.2.1	Le modèle causal contrefactuel.....	45
5.2.2	L'appariement.....	46
5.2.2.1	La sélection des variables d'appariement.....	46
5.2.2.2	Les valeurs manquantes.....	48
5.2.2.3	Spécification des appariements.....	48
5.2.2.4	Diagnostic de l'appariement.....	48
5.2.3	L'effet causal.....	60
6	Conclusion.....	68
	Bibliographie.....	71
	Appendices.....	76
	Appendice 1: Les analyses multiniveaux préliminaires.....	76

Appendice 2: Graphes causaux et sélection des variables	91
Appendice 3: L'analyse de sensibilité.....	98
Appendice 4: Recherches récentes sur l'appariement	105
Appendice 5: Syntaxes de programmation	107

Liste des tableaux

Tableau 1 : Le problème fondamental de l'inférence causale.....	13
Tableau 2 : Détails du problème fondamental de l'inférence causale	15
Tableau 3 : Identification des cas-contrôles aux cas-interventions.....	19
Tableau 4 : Une stratification comme résolution du problème de l'inférence causale	22
Tableau 5: Élaboration de tableaux.....	24
Tableau 6 : Élaboration de tableau dont la variable-test est le sexe.....	32
Tableau 7 : Comparaison des biais obtenus par différentes méthodes d'appariement	38
Tableau 8 : Statistiques descriptives par groupe: moyenne, écart-type, N	42
Tableau 9 : Six cas de figure lors de l'équilibrage de variables.....	51
Tableau 10 : Analyse de l'équilibre des variables	56
Tableau 11 : Effets de l'intervention	63
Tableau 12 : Typologie des modèles de la modélisation du temps.....	83
Tableau 13: Typologie des modèles - Temps et groupe	85
Tableau 14 : Comparaison de trois structures de variance-covariance	86
Tableau 15 : Test de sensibilité de Rosenbaum	104
Tableau 16 : Comparaison des méthodes CEM et génétique.....	106

Liste des figures

Figure 1 : Graphe causal avec une variable de confusion.....	47
Figure 2 : Moyennes brutes du log de Violence	79
Figure 3 : Graphique de croissance empirique avec trajectoires MCO	80
Figure 4 : Trajectoires MCO pour tous les cas et pour chaque groupe.....	81
Figure 5 : Examen de la normalité des résidus	89
Figure 6 : Examen de l'homoscédasticité.....	90
Figure 7 : Graphe causal général.....	92
Figure 8 : Graphe causal avec causes simultanées.....	95
Figure 9 : Graphe causale avec un mécanisme	95

À tous mes proches

Remerciements

Je remercie Marie-Isabelle Martineau pour son soutien indéfectible malgré la maladie qui l'assaille. L'écriture de ce mémoire fût tout autant une épreuve pour elle comme pour moi.

Je remercie également mon directeur de recherche Éric Lacourse, sans qui je n'aurais jamais envisagé d'étudier l'analyse statistique de la causalité. Sans le savoir, Éric me proposait d'étudier scientifiquement un concept philosophique qui m'apparaît des plus importants.

1 Introduction

De nombreuses recherches des sciences sociales nécessitent la comparaison de deux ou plusieurs groupes et plusieurs de ces comparaisons prennent la forme d'une analyse causale, c'est-à-dire l'estimation de l'effet d'un événement sur un groupe d'individus, par opposition à une analyse corrélative ou encore à une analyse prédictive. Cependant, les données dont disposent les chercheurs des sciences sociales sont rarement appropriées pour l'analyse causale puisque ces chercheurs disposent habituellement de données observationnelles au lieu de disposer de données expérimentales. Ce dernier type de données, obtenu à l'aide d'un plan expérimental, permet d'évaluer l'effet sans biais d'une cause sur un groupe d'individus, tandis que l'estimation de l'effet d'une cause à l'aide de données observationnelles sera presque toujours biaisée.

Un plan expérimental prévoit l'assignation aléatoire des cas à un groupe-intervention et à un groupe-contrôle. L'assignation aléatoire des cas garantit que les deux groupes sont similaires sur l'ensemble de leurs caractéristiques, que celles-ci soient observées ou non. Aucune des caractéristiques ne peut donc servir à prédire l'appartenance d'un cas au groupe-intervention ou au groupe-contrôle. Il n'y a pas de corrélation entre la variable qui indique l'appartenance à un des deux groupes et les variables qui caractérisent les cas avant l'intervention.

Dans le cas de données observationnelles, un lien existe souvent entre la variable indicatrice et les caractéristiques des cas. Comment estimer l'effet d'une cause lorsque nous disposons de telles données? L'appariement est l'une des solutions possibles au problème de biais de données observationnelles. L'appariement est conçu pour réduire les écarts entre les deux groupes à l'étude et ainsi reproduire partiellement les conditions de l'assignation aléatoire. L'appariement permet de créer un groupe-contrôle similaire au groupe-intervention selon les caractéristiques observées et mesurées au sein d'un échantillon. Cependant, il existe de nombreuses techniques d'appariement. Comment choisir la bonne technique selon les données à notre disposition? De toute évidence il faut se doter d'un critère qui permet de départager la technique adaptée à nos données de celles qui ne le sont pas. Nous présenterons un critère qui nous permettra de choisir l'une des trois techniques

d'appariement multivarié que nous exposons et qui se sont succédé depuis l'apparition de la première au début des années 70 : distance de Mahalanobis, score de propension et algorithme génétique.

Remarquons que les techniques de régression peuvent être utilisées pour régler les problèmes de biais des données observationnelles. Cependant, il est bien connu qu'un modèle de régression est difficile à établir, notamment à cause des problèmes de la forme des relations entretenues entre la variable dépendante et les variables indépendantes, mais aussi à cause du problème de colinéarité qui peut survenir entre les variables indépendantes. Dans ces circonstances, il est recommandé par certains chercheurs d'effectuer une analyse causale en deux étapes. On utilise une technique d'appariement pour transformer les données observationnelles en données quasi expérimentales, puis on utilise une technique de régression adaptée à la nature des données enfin d'évaluer l'effet causal (Ho, Imai, King et Stuart, 2007; Imai, King, Stuart et Ho, 2011). C'est l'approche que nous avons choisie pour notre travail. Nous reviendrons sur cette question afin de la traiter un peu plus en détail.

Cependant, il ne suffit pas d'utiliser l'appariement pour que les données observationnelles deviennent des données quasi expérimentales, il faut aussi que les données aient été cueillies d'une certaine façon pour s'assurer que la différence entre les deux groupes sur la variable d'intérêt soit interprétée de façon causale. Ainsi, nous présenterons un modèle d'analyse qui permet d'évaluer si oui ou non les données peuvent être utilisées pour une estimation causale. Ce modèle d'analyse se nomme modèle causal contrefactuel ou encore modèle causal de Neyman-Rubin. Nous verrons que nos données s'accordent bien à ce modèle causal puisque celles-ci sont produites par deux groupes dont les membres ont été assignés de façon aléatoire à l'un ou à l'autre.

En principe, il n'est pas nécessaire d'utiliser les techniques d'appariement présentées dans ce mémoire puisque dans notre cas nous disposons de données expérimentales. Cependant, une assignation aléatoire peut être imparfaite, c'est-à-dire que certaines variables ne sont pas bien balancées entre les deux groupes. La découverte de

variables non balancées dans la composante expérimentale de l'Étude longitudinale et expérimentale de Montréal (ÉLEM) a motivé ce projet de recherche d'exploration de différentes techniques d'appariement pour tester la sensibilité de l'effet du programme de prévention à des biais qui pourraient être causés par une randomisation imparfaite.

2 L'analyse causale en sociologie

L'analyse causale en sociologie est fortement influencée par la philosophie des sciences qui domine à une époque (Bernert, 1983). Selon la philosophie des sciences dominante, l'analyse causale sera rejetée parce que trop métaphysique ou sera promue au rang du seul type d'analyse sérieuse pouvant être menée en science. Par exemple, à la fin du XIX^e siècle, sous l'influence du phénoménalisme, dont le positivisme est l'une des incarnations, le terme causalité est de moins en moins utilisé pour être remplacé, sous l'influence du statisticien Pearson, par des termes comme 'corrélacion' ou 'association' (Bernert, 1983, p.233). Le mouvement s'accroît durant la première moitié du XX^e siècle. Le phénoménalisme continue à influencer les sciences sociales, mais vient s'ajouter un autre mouvement philosophique, la phénoménologie. Ceux qui adoptent une version ou une autre du phénoménalisme tendent à favoriser la cueillette de données et les analyses de corrélacion. Ceux qui sympathisent avec la phénoménologie semblent plus enclins aux études de cas et à l'interprétation des motivations des individus (Bunge, 2006, p.88).

Une exception notable en sociologie pendant cette longue période est celle d'Émile Durkheim, lequel tente d'identifier les causes sociales du différentiel du taux de suicide parmi les populations de différents groupes sociaux (Durkheim, 1981). Durkheim fait appel à une technique d'analyse tabulaire multivariée, pratique assez rare à cette époque, afin d'isoler des facteurs tels que le sexe, la religion ou l'état civil. Un phénoménaliste se contenterait d'une telle analyse parce qu'il lui est interdit par sa philosophie de postuler des causes au-delà des données empiriques. Durkheim est un réaliste et cherche les véritables causes qui expliqueraient ce différentiel parmi les facteurs isolés empiriquement. Il postule alors l'existence de deux mécanismes sociaux qui influenceraient le taux de suicide, soit l'intégration sociale et la régulation sociale. Remarquons que ces processus sont transempiriques. Alors que nous pouvons observer le sexe, la religion et l'état civil des individus, il est impossible d'observer directement un processus d'intégration et de régulation. Il faut donc faire appel à des indices afin de vérifier les hypothèses concernant ces inobservables. Autrement dit, Durkheim n'était pas un utilisateur naïf des techniques

multivariées naissantes. Il savait que les données empiriques seules ne peuvent servir à établir des causes puisque ces données sont justement ce qui doit être expliqué par des causes sous-jacentes.

L'analyse causale en sociologie effectue un retour à partir de la Seconde Guerre mondiale et n'a cessé depuis lors d'intéresser les sociologues. Les sociologues développent différentes formes d'analyse tabulaire. Kendall et Lazarsfeld mettent au point l'analyse connue sous le nom d'élaboration de tableaux (Kendall et Lazarsfeld, 1950). Ils proposent une interprétation des résultats d'une analyse entre X et Y lorsqu'on introduit un troisième facteur connu sous le nom de variable-test T . Selon que ce troisième facteur est associé simultanément à X et Y (variable de confusion) ou seulement à l'un ou l'autre, et selon le modèle causal retenu par le chercheur, c'est-à-dire si T précède ou suit X , alors ces auteurs avancent les interprétations possibles en examinant les résultats d'association qui résultent de l'introduction de T (Fox, 1999, Chap. 11; Morgan et Winship, 2007, p.214).

Dans les années 60, les techniques de régression pour l'analyse causale font leur apparition en sociologie alors que Blalock publie *Causal inferences in nonexperimental research* (Blalock, 1964) et Duncan, *Path Analysis: Sociological Examples* (Duncan, 1966). À la fin des années 70, une importante synthèse s'opère entre l'approche par équations structurales et l'analyse factorielle par la publication de *Structural Equation Models in the Social Sciences* (Jöreskog, 1977). Pendant la même période, Campbell et ses associés travaillent dans une perspective d'élaboration de protocoles de recherche pour études observationnelles et se font connaître par la publication de *Experimental and Quasi-Experimental Designs for Research* (Campbell, Stanley et Gage, 1966) et de *Quasi-Experimentation : Design and Analysis Issues for Field Settings* (Cook et Campbell, 1979).

Pendant les années 80 et 90, une insatisfaction grandissante se fait sentir à l'égard de l'utilisation naïve des techniques de régression. Plusieurs auteurs démontrent que l'interprétation causale des coefficients de régression n'est presque jamais justifiée (Sobel, 1996). Autrement dit, peu importe la sophistication des méthodes de régression, une

régression demeure un outil pour analyser des corrélations. Pour qu'un estimateur corrélationnel possède une interprétation causale, il faut injecter des éléments non analytiques. C'est ce que Durkheim et tous les chercheurs qui l'ont suivi ont fait. Ainsi, plusieurs ensembles de règles ont été proposés afin de justifier une telle interprétation causale des données. L'élaboration de tableaux par Kendall et Lazarsfeld n'est pas autre chose qu'un tel ensemble de règles. Les auteurs mentionnés précédemment ont consacré de nombreuses pages au problème de l'interprétation causale des coefficients des différentes techniques avancées par eux.

La règle selon laquelle l'association entre X et Y ne doit pas être due à un troisième facteur a généré une vaste littérature sur les relations fallacieuses. En fait, tous les auteurs vus précédemment ont consacré beaucoup d'énergie à cette question. Cette approche par l'analyse des relations fallacieuses s'est avérée finalement problématique. Sobel tente de démontrer que la stratégie d'inférence causale basée sur cette approche n'est pas valide. Dans un premier temps, il suppose que la majorité des chercheurs intéressés par l'analyse causale soutient implicitement une interprétation contrefactuelle de la causalité. Ainsi, dans le cas d'une relation fallacieuse, il est souvent dit que s'il était possible de faire une expérimentation et de manipuler la variable indépendante, alors nous nous apercevrons que la variable dépendante ne serait pas affectée. L'idée de manipulation est celle de faire varier l'état d'un objet pour voir *si* cette variation provoque un changement d'état sur un autre objet. On veut savoir ce qui se passe lorsque l'état du monde change. Autrement dit, l'expérimentation repose sur l'idée de contrefactualité. C'est ce que Neyman avait découvert dès le début du XXe siècle (Neyman, 1990, pour la trad. en anglais), puis ce qui fût redécouvert à plusieurs reprises et finalement formalisé par Rubin dans un premier article en 1974. Dans un deuxième temps, Sobel introduit le formalisme du modèle causal de Rubin puisque ce modèle traite explicitement de questions contrefactuelles. Finalement, il démontre que les estimateurs standards et les estimateurs dérivés du modèle contrefactuel ne sont pas équivalents en général pour des données observationnelles. Sobel justifie ainsi l'utilisation du modèle contrefactuel pour les recherches en sociologie. Puisque les

chercheurs raisonnent spontanément de façon contrefactuelle, puisque l'interprétation contrefactuelle de la causalité est à la base des études expérimentales et puisque le modèle causal de Rubin est une formalisation des études expérimentales généralisées aux études observationnelles, ce modèle est tout désigné pour la recherche en sociologie.

Le modèle causal contrefactuel est une réponse à ce besoin de disposer d'un formalisme indépendant de toutes techniques d'analyse qui puisse guider le chercheur à recueillir des données auxquelles pourra être appliqué un estimateur causal. Remarquons que le modèle causal contrefactuel n'est pas en soi quantitatif ou qualitatif. Tant les chercheurs en méthodes quantitatives qu'en méthodes qualitatives peuvent bénéficier de ce modèle. Comme nous le verrons, il s'agit d'un formalisme logico-mathématique qui permet de réfléchir à la construction d'un protocole de recherche observationnel afin que les résultats obtenus par celui-ci puissent être interprétés causalement.

3 Le modèle causal contrefactuel

Prenant pour point de départ les études expérimentales, Donald Rubin a formalisé un modèle causal qui permet de dégager les postulats qui sont nécessaires de soutenir afin qu'une étude observationnelle soit similaire à une étude expérimentale (Imbens et Rubin, 2008). Le modèle a le mérite d'attirer l'attention sur le protocole de recherche d'une étude observationnelle. Grâce aux postulats ainsi dégagés, il est possible de prévoir dans un protocole des étapes et des mécanismes de collecte de données qui assureront que l'estimation de l'effet causal sera réalisable, c'est-à-dire une estimation sans biais de cet effet.

Le modèle contrefactuel de Rubin repose sur deux concepts de base: les états causaux et les réponses potentielles. Les états causaux sont aussi nommés traitements dans la tradition de l'analyse causale contrefactuelle. Chaque unité d'analyse d'une étude doit *potentiellement* se trouver dans l'un ou l'autre des états causaux examinés. En pratique, dans le cadre d'une expérimentation, les chercheurs appliquent un traitement à un groupe afin de pouvoir le comparer à un autre groupe qui n'a pas reçu le traitement. L'assignation aléatoire d'une expérimentation garantit que chaque unité d'analyse se retrouvera potentiellement dans l'un ou l'autre des états causaux.

À ces états causaux potentiels correspondent des réponses potentielles. À chaque unité d'analyse est associée une réponse si celle-ci reçoit le traitement et une autre réponse si celle-ci ne reçoit pas le traitement. L'effet causal est alors la différence entre ces deux réponses. Dans les faits, une unité d'analyse se retrouve dans l'un ou l'autre des états causaux. Conceptuellement, il manque cinquante pourcent de l'information afin d'être en mesure d'estimer l'effet causal. Il s'agit d'un problème de données manquantes (Imbens et Rubin, 2008). Cette impossibilité d'observer simultanément deux états causaux pour une même unité d'analyse a été nommé par Holland le problème fondamental de l'inférence causale (Holland, 1986, p.947). Toujours selon Holland, il existe deux solutions à ce problème : la solution scientifique et la solution statistique.

La solution scientifique consiste soit à considérer identique un même objet à deux moments différents ou à considérer identique deux objets différents. Par exemple, un chercheur peut étudier l'effet de la chaleur sur un nouveau matériau. Il peut se procurer deux barres de ce nouveau matériau et les soumettre chacune à une température différente. Il doit alors argumenter que les deux barres sont *identiques* afin de justifier son analyse causale.

En science sociale, il est rare que nous puissions procéder de cette façon. Nous devons faire appel alors à une approche probabiliste. Il s'agit de la solution statistique. Au lieu d'observer l'effet causal sur une seule unité d'analyse, nous mesurons cet effet à un niveau agrégé. Cependant, la façon dont les cas sont assignés à l'un ou l'autre des états causaux doit être modélisée de façon probabiliste pour que l'effet causal soit estimable. Cette modélisation est souvent nommée le mécanisme d'assignation. Pour une étude expérimentale, le mécanisme d'assignation probabiliste est l'assignation aléatoire. L'assignation aléatoire permet en principe de créer des groupes équivalents en moyenne.

Le problème pratique qui se pose alors pour une étude observationnelle est celui de la formation de groupes équivalents et le problème théorique est celui des postulats qui doivent être respectés afin d'obtenir ces groupes selon un modèle probabiliste. Ainsi, pour qu'un effet causal moyen soit estimable, il faut que le mécanisme d'assignation soit probabiliste. Dans les faits, il faut identifier les variables qui permettraient d'assigner les cas à un état causal ou un autre afin de contrôler pour les effets de celles-ci. Une fois ce contrôle effectué, par une technique de régression ou d'appariement, la variabilité résiduelle entre les deux groupes peut être assimilée à une variabilité aléatoire. Nous revenons alors au modèle probabiliste de l'assignation aléatoire, ce qui permet d'utiliser les techniques classiques des études expérimentales, comme une simple différence de moyennes ou une analyse de régression ou de variance, pour évaluer l'effet causal.

À la section suivante, nous présentons de façon plus formelle la discussion qui précède, tandis que le problème pratique de la formation de groupes équivalents sera examiné à la section portant sur L'appariement pour équilibrer les variables.

3.1 Définitions et postulats d'identification de l'effet causal

La façon la plus fructueuse de voir ces postulats est celle qui les considère comme des étapes de l'élaboration d'un protocole de recherche d'une étude observationnelle. Ainsi, la formalisation par Rubin d'un modèle causal pour études observationnelles qui s'appuie sur le modèle causal des études expérimentales a un impact direct sur l'élaboration des protocoles de recherche. Nous sommes en présence d'un exemple clair où la formalisation et la théorisation viennent en aide à la pratique.

Les postulats et les définitions qui suivent ont été tirés de trois sources principales: (Hernan et Robins, en préparation, chap. 1; Imbens et Rubin, 2008; Morgan et Winship, 2007, chap. 1 et 2).

Postulat 1. Existence des états causaux

Le modèle causal contrefactuel présuppose l'existence de deux états causaux nettement délimités auxquels tous les individus de la population d'intérêt peuvent être exposés.

Ces états causaux sont représentés par la variable aléatoire binaire D . Cette variable représente l'exposition des cas à une cause qui produit deux états. Les réalisations de cette variable sont représentées par d_i , dont la valeur est un pour les cas observés dans le groupe-intervention et zéro pour les cas observés dans le groupe-contrôle.

Habituellement, les états causaux sont clairement identifiés dans une étude expérimentale. Pour des études observationnelles, il peut être plus difficile d'identifier ces états, notamment si nous disposons de données pour lesquelles toutes les précautions n'ont pas été prises afin de s'assurer que l'effet causal est identifiable.

Postulat 2. Existence de réponses potentielles

Étant donné l'existence d'états causaux nettement délimités, on stipule l'existence de variables aléatoires de réponses potentielles définies pour tous les individus de la population d'intérêt. Les variables aléatoires de réponses potentielles sont dénotées par Y^1 et Y^0 . Les réalisations de ces variables sont représentées par y_i^1 et y_i^0 .

Ainsi y_i^1 est la réponse potentielle dans l'état causal associé au traitement pour chaque individu et y_i^0 est la réponse potentielle dans l'état causal associé à la situation contrôle pour chaque individu. Autrement dit, le paradigme contrefactuel se représente tous les individus, peu importe s'ils se retrouvent au final dans le groupe-intervention ou dans le groupe-contrôle, comme pouvant produire une réponse comme s'ils se trouvaient soit dans le groupe-intervention soit dans le groupe-contrôle.

Les réponses potentielles existent autant pour les études expérimentales qu'observationnelles. Nous verrons cependant que les réponses potentielles sous conditions expérimentales sont estimables dès qu'on avance un postulat d'indépendance de réponses potentielles des deux états causaux. En ce qui concerne les études observationnelles, un postulat d'indépendance conditionnelle doit être avancé.

Définition 1. L'effet causal individuel

Le paradigme contrefactuel a pour point de départ l'effet causal individuel défini par

$$\delta_i^1 = y_i^1 - y_i^0 \quad (1)$$

Pour un même individu i , l'effet causal est la différence entre ses deux réponses potentielles. Remarquons que l'effet causal est défini sur les réponses potentielles, c'est-à-dire sur des concepts abstraits au sens où ces concepts ne réfèrent pas directement à la réalité. Si nous pouvions observer un individu dans les deux états causaux, alors nous pourrions calculer l'effet causal pour cet individu.

Définition 2. La réponse observée

La réponse observée chez les individus est définie par

$$Y = \begin{cases} Y^1 & \text{si } D = 1 \\ Y^0 & \text{si } D = 0 \end{cases} \quad (2)$$

Autrement dit, pour les cas qui se trouvent effectivement dans le groupe-intervention $D = 1$, leur réponse observée Y est équivalente à leur réponse potentielle dans l'état hypothétique d'intervention Y^1 . De même, pour les cas qui se trouvent effectivement dans le groupe contrôle $D = 0$, leur réponse observée Y est équivalente à leur réponse potentielle dans l'état hypothétique de contrôle Y^0 .

Dans la littérature, cette définition est souvent réécrite de cette façon

$$Y = DY^1 + (1 - D)Y^0 \quad (3)$$

Une autre façon d'écrire cette équation est de la façon suivante

$$Y = Y^0 + D(Y^1 - Y^0) \quad (4)$$

Cette façon de réécrire l'équation permet de voir que l'effet causal est conçu comme l'addition de la différence entre les deux réponses potentielles à la réponse potentielle lorsqu'il n'y a aucune intervention. On peut aussi dire qu'une réponse potentielle est réalisée une fois l'individu exposé à l'un des états causaux. Ceci a pour conséquence que l'autre réponse potentielle sera inobservée. Cette impossibilité a été nommée par Holland *le problème fondamental de l'inférence causale* (Holland, 1986). Ce problème peut être illustré par un tableau croisé :

Tableau 1 : Le problème fondamental de l'inférence causale

Exposition causale	Réponses potentielles	
	Y^1	Y^0
Intervention ($D = 1$)	$Y = Y^1$	$Y = ?$
Contrôle ($D = 0$)	$Y = ?$	$Y = Y^0$

Les valeurs sur la diagonale principale sont *observées* et celles sur la diagonale secondaire sont *inobservables* (cellules grisées), une fois les individus exposés à la cause. Une question *contre-factuelle* est une question qu'on se pose quant à la réponse qu'aurait eue l'individu s'il avait été exposé à l'autre état. Ainsi, pour ceux du groupe-intervention, qu'aurait été leur réponse observée s'ils avaient fait partie du groupe-contrôle? De même, pour ceux du groupe-contrôle, qu'aurait été leur réponse observée s'ils avaient fait partie du groupe-intervention?

L'impossibilité d'observer le même individu dans les deux états causaux entraîne qu'on ne peut mesurer l'effet causal au niveau individuel. Les statisticiens se tournent donc vers l'estimation de divers effets moyens. Cependant, l'estimation de ces effets moyens nécessite de soutenir le postulat suivant :

Postulat 3. *Stable unit treatment value assumption (SUTVA)*

Formalisé par Rubin (Rubin, 1980), ce postulat est constitué de deux postulats. Le premier est le postulat qui stipule qu'aucune interférence ne doit se produire entre les individus des deux groupes. Le deuxième postulat exprime l'idée qu'il n'existe qu'une version du traitement.

La première contrainte incite le chercheur à s'assurer qu'il n'y a pas d'interaction entre les deux groupes. En fait le postulat est moins sévère que ce qui vient d'être énoncé. S'il y a interaction entre les deux groupes, mais que le chercheur peut prouver que cette

interaction n'interfère pas avec les réponses des individus sur la variable d'intérêt, alors le postulat est observé (Hernan et Robins, en préparation, p.5).

La deuxième contrainte oblige le chercheur à s'assurer que le traitement est donné de la même manière à tous les individus du groupe-intervention, mais aussi que tous les individus du groupe-contrôle soient contrôlés de la même façon. S'il y a des variations pour le traitement ou pour le contrôle, alors il faut démontrer que ces variations n'affectent pas les réponses des individus (Hernan et Robins, en préparation, p.6).

Si un chercheur peut soutenir ce postulat, alors il peut estimer différents effets moyens. Le plus simple est l'effet moyen inconditionnel du traitement (*average treatment effect*, ATE) défini de la façon suivante :

Définition 3. Effet moyen inconditionnel du traitement

$$\begin{aligned} E[\delta] &= E[Y^1 - Y^0] \\ &= E[Y^1] - E[Y^0] \end{aligned} \quad (5)$$

Remarquons que cette définition fait appel au concept de réponse potentielle, ce qui entraîne que cette quantité n'est pas plus mesurable que l'effet causal défini pour un individu. Cependant, il est possible de formuler le problème au niveau agrégé. À ce niveau, nous avons la moitié de l'information nécessaire à l'estimation de l'ATE (ATE: average treatment effect; effet moyen de traitement). Regardons le tableau suivant :

¹ Le symbole δ est utilisé dans la littérature pour désigner un effet causal quelconque, que ce soit l'ATE ou un autre effet causal.

Tableau 2 : Détails du problème fondamental de l'inférence causale

	D	Y	Y^1	Y^0
1	1	O	O	?
2	1	O	O	?
3	1	O	O	?
4	0	O	?	O
5	0	O	?	O
6	0	O	?	O

Nous sommes en présence de six individus dont trois se trouvent dans le groupe-intervention $D = 1$ et trois dans le groupe-contrôle $D = 0$. Les cases comprenant le signe d'interrogation représentent les réponses potentielles pour lesquelles nous ne disposons pas d'information et les cases comprenant un O majuscule représentent les réponses observées. La réponse observée Y dépend du groupe auquel appartient l'individu et est égale à la réponse potentielle associée à chaque état causal. Ainsi, pour les individus un, deux et trois, $Y = Y^1$ et pour les individus quatre, cinq et six $Y = Y^0$. Le problème de l'estimation de l'ATE est alors conceptualisé comme un problème de données manquantes. Sous quelles conditions pourrions-nous imputer des valeurs aux données manquantes afin d'estimer l'ATE?

Pour répondre à cette question, décomposons l'ATE :

$$\begin{aligned}
E[\delta] &= E[Y^1 - Y^0] \\
&= E[Y^1] - E[Y^0] \\
&= \left\{ \pi E[\cancel{Y^1|D=1}] + (1-\pi) E[Y^1|D=0] \right\} \\
&\quad - \left\{ \pi E[Y^0|D=1] + (1-\pi) E[\cancel{Y^0|D=0}] \right\} \\
&= \left\{ \pi E[Y|D=1] + (1-\pi) E[Y^1|D=0] \right\} \\
&\quad - \left\{ \pi E[Y^0|D=1] + (1-\pi) E[Y|D=0] \right\}
\end{aligned} \tag{6}$$

Les deux premières égalités constituent la définition de l'ATE telle que représentée par l'équation (5). L'avant-dernière et la dernière égalité sont des moyennes pondérées de cet ATE, l'avant-dernière étant constituée que des réponses potentielles alors que la dernière est constituée de réponses potentielles et de réponses observées.

Nous pouvons lire cette équation à l'aide du Tableau 2. De la deuxième égalité, $E[Y^1] - E[Y^0]$, si nous disposions de toutes les réponses potentielles, nous pourrions directement soustraire la moyenne de la colonne Y^1 de la moyenne de la colonne Y^0 . De là la troisième égalité,

$$\begin{aligned}
&\left\{ \pi E[\cancel{Y^1|D=1}] + (1-\pi) E[Y^1|D=0] \right\} \\
&- \left\{ \pi E[Y^0|D=1] + (1-\pi) E[\cancel{Y^0|D=0}] \right\}
\end{aligned}$$

nous constatons que la moyenne générale $E[Y^1]$ de la deuxième égalité est égale, dans la troisième égalité, à la moyenne pondérée $\pi E[Y^1|D=1]$ du groupe-intervention, additionnée à la moyenne pondérée $(1-\pi) E[Y^1|D=0]$ du groupe-contrôle. La même décomposition s'établit pour la moyenne générale de Y^0 .

Cette décomposition permet de voir que la troisième égalité contient cinq variables dont deux ne peuvent être estimées pour l'instant. Il s'agit de la moyenne de la réponse

potentielle dans la situation contrôle pour les individus dans la situation intervention, $E[Y^0 | D=1]$, et de la moyenne de la réponse potentielle dans la situation intervention pour ceux dans la situation contrôle $E[Y^1 | D=0]$ (les blocs gris pâles du tableau précédent). Plus simplement, ces deux termes représentent les moyennes qu'auraient les groupes d'individus s'ils étaient observés dans le groupe auquel ils n'appartiennent pas. Cependant, les trois autres variables de la troisième égalité peuvent être estimées. La proportion π peut être estimée par la proportion d'échantillonnage. La moyenne de la réponse potentielle Y^1 pour le groupe $D=1$, $E[Y^1 | D=1]$, est simplement la moyenne de Y pour le groupe $D=1$, $E[Y | D=1]$, et la moyenne de la réponse potentielle Y^0 pour le groupe $D=0$, $E[Y^0 | D=0]$ est simplement la moyenne de Y pour le groupe $D=0$ $E[Y | D=0]$. L'identification des réponses potentielles aux réponses observées de Y découle de la Définition 2 par substitution (équation (6)).

Autrement dit, dans la dernière égalité de l'équation (6),

$$\begin{aligned} & \{ \pi E[Y | D=1] + (1-\pi) E[Y^1 | D=0] \} \\ & - \{ \pi E[Y^0 | D=1] + (1-\pi) E[Y | D=0] \} \end{aligned}$$

les réponses potentielles $E[Y^0 | D=1]$ et $E[Y^1 | D=0]$ ne sont pas estimables sans faire appel à un postulat supplémentaire.

Postulat 4. Égalité des réponses potentielles dans les deux groupes

$$E[Y^1 | D=0] = E[Y^1 | D=1] \tag{7}$$

$$E[Y^0 | D=1] = E[Y^0 | D=0] \tag{8}$$

L'égalité (7) se lit de la façon suivante : en moyenne, les deux groupes réagissent de la même façon au traitement. De même, l'égalité (8) se lit de la façon suivante : en moyenne, les deux groupes réagissent de la même façon s'ils ne sont pas traités.

Si nous remplaçons ces deux égalités dans (6), alors les deux dernières inconnues sont estimables. En fait, la formule (6) se réduit alors à la différence de moyenne de Y entre les deux groupes :

$$E[\delta] = E[Y | D = 1] - E[Y | D = 0] \quad (9)$$

Toujours à l'aide du tableau Tableau 2, nous remarquons que cette égalité est tout simplement la différence de moyenne de Y entre les deux groupes (colonne Y). Nous constatons aussi que les réponses potentielles ont disparu. Ceci revient à dire que nous considérons les deux groupes similaires en moyennes, ce qui nous permet d'imputer les réponses observées d'un groupe aux réponses potentielles de l'autre groupe dans l'état inobservé de ce groupe. Ainsi, $E[Y | D = 0]$ peut être imputé à $E[Y^0 | D = 1]$ et $E[Y | D = 1]$ peut être imputé à $E[Y^1 | D = 0]$. Dans le Tableau 2, ceci permet de copier les valeurs des cellules moyennement grisées dans les cellules faiblement grisées, en dessous dans le cas de Y^1 et au-dessus dans le cas de Y^0 . Nous constatons déjà qu'il s'agit d'une forme d'appariement des cas. Lorsque nous soustrayons les moyennes entre deux groupes dans une étude avec assignation aléatoire, nous prétendons que les individus des deux groupes sont les mêmes, mais observés dans deux états différents. Pour bien voir cette situation, restructurons le Tableau 2 :

Tableau 3 : Identification des cas-contrôles aux cas-interventions

	D	Y^1	Y^0	Y^0	D	
1	1	O	?	O	0	4
2	1	O	?	O	0	5
3	1	O	?	O	0	6

Nous avons aligné les cas quatre, cinq et six aux cas un, deux et trois. Ce qui revient à dire que nous les avons appariés. Ainsi la réponse potentielle Y^0 des cas quatre, cinq et six (cellule moyennement grisées à droite), observée pour ces cas, est assimilée à la réponse potentielle Y^0 des cas un, deux et trois (cellules faiblement grisées au centre), inobservée pour ces cas. Rappelons que ceci est possible grâce au Postulat 4.

Maintenant, dans quelles circonstances pouvons-nous adopter ce postulat? Autrement dit, dans quelles circonstances les réponses potentielles sûres Y^1 sont-elles identiques en moyenne dans les deux groupes, égalité (7), et dans quelles circonstances les réponses potentielles sur Y^0 sont-elles identiques en moyenne dans les deux groupes, égalité (8). Il faut que les réponses potentielles soient indépendantes de l'assignation des cas aux deux groupes.

Postulat 5. Indépendance du statut de traitement et des réponses potentielles

$$(Y^0, Y^1) \perp D \quad (10)$$

Dans le jargon du paradigme contrefactuel, on dit que le mécanisme d'assignation peut être ignoré (*ignorable assignment mechanism*). **Le seul type de mécanisme d'assignation qui rencontre le Postulat 5 est le mécanisme d'assignation aléatoire. C'est pour cette raison qu'on peut comparer directement la moyenne des deux groupes lors d'une étude avec assignation aléatoire.** Remarquons que l'assignation

aléatoire peut être conceptualisée comme une forme d'appariement. Si nous disposons d'une population, nous pouvons sélectionner au hasard un individu. À cet individu, nous pouvons appairer un autre individu sélectionné au hasard. Nous poursuivons de cette façon jusqu'à ce qu'on atteigne le nombre désiré d'individus dans chaque groupe. En principe, nous disposons alors de deux groupes identiques en moyenne.

À cette étape-ci de notre exposé il est utile de remarquer que les définitions et postulats précédents caractérisent les études expérimentales. Autrement dit, jusqu'ici, tout ce qui a été présenté est une formalisation de la pratique de l'assignation aléatoire. Les prochains postulats et définitions devront établir les conditions nécessaires afin qu'une étude observationnelle dispose de caractéristiques semblables à celles d'une étude expérimentale. Notre exposé peut être vu comme une montée puis une descente où le Postulat 5 est le point culminant et la Définition 4, ci-dessous, le point d'arrivée.

Le Postulat 5 est très restrictif. Les études observationnelles se caractérisent en général par l'absence d'une assignation aléatoire. Existe-t-il alors un postulat d'indépendance moins restrictif que le Postulat 5? Si nous disposons d'une ou plusieurs variables qui influencent l'assignation des cas aux deux groupes, alors, en contrôlant pour ces variables, nous brisons le lien entre (Y^0, Y^1) et D , ce qui les rend indépendants.

Postulat 6. Indépendance conditionnelle du statut et des réponses potentielles

$$(Y^0, Y^1) \perp D | S \quad (11)$$

Le Postulat 6 introduit explicitement un mécanisme d'assignation représenté par S . Autrement dit, l'appartenance à l'un des deux groupes représentés par D dépend d'un mécanisme d'assignation représenté par S . Si ce mécanisme est connu et mesuré, alors nous pouvons contrôler ces variables et rendre D statistiquement indépendant des réponses potentielles.

Le Postulat 6 nous permet de réécrire le Postulat 4 sous une variante conditionnelle.

Postulat 7. Égalité conditionnelle des réponses potentielles

$$E[Y^1 | D = 0, S] = E[Y^1 | D = 1, S] \quad (12)$$

$$E[Y^0 | D = 1, S] = E[Y^0 | D = 0, S] \quad (13)$$

Finalement, ce dernier postulat nous permet de réécrire la Définition 3 sous sa variante conditionnelle.

Définition 4. Effet moyen conditionnel du traitement

$$\begin{aligned} E[\delta | S] &= E[Y^1 - Y^0 | S] \\ &= E[Y^1 | S] - E[Y^0 | S] \end{aligned} \quad (14)$$

Si nous procédions à la dérivation complète de cette formule, en substituant aux bons endroits les égalités (12) et (13), nous obtiendrions cet estimateur :

$$E[\delta | S] = E[Y | D = 1, S] - E[Y | D = 0, S] \quad (15)$$

Nous remarquons que les réponses potentielles ont disparu. Si nous supposons que S est une variable nominale, cette formule nous dit que nous pouvons comparer les deux groupes à l'intérieur de chacune des catégories de S . Autrement dit, les cas des deux groupes sont similaires à l'intérieur des catégories de S . La formule (15) est un peu plus compliquée si le nombre de cas par catégorie diffère. Il faut alors pondérer le calcul. Nous pouvons lire l'équation (15) à l'aide du Tableau 4:

Tableau 4 : Une stratification comme résolution du problème de l'inférence causale

	D	S	Y	Y^1	Y^0
1	1	0	O	O	?
2	1	0	O	O	?
3	1	1	O	O	?
4	1	1	O	O	?
5	0	0	O	?	O
6	0	0	O	?	O
7	0	1	O	?	O
8	0	1	O	?	O

Dans ce tableau, S possède deux catégories représentées par zéro et un. Nous pouvons comparer les cas un et deux avec les cas cinq et six et nous pouvons comparer les cas trois et quatre avec les cas sept et huit puisqu'à l'intérieur de chaque catégorie de S , les cas ne diffèrent que par leur appartenance à un des deux groupes. Il s'agit ici en fait de la stratégie d'appariement par stratification, à la base de la logique de toutes les stratégies d'appariement. L'idée derrière la stratification est celle de considérer que si S est vraiment le mécanisme d'assignation des cas à D , alors, une fois la stratification effectuée, les différences qui devraient subsister entre les deux groupes pour chaque strate de S ne seraient dues qu'au hasard.

Le modèle causal contrefactuel est un formalisme qui justifie ce que les chercheurs pratiquaient intuitivement. L'avantage de ce formalisme est de mettre à jour les postulats qu'on doit défendre à chaque étape d'une étude. Ce formalisme attire aussi l'attention sur la nécessité de bien monter un protocole de recherche, de prévoir la façon de satisfaire les postulats par la collecte d'information, de prévoir les défauts possibles afin de recueillir des

informations sur ces défauts, informations qui serviront à corriger les données à l'étape de l'analyse afin d'obtenir une estimation sans biais de l'effet causal.

De ce qui précède, nous constatons qu'il est crucial d'identifier le mécanisme d'assignation S , c'est-à-dire les variables pour lesquelles nous avons de bonnes raisons de penser qu'elles influencent l'appartenance à l'un ou l'autre groupe à l'étude. Une fois identifiées ces variables, nous devons choisir les techniques que nous utiliserons pour estimer l'effet causal. Une pratique largement répandue consiste à utiliser l'appariement afin d'équilibrer les deux groupes sur les variables à contrôler, puis à utiliser un modèle de régression adapté à la nature des données. Par exemple, dans notre cas, une fois équilibrés les deux groupes sur les variables identifiées, nous utiliserons un modèle de régression multiniveaux pour évaluer l'effet causal parce que nos données sont de nature longitudinale. Les avantages de cette approche sont nombreux : elle est facile à comprendre et à mettre en application; les chercheurs n'ont pas à mettre de côté l'expérience qu'ils ont développée pour la construction de modèles de régression; les variables utilisées lors de l'appariement n'ont pas à être introduites dans un modèle de régression, ce qui réduit l'impact des problèmes de colinéarité et de forme fonctionnelle du modèle (Ho *et coll.*, 2007; Imai *et coll.*, 2011; Morgan et Winship, 2007, p.69). Nous examinerons dans la section suivante les caractéristiques de trois méthodes d'appariement et nous les appliquerons à nos données à la dernière section.

4 L'appariement pour équilibrer les variables

L'appariement repose sur l'idée d'homogénéité (Sekhon, 2008). Ceci se traduit par la formule populaire, comparer ce qui est comparable. L'idée remonte à loin. Francis Bacon traite de celle-ci dans son *Novum Organum* paru en 1640 (Bacon, 1986) et John Stuart Mills fait reposer une partie de sa méthode sur celle-ci dans son *System of Logic* paru en 1843 (Mill, 1970). Nous pouvons retracer l'utilisation de l'appariement jusqu'à au moins Fechner, le père de la psychophysique, dans son *Elements of psychophysics, Vol 1*, paru en 1860. Fechner utilisait l'appariement exact (Sekhon, 2008). Dans certaines disciplines on nomme l'appariement exact une stratification ou une sous-classification (*subclassification*) et dans d'autres, notamment en sociologie, on connaît ce type d'appariement sous le nom de la méthode de l'élaboration de tableaux.

Prenons l'exemple général suivant. Nous disposons de trois variables dichotomiques : Y , la variable d'intérêt ou variable dépendante, D , la variable indépendante principale ou encore une intervention quelconque et S , une variable indépendante ou encore une variable-contrôle ou variable-test. S est aussi une variable antécédente mesurée avant l'intervention. Nous voulons donc analyser l'effet de D sur Y en contrôlant S . Le tableau final d'une élaboration de tableau de cette analyse se présenterait ainsi :

Tableau 5: Élaboration de tableaux

		S			
		1		0	
		D		D	
Y		1	0	1	0
1					
0					

Nous avons stratifié la relation entre Y et D selon S . Ou encore, nous avons effectué un appariement exact des cas selon les valeurs de S , peu importe leur valeur en D ,

i.e. que tous les cas-intervention et tous les cas-contrôles qui ont $S = 1$ ont été regroupés (appariés) et de même pour les cas qui ont $S = 0$. De cette façon nous pouvons comparer les cas-contrôles et les cas-intervention dans chaque strate de S parce que les cas sont identiques selon les strates de S ou encore il y a homogénéité des caractéristiques des cas selon les strates de S . Si S représente toutes les caractéristiques qui différencieraient les cas avant l'intervention, maintenant, après l'appariement exact ou la stratification, les cas ne diffèrent plus que sur D . Comme on peut le constater, l'expression « appariement » (*matching*) est mal choisie. Dans le cas de l'appariement exact, il n'y a pas d'appariement au sens strict, c'est-à-dire l'appariement des cas un à un. On forme plutôt des sous-groupes les plus homogènes possible, c'est-à-dire pour lesquels les variables-tests sont distribuées sensiblement de la même façon entre le groupe-intervention et le groupe-contrôle. Ceci vaut pour les trois méthodes d'appariement que nous allons examiner. Il n'y a pas formation de données appariées, mais plutôt constitution d'un groupe-intervention et d'un groupe-contrôle de telle sorte que les caractéristiques des deux groupes sont semblables en moyenne.

L'appariement exact impose de sérieuses limites quant au nombre de variables pour lesquelles nous pouvons contrôler. Le nombre de cellules du tableau augmente de façon exponentielle avec l'ajout de chaque nouvelle variable (Rosenbaum, 2010, p.72). Dans notre exemple, nous avons trois variables à deux valeurs chacune et nous sommes déjà à huit cellules (2^3). Cinq variables dichotomiques donneraient trente-deux cellules. Le danger est de se retrouver avec des cellules vides pour lesquelles aucun cas n'existe selon les caractéristiques retenues pour l'appariement ou la stratification. Ce problème est connu sous le nom du problème de multidimensionnalité. De plus, l'appariement exact ne permet pas de contrôler pour des variables continues à moins de former des catégories avec ces variables.

La régression est une technique puissante et flexible. Un problème qui peut être résolu à l'aide d'un tableau peut aussi l'être à l'aide d'une régression. En fait, il y a un lien direct entre un tableau et une régression. Prenons le cas d'un tableau 2×2 , i.e. le cas de deux

variables dichotomiques. Nous pouvons calculer le rapport de cote directement à partir des fréquences du tableau, mais nous pouvons aussi obtenir ce rapport de cote à l'aide d'une régression logistique. Cette constatation se généralise à des tableaux à plusieurs variables, elles-mêmes constituées de plusieurs catégories. De plus, la régression logistique nous fournira de l'information supplémentaire qui servira à prédire (au sens statistique) et elle permet aussi d'analyser les corrélations partielles entre la variable dépendante et les variables indépendantes (sans termes d'interaction). Tout ceci est impossible à l'aide d'un tableau. Par exemple, un tableau analyse toujours des interactions. Ajouter une variable-test à un tableau et ainsi créer des sous-tableaux, exercice typique de l'élaboration de tableaux, c'est en fait analyser une interaction. Ceci se traduit par des termes d'interaction dans une équation de régression (Morgan et Winship, 2007). Par exemple, le Tableau 5 en page 24, se traduit de la façon suivante sous forme d'équation de régression :

$$Y = \beta_0 + \beta_1(D) + \beta_2(S) + \beta_3(D * S) \quad (16)$$

Cette équation peut être résolue à l'aide d'une régression logistique puisque notre variable d'intérêt est dichotomique. La variable D doit demeurer dichotomique si nous voulons que cette variable représente une intervention sur deux groupes. Par contre, S peut maintenant être une caractéristique mesurée sur une échelle d'intervalle-ratio. La régression n'est cependant pas exempte du problème de multidimensionnalité. Implicitement, les algorithmes de régression créent des « cellules », lesquelles peuvent se retrouver vides s'il y a un trop grand nombre de variables-contrôles.

Malgré ces limitations, les techniques de régression et l'élaboration de tableaux (stratification ou appariement exact) demeurent des outils utiles et puissants. Un tableau, tout comme un graphique, donne une certaine image des données qu'une équation de régression ne peut rendre, tandis qu'une régression peut fournir des informations supplémentaires utiles au chercheur.

Les méthodes d'appariement que nous examinerons ont été élaborées en partie pour résoudre le problème de multidimensionnalité de la stratification ainsi que le problème de

la forme des relations de la régression. Ces méthodes sont souvent qualifiées d'appariements multivariés et non paramétriques, mais ce n'est pas ce qui les différencie de la stratification ou de la régression. L'élaboration de tableaux est une méthode non paramétrique par définition puisqu'il n'est pas question d'estimer des paramètres à l'aide de cette méthode. Les techniques de régression dont les variables indépendantes sont toutes des variables factices, i.e., des variables obtenues en nominalisant puis en dichotomisant des variables continues, ordinales ou nominales, sont équivalentes à des techniques non paramétriques bien qu'on qualifie souvent cette approche de surparamétrisation. Et puisqu'une variable dichotomique indépendante entretient toujours une relation linéaire avec la variable d'intérêt, il n'y a plus de problème de forme de la relation entre celle-ci et la variable d'intérêt, tout comme pour une technique non paramétrique.

Ce qui caractérise les méthodes d'appariement multivariées, c'est qu'elles sont à la fois non paramétriques, qu'elles résolvent le problème de multidimensionnalité et qu'enfin elles peuvent s'accommoder des variables continues sans créer un problème de forme de la relation entre les variables indépendantes et la variable d'intérêt. Les trois méthodes que nous examinerons sont, dans leur ordre d'apparition historique, l'appariement par la distance de Mahalanobis (Mahalanobis, 1927), l'appariement par le score de propension (Rosenbaum et Rubin, 1983) et l'appariement génétique (Diamond et Sekhon, 2005). Mais avant d'étudier ces méthodes une à une, examinons ce qu'elles ont en commun.

4.1 Caractéristiques générales des méthodes d'appariement multivariées

Les trois méthodes que nous présenterons se caractérisent par un type d'appariement affinement invariant (*affinely invariant matching method*) (Sekhon, 2008). Si les distributions des variables sur lesquelles nous voulons appairer nos cas (afin d'équilibrer ces variables), habituellement des variables mesurées avant l'intervention, sont distribuées de façon ellipsoïdale (comme la distribution normale ou la distribution de Student), alors

les trois méthodes d'appariement acquièrent la propriété de réduire le biais de ces variables de façon égale pour toutes leurs combinaisons linéaires (*equal percent bias reduction; EPBR*) (Sekhon, 2008).

L'EPBR est une propriété qui vaut pour une population infinie. Autrement dit, la distribution des variables d'appariement dans une population peut être distribuée de façon ellipsoïdale, mais ne pas l'être dans un échantillon. Si cette situation se présente, alors l'appariement par la distance de Mahalanobis et par le score de propension ne sera pas EPBR. Lorsqu'une méthode d'appariement n'est pas EPBR, alors l'algorithme crée des biais pour certaines fonctions linéaires des variables d'appariement même si les moyennes des deux groupes pour chaque variable sont plus près l'une de l'autre après l'appariement (Rubin, 1976). Dans le cas de l'appariement par le score de propension, si ce dernier n'a pas la caractéristique d'être EPBR à cause de la distribution des variables, il est possible de remédier à la situation seulement si on connaît le modèle du vrai score de propension (et non pas une estimation de celui-ci) et que l'échantillon est large (Sekhon, 2008). En général, on connaît le modèle du vrai score de propension seulement dans un contexte expérimental et quelquefois dans un contexte observationnel lorsqu'on suit les recommandations pour l'élaboration d'un protocole de recherche observationnel (Rubin, 2007).

L'appariement génétique, le dernier né des méthodes d'appariement affinement invariant, est pratiquement insensible aux distributions des variables, ce qui revient à dire que la propriété EPBR n'a pas à être présente pour que l'algorithme apparie les cas de façon efficiente, bien que la présence de cette propriété augmente l'efficacité de l'algorithme génétique :

Genetic Matching is shown to have better properties than the usual alternative matching methods both when the EPBR property holds and when it does not [...]. Even when the EPBR property holds and the mapping from X to Y is linear, Genetic Matching has better efficiency — i.e., lower mean square error (MSE)—in finite samples. When the EPBR property does not hold as it generally does not, Genetic Matching retains appealing properties

and the differences in performance between Genetic Matching and the other matching methods can become substantial both in terms of bias and MSE reduction. In short, at the expense of computer time, Genetic Matching dominates the other matching methods in terms of MSE when assumptions required for EPBR hold and, even more so, when they do not (Sekhon, 2008, p.16).

Les trois méthodes que nous avons examinées se suivent historiquement, l'appariement par la distance de Mahalanobis étant la première, suivi de l'appariement par le score de propension puis finalement nous retrouvons l'appariement génétique. Chaque méthode est censée être meilleure que la précédente, dans le sens où l'ensemble des variables servant à l'appariement doit être mieux équilibré entre les deux groupes lorsqu'on utilise une méthode plus récente qu'une autre. On s'attend donc à ce résultat lorsque nous appliquerons ces méthodes à nos données.

Ces trois méthodes d'appariement ont un postulat en commun : l'absence de biais de sélection sur les variables inobservées. Autrement dit, les trois méthodes postulent que toutes les variables pertinentes à l'appariement ont été identifiées. L'assignation aléatoire assure que les groupes sont balancés en fonction des différentes variables, qu'elles soient observées ou non, tandis que l'appariement ne peut qu'assurer l'équilibre des variables mesurées. Examinons maintenant les trois méthodes d'appariement une à une.

4.2 L'appariement par la distance de Mahalanobis

La distance de Mahalanobis fût conçue par Prasanta C. Mahalanobis en 1927 (Mahalanobis, 1927). Celle-ci est une généralisation de la distance euclidienne. La formule de la distance euclidienne entre deux variables est la suivante :

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (17)$$

Nous pouvons imaginer deux points dans un plan cartésien à deux dimensions. Cependant, la distance euclidienne n'est pas invariante aux échelles de mesure. En utilisant l'écart-type combiné des variables, nous obtenons la distance euclidienne normalisée :

$$d(x, y) = \sqrt{\sum_{i=1}^n \frac{(x_i - y_i)^2}{s_i^2}} \quad (18)$$

Cette dernière distance est invariante aux échelles de mesure. La distance de Mahalanobis est une généralisation de cette dernière mesure. Si nous considérons une matrice X de variables et une matrice S de variance-covariance de ces variables, nous formulons la distance de Mahalanobis de la façon suivante :

$$d(X_i, X_j) = \sqrt{(X_i - X_j)^T S^{-1} (X_i - X_j)} \quad (19)$$

Si S est la matrice identité, alors nous retrouvons la distance euclidienne et si en plus la matrice identité est diagonale, nous retrouvons la distance euclidienne normalisée. Ainsi, la distance entre deux observations peut être calculée et ce peu importe le nombre de variables et les échelles de mesure utilisées par celles-ci. Remarquons que la distance est calculée entre chaque cas-intervention et tous les cas-contrôle, un à un. Ainsi, si nous disposons de cinq cas-interventions et de dix cas-contrôles potentiels, nous obtenons cinquante mesures de distance. Ensuite, nous pouvons apparier les cas sur la base de ces distances. En général, l'approche utilisée est celle de l'appariement prédateur (*greedy matching*) sans remplacement. L'algorithme ordonne les cas-intervention de façon aléatoire. Le premier cas-intervention est pigé puis apparié au cas-contrôle dont la distance est la plus petite de ce cas-intervention. Les deux cas sont ensuite retirés de l'ensemble. Le processus se poursuit ainsi pour tous les cas-intervention, un à la suite de l'autre (Guo et Fraser, 2010, p.145).

Cette approche a au moins deux limitations théoriques. La première concerne la distance de Mahalanobis et la seconde concerne l'appariement prédateur. Dans le cas de la distance de Mahalanobis, si la matrice X contient plus d'une variable continue,

l'estimateur de l'effet causal contient alors un terme de biais qui ne converge par vers zéro même asymptotiquement (Abadie et Imbens, 2006). Quant à l'appariement prédateur, il se peut que celui-ci ne produise pas un appariement optimal au sens où la distance moyenne entre les cas appariés ne soit pas la plus petite possible. Par exemple, il se peut que la distance du cas-contrôle apparié au premier cas-intervention soit plus grande qu'entre ce cas-contrôle et un autre cas-intervention. En réalité, ce cas-contrôle aurait dû être apparié au deuxième cas-intervention, mais comme la distance entre ce cas-contrôle et le premier cas-intervention était la plus petite des distances entre le premier cas-intervention et tous les cas-contrôles, alors le cas-contrôle en question fût apparié au premier cas-intervention. Malgré ces limitations théoriques, nous verrons lors de l'examen des résultats des appariements que dans certaines circonstances la distance de Mahalanobis performe mieux que le score de propension, du moins selon les critères retenus et les données en main. Examinons maintenant l'appariement par le score de propension.

4.3 L'appariement par le score de propension²

Le score de propension fût introduit par Rubin et Rosenbaum (Rosenbaum et Rubin, 1983). Pour bien comprendre le rôle du score de propension dans une étude observationnelle, il est utile d'examiner sa nature dans une étude expérimentale. Dans une étude expérimentale, le score de propension est déterminé par le chercheur, alors que dans une étude observationnelle celui-ci doit être estimé. Si un chercheur, dans une étude expérimentale, désire former deux groupes d'environ le même nombre de sujets, il procédera à une assignation aléatoire simple et de cette façon chaque cas aura une chance sur deux de se retrouver dans le groupe-intervention. Le score de propension est alors de 0,5. Symboliquement, nous écrivons,

² Je suis de prêt l'exposé de S. L. Morgan et C. Winship, *Counterfactuals and causal inference: Methods and principles for social research* (New York NY: Cambridge University Press, 2007), p. 74-77, 98-105, 42-51. Cet exposé est l'un des plus éclairants sur le sujet. Je mets de côté cependant les références au modèle contrefactuel de la causalité, ce qui n'empêche pas selon nous de comprendre l'essentiel sur le score de propension.

$$\Pr[D = 1] = 0,5, \quad (20)$$

où D est la variable dichotomique qui indique l'appartenance au groupe-intervention ($D = 1$) ou au groupe-contrôle ($D = 0$).

Ce chercheur peut aussi mettre sur pied un schéma d'assignation aléatoire plus complexe. Par exemple, il peut diviser (stratifier) le groupe selon le sexe puis assigner une probabilité différente aux hommes et aux femmes d'appartenir au groupe-intervention. Symboliquement, nous aurions, par exemple,

$$\begin{aligned} \Pr[D = 1 | \text{Sexe} = \text{Femme}] &= 0,7 \\ \Pr[D = 1 | \text{Sexe} = \text{Homme}] &= 0,5 \end{aligned} \quad (21)$$

Nous obtenons deux scores de propension, un pour les femmes et un autre pour les hommes et à l'intérieur de chaque strate l'assignation est aléatoire, mais différente pour les deux sexes. Peu importe le schéma d'assignation aléatoire utilisé, simple ou stratifié, le score de propension est déterminé par le chercheur. Bien que nous nous trouvions dans un cadre expérimental, la différence de moyenne entre le groupe-intervention et le groupe-contrôle n'est plus possible parce que la variable sexe joue un rôle central dans le mécanisme d'assignation ou de sélection. Dans le langage de l'élaboration de tableaux, la variable sexe devient une variable-test afin de tenir compte de la propension de chaque sexe d'appartenir au groupe-intervention. Nous reproduisons cette situation dans le Tableau 6, où Y est une variable continue quelconque et dont les cellules contiennent la distribution des cas et la moyenne de Y pour chaque sous-groupe.

Tableau 6 : Élaboration de tableau dont la variable-test est le sexe

	Sexe			
	Femme=1		Homme=0	
	D		D	
	1	0	1	0
Y	42%	18%	20%	20%
	4	2,5	3	2

Nous savons que sept femmes sur dix se retrouveront dans le groupe-intervention et qu'un homme sur deux s'y retrouvera aussi. Afin d'obtenir une estimation sans biais de l'effet causal, il faut ajuster le calcul afin de tenir compte de cette assignation aléatoire stratifiée. Nous ajustons ce calcul en pondérant par les probabilités marginales de S ³, lesquelles proviennent des probabilités conjointes de D et S . Le calcul se fait ainsi pour la probabilité conjointe de la première cellule en postulant que les femmes constituent soixante pour cent de l'échantillon :

$$\Pr[D = 1, S = 1] = \Pr[D = 1 | S = 1] * \Pr[S = 1] = 0,7 * 0,6 = 0,42. \quad (22)$$

Nous poursuivons ainsi pour chaque cellule et obtenons les valeurs inscrites au Tableau 6. Le calcul marginal pour la strate $S = 1$ se fait ainsi :

$$\Pr[S = 1] = \Pr[D = 1, S = 1] + \Pr[D = 0, S = 1] = 0,42 + 0,18 = 0,60 \quad (23)$$

Un calcul semblable pour la strate $S = 0$ nous donne $\Pr[S = 0] = 0,40$. Finalement, nous pouvons estimer l'effet causal par la différence entre le groupe-intervention et le groupe-contrôle pour chaque strate de *sexe*, mais pondérer par leur distribution marginale de chaque strate, puis additionner ces résultats :

$$\begin{aligned} (\bar{y}_{D=1, S=1} - \bar{y}_{D=0, S=1}) \Pr[S = 1] + (\bar{y}_{D=1, S=0} - \bar{y}_{D=0, S=0}) \Pr[S = 0] = \\ (4 - 2),5)(,6) + (3 - 2)(,4) = 1,3 \end{aligned} \quad (24)$$

Ainsi, l'effet causal est de 1,3 dans cet exemple hypothétique et cette estimation est sans biais parce que nous avons ajusté les données de façon appropriée. Puisqu'il s'agit d'une étude expérimentale, nous avons le contrôle du mécanisme de sélection. Le mécanisme de sélection est une assignation aléatoire stratifiée sur une variable. Le sexe est la seule caractéristique des individus qui ait un rôle dans le mécanisme de sélection. Toutes les autres caractéristiques n'ont aucun rôle dans le mécanisme de sélection puisqu'elles sont

³ Techniquement, nous arrivons à la même réponse si nous utilisons les probabilités marginales de *sexe*, mais il est plus intuitif d'utiliser D puisque nous voulons comparer un groupe-intervention à un groupe-contrôle.

aléatoirement distribuées à l'intérieur des strates pour le sexe. Nous devons donc tenir compte du sexe lors du calcul de l'effet causal afin d'obtenir un résultat sans biais. Remarquons qu'un premier estimateur naïf est égal à 1,5. Il s'agit simplement de la différence des moyennes entre le groupe-intervention et le groupe-contrôle, sans tenir compte de la proportion de femmes et d'hommes dans chaque groupe :

$$\bar{y}_{D=1} - \bar{y}_{D=0} = 3,7 - 2,2 = 1,5 \quad (25)$$

Un deuxième estimateur naïf est égal à 2,5. Il s'agit de l'addition de la différence des moyennes entre le groupe-intervention et le groupe-contrôle à l'intérieur de chaque strate et ce, sans pondération :

$$\begin{aligned} (\bar{y}_{D=1,S=1} - \bar{y}_{D=0,S=1}) + (\bar{y}_{D=1,S=0} - \bar{y}_{D=0,S=0}) = \\ (4 - 2,5) + (3 - 2) = 2,5 \end{aligned} \quad (26)$$

Dans les deux cas, l'estimation est biaisée positivement.

Bien que nous ayons créé une homogénéité entre le groupe-intervention et le groupe-contrôle à l'intérieur de chaque strate, nous avons par le fait même créé une hétérogénéité entre les strates. Ainsi, nous avons regroupé les femmes et les hommes dans deux groupes distincts, ce qui fait en sorte que l'homogénéité intragroupe est très forte et l'homogénéité intergroupe est très faible. Autrement dit, le groupe-intervention et le groupe-contrôle sont comparables à l'intérieur de chaque strate, mais les strates demeurent incomparables si nous n'ajustons ces strates selon leur poids respectif dans l'échantillon. C'est cette hétérogénéité que le score de propension permet d'ajuster (Morgan et Winship, 2007, p.99). Rappelons-nous la formule (22). Celle-ci comporte le terme $\Pr[D=1|S=1]$ qui n'est autre que le score de propension pour les femmes, 0,7. Seulement, pour la majorité des études observationnelles, le score de propension n'est pas connu. Celui-ci doit être estimé. Par exemple, un chercheur ne connaissant pas le score de propension des femmes ayant servi à construire le Tableau 6, pourraient s'y prendre de cette façon en utilisant les fréquences ou les proportions des cellules du tableau :

$$\Pr[D = 1 | S = 1] = \frac{\Pr[D = 1, S = 1]}{\Pr[S = 1]} = \frac{0,42}{0,6} = 0,7 \quad (27)$$

Cet exemple illustre bien le procédé, mais il est difficile d'application. Si le chercheur désire inclure une multitude de variables-tests ou de variables d'appariement, même si celles-ci sont toutes dichotomiques, le calcul devient difficile et le problème de multidimensionnalité refait surface. Certaines cellules peuvent se retrouver seulement avec des cas-interventions, seulement avec des cas-contrôles ou sans aucun cas. Plus important, s'il désire utiliser des variables continues, le calcul est impossible avec la méthode des tableaux. C'est ici qu'intervient le score de propension. Depuis l'article fondateur de l'analyse par le score de propension (Rosenbaum et Rubin, 1983), la technique la plus largement répandue pour estimer le score de propension est la régression logistique. La régression logistique permet d'inclure autant de variables qu'un chercheur le désire, que celles-ci soient catégorielles ou continues. Les probabilités prédites de la régression logistique sont assimilées au score de propension évalué pour chaque cas. Une fois les scores de propension obtenus pour chaque cas, ces scores servent de mesures de distance. Ainsi, nous pouvons appairer les cas sur la base de ces distances. En général, tout comme pour la distance de Mahalanobis, la méthode d'appariement la plus souvent utilisée est celle de l'appariement prédateur (greedy matching) sans remplacement, ce qui a pour conséquence que les défauts de l'appariement prédateur, notamment la sous-optimisation de la distance moyenne entre les cas, sont encore présents.

La dernière mesure de distance que nous examinons se veut une généralisation de la distance de Mahalanobis et du score de propension. Elle permet entre autres de se passer de la propriété EPBR afin de produire des appariements efficaces. De plus, l'appariement est indissociable du calcul de la distance, contrairement aux deux autres techniques vues précédemment. La distance de Mahalanobis et la distance par le score de propension peuvent être associées à une multitude de techniques d'appariement, bien que l'appariement prédateur soit de loin la technique la plus utilisée. Dans le cas de l'appariement génétique, l'appariement est fait de façon telle que la mesure de distance soit optimale. Le prix à payer

pour ce gain d'efficacité est seulement du temps de calcul. Un appariement génétique peut prendre des dizaines de fois plus de temps qu'un appariement par la distance de Mahalanobis ou celui par la distance du score de propension, selon le degré d'efficience désiré par le chercheur, le nombre de variables en jeu ainsi que le nombre d'observations.

4.4 L'appariement génétique

Nous avons vu que l'appariement par la distance de Mahalanobis et l'appariement par le score de propension étaient optimaux lorsque les variables d'appariement étaient distribuées de façon ellipsoïdale. Lorsque les variables sont distribuées de la sorte, les deux méthodes acquièrent la propriété d'être EPBR. Nous aimerions disposer d'une méthode qui est peu ou pas sensible à la distribution des variables d'appariement. C'est ce que propose Sekhon et Diamond (Diamond et Sekhon, 2005). Il nomme cet appariement, l'appariement génétique.

Les deux mesures de distance que sont la distance de Mahalanobis et la distance assimilée au score de propension sont des mesures de distance particulière. Il existe d'innombrables mesures de distance. Comment savoir quelle est la mesure de distance appropriée pour un ensemble de données particulier ? La plupart des ensembles de données ne disposent pas de variables d'appariement correctement distribuées pour que les deux méthodes d'appariement vues jusqu'à maintenant soient efficaces. Sekhon et Diamond ont donc proposé une mesure de distance générale. Cette mesure est une généralisation de la distance de Mahalanobis et du score de propension.

La formule de la distance utilisée par l'appariement génétique s'écrit de cette façon :

$$d(X_i, X_j) = \sqrt{(X_i - X_j)^T (S^{-1/2})^T W S^{-1/2} (X_i - X_j)} \quad (28)$$

La différence fondamentale entre cette mesure et la mesure de Mahalanobis est la présence de la matrice W . Il s'agit d'une matrice de pondération dont toutes les cellules sont égales zéro sauf les cellules de la diagonale. Cette matrice est la seule inconnue de

cette formule puisque la matrice X contient les variables d'appariement et la matrice S les covariances associées à ces variables. C'est donc cette matrice W que cherchera à estimer l'algorithme génétique. Il cherche un poids à donner à chaque variable de façon telle qu'on optimise une distribution cumulative d'un test statistique. Si les conditions d'application de l'appariement par la distance de Mahalanobis sont réunies, alors l'algorithme génétique estimera une matrice W dont la diagonale a des valeurs près de un, ce qui revient à une matrice identité et nous retrouvons la formule (19) de la distance de Mahalanobis. Si les conditions d'application de l'appariement par le score de propension sont réunies, alors l'appariement génétique retrouvera des mesures de distance proche de celles qu'obtiendraient d'autres techniques telle que la régression logistique.

Examinons maintenant les résultats de deux simulations menées par Sekhon afin de comparer les performances des trois méthodes d'appariement examinées dans ce mémoire.

4.5 Une évaluation des méthodes d'appariement par simulation

Afin de bien fixer les idées à propos des considérations théoriques précédentes, nous rendons compte ici des résultats de deux simulations de Monte-Carlo exécutées par Sekhon et reproduits dans un rapport de travail (Sekhon, à paraître, p.7-10).

Nous reproduisons une partie des résultats des simulations dans le tableau suivant :

Tableau 7 : Comparaison des biais obtenus par différentes méthodes d'appariement

	<i>Biais de la Simulation 1</i>	<i>Biais de la Simulation 2</i>
<i>Estimateurs</i>	Distribution normale multivariée	Distribution des données de Lalonde
Génétique	-2,47	25,6
Score de propension	-2,45	512
Mahalanobis	-8,63	-717
Sans appariement	-60,4	485

La première simulation est conçue de façon telle que les conditions d'application de l'appariement par la distance de Mahalanobis et de l'appariement par le score de propension sont respectées. Ces conditions sont détaillées dans deux articles de Rubin et Thomas (Rubin, D. B. et Neal Thomas, 1992; Rubin, Donald B. et Neal Thomas, 1992). Essentiellement, les variables d'appariement sont normales multivariées, la relation entre ces variables et la variable d'intérêt est linéaire et il n'existe pas de variables inobservées influentes. La propriété EPBR est donc présente, ce qui devrait favoriser l'appariement par la distance de Mahalanobis et l'appariement par le score de propension. La vraie valeur de l'effet causal pour cette simulation est zéro. Nous constatons que l'estimation de l'effet causal sans appariement produit un biais assez large (-60,4). L'appariement génétique et l'appariement par le score de propension produisent le plus petit biais, marginalement mieux dans le cas du score de propension (-2,47 et -2,45, respectivement).

La deuxième simulation est conçue de façon à reproduire le plus fidèlement possible les caractéristiques des variables d'une étude réelle. Ces variables se trouvent être un sous-ensemble des variables de l'ensemble de données de Lalonde (LaLonde, 1986) utilisées par Dehejia and Wahba (Dehejia et Wahba, 1998). Les variables ne sont pas distribuées de

façon idéale, c'est-à-dire qu'elles n'ont pas une distribution ellipsoïdale. La relation entre ces variables et la variable d'intérêt n'est pas linéaire et certaines de ces variables sont nominales. Cependant, il n'existe pas de variables inobservées influentes pour cette simulation comme pour la première. Sous ces conditions, la propriété EPBR ne tient pas. La vraie valeur de l'effet causal pour cette simulation est mille. Plusieurs constats ressortent des résultats de cette simulation. L'appariement génétique obtient le plus petit biais (25,6). Le résultat le plus dérangeant cependant est celui de constater que l'appariement par la distance de Mahalanobis et l'appariement par le score de propension aggravent le biais par rapport à l'estimation de l'effet causal sans appariement (-717, 512 et 485, respectivement).

Ainsi, il apparaît clairement que tant les considérations théoriques que les simulations incitent à l'utilisation de l'appariement génétique. Ajoutons que l'appariement génétique repose sur un algorithme qui appartient à une famille d'algorithmes du nom d'algorithmes d'apprentissage automatique (*machine learning algorithms*). Dans un autre document de travail (Diamond et Sekhon, 2005), Diamond et Sekhon comparent l'algorithme génétique à deux autres algorithmes d'apprentissage automatique, les forêts d'arbres décisionnels (*random decision forest*) et les arbres de régression (*boosted regression trees*) et ce sous toutes sortes de contraintes (sept scénarios en tout). Dans tous les cas, sans exception, l'appariement génétique produit les plus petits biais de l'estimation de l'effet causal.

5 Application de l'analyse causale

Ce chapitre appliquera les notions de l'analyse causale sur des données obtenues dans le cadre d'une étude du GRIP (Groupe de recherche sur l'inadaptation psychosociale chez l'enfant), l'Étude longitudinale et expérimentale de Montréal (ÉLEM). Après un bref aperçu de l'ÉLEM, nous présenterons la composante expérimentale de l'ÉLEM, de laquelle sont tirées les données de notre analyse. Nous poursuivrons par une discussion sur les considérations théoriques examinées dans la première partie de ce travail, mais appliquées à nos données. Finalement, nous commenterons les résultats de nos analyses, lesquelles sont divisées en deux parties. Dans un premier temps nous examinerons les résultats de l'équilibrage des variables, puis nous évaluerons les résultats de l'effet de traitement.

5.1 L'Étude longitudinale et expérimentale de Montréal (ÉLEM)

Les participants à l'ÉLEM sont au nombre de 1 037, tous des garçons d'âge préscolaire, recrutés dans 53 écoles primaires de quartiers défavorisés de Montréal au Canada. Ces garçons furent évalués par les professeurs à l'aide du *Questionnaire d'évaluation du comportement au préscolaire* (Tremblay, 2004, p.2). Ce questionnaire collecte de l'information sur les caractéristiques des enfants. Le questionnaire est composé de 38 items conçus spécialement pour des enfants de cet âge. De ces 38 items, 32 formèrent quatre composantes orthogonales et furent regroupés en quatre échelles : comportement disruptif (treize items), anxiété (cinq items), inattention (quatre items) et comportement prosocial (dix items) (Tremblay, Loeber, Gagnon, Charlebois, Larivée et LeBlanc, 1991). Afin de créer un échantillon le plus homogène possible et ainsi réduire un biais culturel potentiel, seuls des enfants de parents nés au Canada et dont la langue maternelle est le français furent recrutés (Tremblay, 2004, p.2).

5.1.1 Présentation de la composante expérimentale

5.1.1.1 Sélection des cas

Sur l'ensemble des 1 037 enfants de l'étude longitudinale, 250 jugés disruptifs furent sélectionnés pour participer à la composante expérimentale (Tremblay, 2004). La sélection fût effectuée à l'aide de l'échelle du comportement disruptif (13 items) du *Questionnaire d'évaluation du comportement au préscolaire* : intimide; donne des coups de pied, mord, frappe; se bat; désobéit; blâme les autres; irritable; détruit des objets; agité; manque d'égard envers autrui; ment; ne tient pas en place; refuse de partager; peu apprécié de ses pairs (Tremblay, 2004, p.2). Sur cette échelle, les garçons qui se situaient au-delà du 70^e percentile furent sélectionnés. Ces garçons furent assignés de façon aléatoire à trois groupes : un groupe-intervention (n=69), un groupe-contrôle (n=58) et un groupe-témoin (n=123)⁴. Le groupe-contrôle et le groupe-témoin furent regroupés puisqu'aucune caractéristique n'était significativement différente entre ces deux groupes (Boisjoli, Vitaro, Lacourse, Barker et Tremblay, 2007; Lacourse, Coté, Nagin, Vitaro, Brendgen et Tremblay, 2002; Tremblay, 2004). C'est de ce nouveau groupe-contrôle (n=181) que les cas-contrôles potentiels seront échantillonnés par les différentes méthodes d'appariement.

Le variable d'intérêt est une mesure d'auto-évaluation de la violence qui représente la fréquence de sept comportements sur une période de 12 mois. Une échelle de quatre points (jamais, une ou deux fois, souvent et très souvent) a été utilisée pour déterminer le niveau de comportement violent : menacer d'attaquer une personne, se battre à coup de poing, attaquer une personne innocente, participer à une bataille au sein d'un groupe, lancer des objets vers d'autres personnes, porter une arme et faire usage d'une arme lors d'une bataille. La somme des scores est utilisée comme échelle de violence (coefficient alpha de Cronbach = 0,77) (Lacourse *et coll.*, 2002). Le logarithme naturel a été utilisé en raison de la distribution asymétrique de cette variable. Le Tableau 8 montre les statistiques

⁴ Le groupe-témoin sert à vérifier les effets du suivi sur le groupe-contrôle. Le groupe-contrôle n'a pas reçu de traitement, mais il a été suivi pour la collecte de données.

descriptives pour la variable d'intérêt (violence) et les variables d'appariement. Ces dernières variables, mesurées à l'âge de six ans, ont très peu de valeurs manquantes, ce qui donne un total de 2240 valeurs sur une possibilité de 2250 (250 garçons et 9 variables). Ceci représente 99,6% des valeurs possibles. Par contre, les sept mesures de la violence entre 11 et 17 ans ont une forte proportion de valeurs manquantes, ce qui fait un total de 1379 valeurs sur une possibilité de 1750 (250 garçons pendant sept vagues), s'il n'y avait pas eu un problème de suivi ou de désistement de la part des participants. Ceci représente 78,8% des valeurs obtenues à la fin de la collecte (Paquin, à venir).

À première vue, les variables les plus dissemblables entre les deux groupes sont les trois premières variables du Tableau 8 : *Dossier criminel – Mère*, *Dossier criminel – Père* et *Anxiété* (grisées dans le tableau). Nous devons porter une attention particulière à ces variables lors de l'examen des résultats des appariements.

Tableau 8 : Statistiques descriptives par groupe: moyenne, écart-type, N

Variables	Groupes						Total		
	Contrôle			Intervention					
	Moyenne	Écart-type	N	Moyenne	Écart-type	N	Moyenne	Écart-type	N
<i>Variables d'appariement (6 ans)</i>									
Dossier criminel - Mère	0,08	(0,27)	180	0,03	(0,17)	68	0,06	(0,25)	248
Dossier criminel - Père	0,08	(0,28)	180	0,03	(0,17)	68	0,07	(0,25)	248
Anxiété	1,75	(1,73)	181	2,30	(1,73)	69	1,90	(1,74)	250
Agression physique	3,52	(1,59)	181	3,48	(1,54)	69	3,51	(1,58)	250
Opposition	5,62	(2,19)	181	5,81	(1,93)	69	5,68	(2,12)	250
Hyperactivité	2,79	(1,21)	180	2,96	(1,19)	68	2,84	(1,20)	248
Inattention	2,18	(1,33)	179	2,26	(1,38)	68	2,20	(1,34)	247
Pro-social	6,52	(4,79)	181	6,99	(4,51)	69	6,65	(4,71)	250
Adversité familiale	0,43	(0,24)	181	0,43	(0,27)	68	0,43	(0,25)	249
<i>Violence (âge)</i>									
11 ans	2,23	(0,25)	160	2,21	(0,28)	62	2,22	(0,26)	222
12 ans	2,22	(0,27)	154	2,19	(0,24)	62	2,21	(0,26)	216
13 ans	2,21	(0,26)	149	2,13	(0,22)	56	2,19	(0,25)	205
14 ans	2,26	(0,29)	138	2,21	(0,34)	55	2,24	(0,31)	193
15 ans	2,26	(0,32)	136	2,19	(0,31)	53	2,24	(0,32)	189
16 ans	2,23	(0,32)	130	2,12	(0,25)	50	2,20	(0,31)	180
17 ans	2,15	(0,25)	124	2,12	(0,24)	50	2,14	(0,25)	174

Note : Dossier criminel, père et mère, sont des variables dichotomiques codées 0 et 1. La moyenne est donc une proportion.

Les variables indépendantes sont la possession d'un dossier criminel pour le père et pour la mère, l'anxiété, l'agression physique, l'opposition, l'hyperactivité, l'inattention, le comportement prosocial et l'adversité familiale. L'anxiété a été mesurée en utilisant cinq

items : craintif, avoir peur de nouvelles situations; misérable, ne semble pas être heureux; être inquiet; solitaire, tendance à faire les choses seul; être en pleurs facilement (Alpha de Cronbach=0,76). L'agression physique a été mesurée en utilisant trois items : se bat avec les autres enfants, donne des coups de pied, mord ou frappe les autres enfants et intimide d'autres enfants (Alpha de Cronbach=0,84). L'opposition a été mesurée à partir de cinq items : ne partage pas son matériel, est irritable, est désobéissant, blâme les autres et est peu attentif aux autres (Alpha de Cronbach=0,84). L'hyperactivité a été mesurée à partir de deux items: remue continuellement et est très agité (Alpha de Cronbach=0,87). L'inattention a été mesurée à partir de deux items : est facilement distrait et a une faible concentration (Alpha de Cronbach=0,74). Le comportement prosocial a été mesuré à partir de dix items: essaie d'arrêter les querelles, invite les autres à se joindre, essaie d'aider quelqu'un qui est blessé, aide à ramasser les objets échappés par quelqu'un d'autre, encourage le travail des enfants moins capables, montre de la sympathie envers quelqu'un qui a fait une erreur, aide les enfants qui ont de la difficulté avec une tâche, aide les enfants qui sont malades, reconforte les enfants en pleurs ou fâchés et aide nettoyer le désordre causé par quelqu'un d'autre (Alpha de Cronbach=0,92). L'indice d'adversité familiale utilise six variables dichotomiques: âge de la mère à la première naissance, le niveau d'éducation de la mère et du père, le niveau socioprofessionnel du père et de la mère et la structure familiale (intacte ou non). Les garçons se voyaient attribuer la valeur de zéro sur les trois premiers critères s'appliquant aux parents si ceux-ci se trouvaient au-dessus du 30e percentile et de un s'ils étaient au-dessous. Le rang percentile a été calculé à partir de l'échantillon complet (N=1161). Pour la structure familiale, une valeur de zéro était attribuée aux garçons qui vivaient avec leurs deux parents biologiques et de un pour tous les autres types de structure familiale. Puisque le maximum de l'échelle diffère selon qu'un garçon est dans une famille avec deux parents ou monoparentale, le score total obtenu a été divisé par sept pour les garçons ayant deux parents et par quatre pour ceux d'une famille monoparentale. Ainsi, une valeur élevée représente un haut niveau d'adversité familiale (Nagin et Tremblay, 2001; Paquin, 2010; Tremblay *et coll.*, 1991; Vitaro, Brendgen et Tremblay, 2001).

5.1.1.2 Le programme d'intervention auprès des enfants et des parents

Conçu pour réduire les comportements violents chez les garçons, le programme de prévention a été mené entre 1985 et 1987 alors que les garçons étaient âgés de sept à neuf ans. L'intervention était destinée non seulement aux garçons mais aussi à leurs parents. Les enfants ont fait l'objet d'un apprentissage de compétences sociales à l'école et les parents d'un apprentissage de parents efficaces (Barker, Vitaro, Lacourse, Fontaine, Carbonneau et Tremblay, 2010).

Les enfants assignés au groupe expérimental ont pris part à des séances de perfectionnement des compétences sociales qui ont eu lieu en groupes de quatre à sept participants à l'école. Un ratio de trois enfants prosociaux pour un enfant perturbateur a été maintenu. La présence des enfants prosociaux avait pour but de créer un effet positif sur les garçons jugés à risque. Un total de 19 séances a eu lieu et en moyenne les garçons se sont présentés à 14,5 de ces sessions. Les sessions furent dirigées par quatre professionnels, dont deux éducateurs, un travailleur social et un psychologue. Des instructions verbales, du renforcement positif, la présentation de modèles à suivre et la répétition de certains comportements furent utilisées pour l'enseignement des compétences sociales recherchées (Barker *et coll.*, 2010).

La composante parentale visait à développer des pratiques parentales efficaces selon le modèle de *l'Oregon Social Learning Center* (Patterson, Reid, & Jones, 1975). Les mêmes professionnels qui dirigeaient les séances pour les garçons dirigèrent aussi la formation des parents à la maison. Dans un premier temps, on apprenait aux parents à reconnaître, à observer et à noter les comportements problématiques de leurs enfants. Ensuite, on leur apprenait à définir ce qu'étaient des comportements appropriés et à fixer des objectifs clairs à atteindre pour leurs enfants. En troisième lieu, ceux-ci apprenaient à utiliser le renforcement verbal et physique dans des circonstances appropriées et de façon systématique afin de favoriser l'acquisition de comportements appropriés par leurs enfants. Les parents apprenaient aussi à punir les comportements inappropriés de façon systématique et modérée. Les parents furent encouragés à superviser les travaux d'école de

leurs enfants et à surveiller leur comportement à l'extérieur de la maison. Finalement, les parents apprenaient à gérer les crises familiales par la résolution de problèmes et aussi à utiliser la négociation dans les situations de la vie quotidienne (Barker *et coll.*, 2010).

5.2 Effet causal d'une intervention psychosociale sur le comportement agressif

5.2.1 Le modèle causal contrefactuel

Si nous passons en revue les postulats du modèle causal contrefactuel, nous avons de bonnes raisons de croire que les cinq premiers postulats sont rencontrés par nos données, en grande partie parce que celles-ci proviennent d'une étude expérimentale avec assignation aléatoire. L'existence d'états causaux nettement délimités, stipulée par le Postulat 1, est assurée par la formation d'un groupe-intervention et d'un groupe-contrôle. En ce qui concerne le Postulat 3, sa première composante stipule qu'il n'y a aucune interaction entre les deux groupes et sa deuxième composante stipule qu'il n'y a pas de variation dans l'administration du traitement ni dans la façon de contrôler le groupe-témoin. Nous n'avons aucune raison de croire que les deux groupes de notre étude aient pu interagir fortement ou que le protocole d'intervention et de contrôle ne fût pas suivi rigoureusement. En général, il est assez facile, pour une étude comme l'ÉLEM de s'assurer que ces deux contraintes soient respectées. Le Postulat 2 est automatiquement rencontré si le Postulat 1 l'est (il s'agit d'un corollaire en fait). En effet, si tous les individus ont une probabilité déterminée d'appartenir soit au groupe-intervention soit au groupe-contrôle, peu importe le mécanisme d'assignation à l'origine de cette affectation, le Postulat 2 est rencontré, c'est-à-dire que chaque individu est considéré posséder une réponse potentielle bien définie s'il se trouve dans le groupe-intervention et une réponse potentielle bien définie s'il se trouve dans le groupe-contrôle.

Le Postulat 4 stipule qu'en moyenne les deux groupes réagissent de la même façon au traitement ainsi qu'au contrôle. En principe, ce postulat est rencontré puisque les deux

groupes proviennent d'une assignation aléatoire. C'est ici que les problèmes surgissent. Puisque certaines variables ne sont pas équilibrées, puisque les deux groupes ne sont pas équivalents sur ces variables, nous ne pouvons être certains que les deux groupes réagiraient de la même façon au traitement et au contrôle. Le Postulat 4 nécessite l'adjonction du Postulat 5. Ce dernier postulat stipule que les réponses potentielles sont indépendantes du statut du traitement (intervention ou contrôle). L'assignation aléatoire assure en général que ce dernier postulat est rencontré. Mais puisque l'assignation aléatoire semble défectueuse, nous devons faire appel à deux autres postulats, comme si notre étude était une étude sans assignation aléatoire (Guo et Fraser, 2010, p.113, 206).

Nous pouvons alors soutenir le Postulat 6 qui stipule que les réponses potentielles sont indépendantes du statut du traitement (intervention ou contrôle) si nous tenons compte du mécanisme d'assignation. De même, si nous tenons compte du mécanisme d'assignation, nous pouvons alors soutenir le Postulat 7, lequel stipule qu'en moyenne les deux groupes réagissent de la même façon au traitement ainsi qu'au contrôle. Finalement, nous pouvons avancer la Définition 4 d'un effet conditionnel du traitement, c'est-à-dire de l'effet d'un traitement compte tenu du mécanisme d'assignation. Ainsi, l'effet de traitement est identifié, dans le langage des statisticiens. Comme nous le verrons à la prochaine section, le mécanisme d'assignation retenu est assez simple et sera composé des neuf variables d'appariement présentées plus haut.

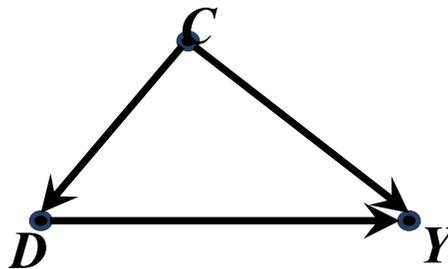
5.2.2 L'appariement

5.2.2.1 La sélection des variables d'appariement

Les neuf variables d'appariement retenues nous ont été proposées en partie pour permettre des comparaisons avec une étude antérieure sur l'ajustement des données manquantes de la variable d'intérêt (Paquin, 2010). Toutes les variables se retrouvent dans un seul ensemble et sont considérées de même nature causative. Nous représentons à la Figure 1 un graphe causal qui représente une telle situation. Ces variables sont dites des variables de confusion. Dans le cadre d'une analyse causale, une variable de confusion C

est considérée comme une cause commune à la variable d'intérêt Y et à la cause à l'étude D . Autrement dit, on considère que l'ensemble C affecte simultanément l'assignation des cas et le résultat de l'intervention sur la caractéristique à l'étude (Hernan et Robins, en préparation, p.83-93; Morgan et Winship, 2007, p.67-70).

Figure 1 : Graphe causal avec une variable de confusion



Notons finalement qu'en principe nos données ne comportent pas de variables de confusion puisque ces données proviennent d'une assignation aléatoire. Avec ce type d'étude, il suffit de calculer la différence des moyennes entre les deux groupes pour obtenir l'effet causal. Bien entendu, si le plan expérimental est complexe, on doit faire appel à l'une ou l'autre des techniques d'ANOVA ou de régression afin d'évaluer l'effet causal. Il est courant malgré tout, lors d'études avec assignation aléatoire simple, de faire appel à l'une de ces techniques afin de réduire la variabilité due à l'assignation aléatoire, notamment si l'échantillon est de taille réduite. Cette variabilité peut être assez sévère pour considérer que l'assignation aléatoire est problématique. Dans ce cas, on peut toujours relancer l'assignation aléatoire jusqu'à ce que nous obtenions deux groupes assez semblables sur l'ensemble des variables disponibles avant tout traitement (*baseline variables*). Si le déséquilibre est constaté après la collecte de données qui a suivi le traitement, alors il est courant d'avoir recours aux techniques d'ajustement pour données observationnelles afin d'équilibrer les variables. Nous verrons au Tableau 10, ci-dessous, que selon le critère retenu pour évaluer l'équilibre des variables, cinq des neuf variables à notre disposition sont déséquilibrées avant l'application d'une ou l'autre des techniques d'appariement.

5.2.2.2 Les valeurs manquantes

Les valeurs manquantes sur les variables d'appariement doivent être imputées puisque tous les algorithmes d'appariement utilisés exigent qu'aucune valeur manquante ne soit présente sur les variables servant à l'appariement. Puisqu'il y a très peu de valeurs manquantes pour ces variables (V. Tableau 8), pratiquement n'importe quelle technique d'imputation fera l'affaire. Nous avons opté pour une imputation simple par la régression à l'aide du module MVA de PASW Statistics 18 (SPSS Inc., 2009).

5.2.2.3 Spécification des appariements

Le module MatchIt sous R fût utilisé pour les appariements (Imai *et coll.*, 2011). Pour les trois méthodes d'appariement exposées dans ce travail, nous avons choisi quatre schèmes d'appariement : 1:1 sans remise, 1:1 avec remise, 2:1 sans remise et 2:1 avec remise. Ces schèmes sont d'utilisation courante dans la littérature. Le premier chiffre de chaque schème indique le nombre de cas-contrôles appariés à un cas-intervention. Un schème sans remise indique qu'un cas-contrôle ne peut être apparié qu'à un seul cas-intervention alors qu'un schème avec remise indique qu'un cas-contrôle peut être apparié à plusieurs cas-interventions. Autrement dit, avec un schème sans remise le cas-contrôle est retiré de l'ensemble (avec le cas-intervention auquel il est apparié), alors qu'avec un schème avec remise le cas-contrôle est retourné dans l'ensemble des cas-contrôles (le cas-intervention est retiré) afin qu'il soit disponible à être apparié à un autre cas-intervention. Le module Matchit tient compte de ces différents schèmes lors de ses calculs en produisant notamment une variable de pondération qui sera utilisée lors de l'analyse causale. Nous nous retrouvons donc avec douze ensembles appariés, un ensemble par méthode d'appariement et schème d'appariement. Le Tableau 10 présente ces douze schèmes d'appariement ainsi que les résultats du diagnostic sur l'équilibre des variables.

5.2.2.4 Diagnostic de l'appariement

Le diagnostic des appariements est l'étape la plus importante de l'analyse causale. Même si toutes les conditions d'une analyse causale semblent avoir été remplies, si les

algorithmes d'appariement ne réussissent pas à équilibrer les variables entre les deux groupes, alors l'analyse causale est impossible. Ce diagnostic sert à choisir l'ensemble de données appariées qui sera utilisé à l'étape de l'analyse causale proprement dite. Autrement dit, il faut choisir lequel des ensembles de données produits par les appariements servira à l'évaluation de l'effet causal.

Afin d'évaluer les résultats des appariements, nous devons disposer de critères. Ces critères permettent de constater si les deux groupes sont équilibrés sur les variables ayant servi aux appariements. Qu'entendons-nous par "équilibré" (*balance* en anglais) ? Deux groupes sont équilibrés si les distributions statistiques de chacune des variables les concernant, que celles-ci aient été mesurées ou pas, sont identiques pour les deux groupes. Autrement dit, toutes les statistiques descriptives devraient être identiques en moyenne entre les deux groupes pour une même variable. Seule l'assignation aléatoire permet un équilibre aussi complet. Dans le cas de données observationnelles, nous devons nous contenter d'évaluer les variables à notre disposition selon des critères partiels, c'est-à-dire pour un petit nombre de statistiques descriptives. La plus utilisée de ces statistiques est la différence des moyennes standardisées. Un chercheur pourrait aussi examiner le minimum, le maximum, l'écart-type, la médiane ainsi qu'une multitude d'autres mesures descriptives.

Remarquons que plusieurs auteurs font appel à un test statistique afin de vérifier si les variables sont équilibrées entre deux groupes. Cependant, il a été démontré qu'un test statistique tel que le test t peut très bien indiquer que deux groupes sont semblables pour chacune des variables observées alors que dans les faits ces deux groupes diffèrent pour certaines mesures descriptives (Ho et coll., 2007, p.220-23; Imai, King et Stuart, 2008, p.494-99). Ces auteurs nous rappellent que l'équilibre des variables est une notion qui s'applique à un échantillon et non à une population. Autrement dit, il n'est pas question d'inférence statistique lorsqu'il s'agit de s'assurer que deux groupes d'un même échantillon sont semblables en moyenne sur l'ensemble des variables mesurées. Ainsi, même si un test t nous indique que deux groupes ne sont pas différents, ceci vaut en fait seulement pour ces deux groupes au niveau de la population de laquelle ils ont été tirés. Pour la population,

l'échantillon nous permet de conclure que les deux groupes ne diffèrent pas malgré que nous constatons que les deux groupes de l'échantillon diffèrent pour certaines variables sur certaines mesures descriptives. Par exemple, la moyenne de la variable z est de 31 pour le groupe A et de 32 pour le groupe B. Le test t nous permet d'inférer qu'il n'y a pas de différence entre les deux groupes au niveau de la population. Mais ce n'est pas la question qui nous intéresse. La question qui nous intéresse est du même ordre que celle qu'on se pose dans un protocole expérimental. Est-ce qu'une variable x (traitement, intervention) cause une réaction sur une variable d'intérêt ? Il n'est pas question ici d'inférence statistique à une population à partir d'un échantillon. En fait, très souvent, les groupes ne proviennent pas d'un échantillonnage aléatoire au sein d'une population. Ainsi, la différence de moyenne de la variable z ($32-31=1$) est une différence réelle entre les deux groupes et il est toujours avantageux de réduire les différences qui peuvent exister entre les deux groupes afin qu'ils soient les plus semblables possibles.

Pour ces raisons, la différence des moyennes standardisées entre les deux groupes est recommandée par les auteurs du package `MatchIt` sous R (Imai *et coll.*, 2011). Parfois, ce seul critère ne permet pas de départager différentes méthodes d'appariement. Ces auteurs proposent alors un deuxième critère complémentaire, soit l'écart des moyennes absolues standardisées en pourcentage entre l'ensemble de données avant l'appariement et chacun des ensembles de données produits par les méthodes d'appariement. Le premier critère indique lesquelles des variables ont été équilibrées et le deuxième critère indique si l'écart entre les deux groupes a diminué ou augmenté pour chacune des variables utilisées pour les appariements, que celles-ci soient équilibrées ou non avant et après l'appariement. Six situations sont alors possibles, représentées au Tableau 9.

Tableau 9 : Six cas de figure lors de l'équilibrage de variables

Situation No	Premier critère <i>Différence des moyennes standardisées</i>		Deuxième critère <i>Augmentation/Diminution de l'écart</i>
	Avant l'appariement	Après l'appariement	
1	Équilibrée	Équilibré	Diminution
2	Équilibrée	Équilibrée	Augmentation
3	Équilibrée	Déséquilibrée	Augmentation
4	Déséquilibrée	Déséquilibrée	Augmentation
5	Déséquilibrée	Déséquilibrée	Diminution
6	Déséquilibrée	Équilibrée	Diminution

Les deux premières colonnes indiquent un équilibre ou un déséquilibre selon le premier critère, soit la différence des moyennes standardisée, et la troisième colonne indique si ces différences ont subi une augmentation ou une diminution de l'écart suite aux appariements. La situation qui intéresse le plus le chercheur est la situation six. Une variable est déséquilibrée au départ puis devient équilibrée après l'appariement, ce qui a pour conséquence une diminution de l'écart entre les deux groupes sur cette variable. Les situations un et deux peuvent être considérées inoffensives. Dans les deux cas, la variable demeure équilibrée après l'appariement, mais peut subir soit une diminution soit une augmentation de l'écart entre les deux groupes. Les situations trois, quatre et cinq sont celles qu'on désire éviter. Pour ces situations, la variable devient soit déséquilibrée (situation trois) ou demeure déséquilibrée en subissant une augmentation du déséquilibre (situation quatre) ou une diminution du déséquilibre (situation cinq).

Le premier critère, la différence des moyennes standardisées, se calcule de la façon suivante:

$$\frac{\text{Moyenne du groupe-intervention} - \text{Moyenne du groupe-contrôle}}{\text{Écart-type du groupe-intervention}} \quad (29)$$

Le deuxième critère, relatif aux augmentations ou aux diminutions des moyennes standardisées, est réécrit sous forme de pourcentage selon la formule suivante :

$$\frac{\left| \text{Différence des moyennes standardisées brutes} \right| - \left| \text{Différence des moyennes standardisées appariées} \right|}{\left| \text{Différence des moyennes standardisées brutes} \right|} * 100 \quad (30)$$

Une valeur positive indique une diminution de l'écart entre les deux groupes et une valeur négative indique une augmentation de l'écart.

Maintenant, comment déterminons-nous si une variable est équilibrée ou non? Nous devons choisir une valeur pour la différence des moyennes standardisées qui servira à déterminer si la variable est équilibrée ou non entre les deux groupes. En ce qui concerne des données observationnelles, il est fréquent d'utiliser une différence de moyenne standardisée de 0,25 comme critère d'équilibrage. Une variable est considérée déséquilibrée si la différence de la moyenne standardisée de cette variable est égale ou supérieure à 0,25. Nous verrons que ce critère est trop permissif dans le cas de données provenant d'une assignation aléatoire. Dans le cas d'une assignation aléatoire, même si cette assignation est défectueuse, les deux groupes ne peuvent être tellement différents l'un de l'autre. Dans ce cas, on suggère d'utiliser un critère plus restrictif. Une variable est considérée déséquilibrée si la différence de la moyenne standardisée de cette variable est égale ou supérieure à 0,10 (Harder, Stuart et Anthony, 2010, p.238).

Muni de ces critères, nous pouvons maintenant évaluer les résultats des différentes méthodes d'appariement sous les différents schèmes retenus. Le Tableau 10 présente les résultats de l'équilibre des variables après les appariements. Il est divisé en quatre sous-tableaux, un tableau par schème d'appariement. Chaque sous-tableau est à son tour divisé en deux sous-tableaux, lesquels correspondent aux deux principaux critères retenus pour l'examen de l'équilibre des variables, soit la différence des moyennes standardisées (sous-tableau *Différences (moyennes standardisées)*) et l'écart des moyennes standardisées en pourcentage (sous-tableau *Réduction/Augmentation (%)*). Prenons la variable *Adversité* sous le schème d'appariement 1:1 sans remise. La colonne «*Brutes*» nous indique qu'au départ le groupe-intervention est constitué de 69 cas et le groupe-contrôle est constitué de

181 cas. Il s'agit des cas originaux avant d'appliquer une méthode d'appariement. En ce qui concerne la variable *Adversité*, nous constatons que la différence des moyennes standardisées entre les deux groupes est de 0,0318. Cette variable n'est donc pas déséquilibrée même selon notre critère plus restrictif de 0,10. Les trois colonnes suivantes nous indiquent que 69 cas-contrôles ont été appariés aux 69 cas-intervention puisque le schème d'appariement est un schème 1:1 sans remise. Sous ces trois colonnes, se trouvent les différences des moyennes standardisées obtenues une fois l'appariement effectué à l'aide des trois méthodes d'appariement. Nous constatons que cette différence s'est réduite (nous nous rapprochons de zéro en valeur absolue) dans le cas de l'appariement par la distance de Mahalanobis et dans le cas de l'appariement génétique (-0,0225 et -0,0219), mais que cette différence a augmenté (nous nous éloignons de zéro) dans le cas de l'appariement par le score de propension (-0,0818). Remarquons que ce dernier cas est représenté par la situation deux au Tableau 9, c'est-à-dire que malgré l'augmentation de l'écart, la variable demeure équilibrée après l'appariement par le score de propension.

Le sous-tableau *Réduction/Augmentation (%)* nous indique si les appariements ont permis de diminuer la différence des moyennes standardisées. Ainsi, pour la variable *Adversité*, nous enregistrons une diminution de 29,2% et de 31,2% dans le cas de l'appariement par la distance de Mahalanobis et de l'appariement génétique, tandis que l'appariement par le score de propension enregistre une augmentation de 157,2% (indiqué par le signe négatif). Ce dernier résultat ne peut être interprété seul, sans faire référence au premier critère retenu pour l'examen de l'équilibre des variables. L'analyse repose sur ce premier critère. Nous voyons que cette augmentation de 157,2% de l'écart ne fait pas en sorte que la variable est déséquilibrée après l'appariement puisque la différence des moyennes standardisées est de -0,0818, soit en deçà de notre critère de 0,10.

Revenons maintenant à la différence qui existe entre le critère de 0,25 et de 0,10. Selon le critère de 0,25, trois variables seraient déséquilibrées au départ (colonne «*Brutes*», cellules en gris pâle), *Anxiété*, *Doss. crim. – Père et Doss. crim.- Mère*, avec une différence standardisée respective de 0,3203, -0,3189 et -0,2862. Une fois les appariements effectués,

une seule valeur se trouve au-dessus de ce critère, soit la valeur de la variable *Anxiété* pour l'appariement par la distance de Mahalanobis, 0,3023 (colonne «Mahalanobis»). Et toujours selon cette valeur de 0,25, toutes les variables seraient équilibrées dans le cas de l'appariement par le score de propension et de l'appariement génétique. Ainsi, il serait pratiquement impossible de distinguer le mérite respectif de ces deux méthodes d'appariement si on se limitait à ce critère. Maintenant, selon ce critère de 0,10, nous constatons que cinq variables sont déséquilibrées avant l'appariement des cas (colonne «*Brutes*», cellules en gris pâle et en gris foncé; le nombre de variables déséquilibrées sont indiquées sur l'avant dernière ligne de chacun des sous-tableaux). Sur l'avant-dernière ligne du premier sous-tableau, nous constatons que seul l'appariement génétique a réussi à équilibrer toutes les variables (indiqué par un zéro) alors que la distance de Mahalanobis produit deux variables déséquilibrées et le score de propension en produit une. Nous distinguons maintenant plus facilement le mérite respectif des trois méthodes d'appariement.

Cependant, qu'arriverait-il si deux ou plusieurs méthodes d'appariement réussissaient à équilibrer toutes les variables (zéros sur l'avant-dernière ligne)? C'est à ce moment qu'on peut faire appel au deuxième critère, lequel dérive du premier comme nous l'avons vu ci-dessus à la formule (30). Examinons l'avant-dernière ligne du deuxième sous-tableau, *Réduction/Augmentation(%)*. Nous retrouvons ici le nombre de variables dont le déséquilibre a augmenté, représenté par les valeurs négatives dans les colonnes de ce sous-tableau (nous avons grisé les cellules qui représentent des situations d'augmentation de l'écart). Nous voyons que seul l'appariement génétique a diminué l'écart entre les deux groupes sur toutes les variables (zéro sur l'avant dernière ligne), alors que la distance de Mahalanobis a augmenté l'écart sur une variable (*Opposition*, -16,1) et le score de propension sur trois variables (*Adversité*, -157,2, *Bataille*, -24,4, *Inattention*, -148,9). Ces augmentations de l'écart ne sont pas nécessairement inattendues. Comme nous l'avons vu, ces résultats doivent être interprétés en regard de notre critère initial. Dans le cas du score de propension, pour les trois variables ayant subi une augmentation de l'écart, si nous examinons les différences des moyennes standardisées, nous constatons que seule

Inattention est déséquilibrée après l'appariement, -0,1130. Ainsi, bien que l'appariement par le score de propension ait augmenté l'écart de 157,2% sur *Adversité*, cet écart demeure en deçà de notre critère de 0,10, soit -0,0818. Ce cas est une illustration de la situation deux du Tableau 9, laquelle est considérée inoffensive. Mais bien que cette situation soit inoffensive, l'appariement génétique a fait mieux selon le critère retenu. Sur *Adversité*, l'appariement génétique a diminué l'écart de 31,2%. Si nous revenons à notre question de départ, même si deux méthodes d'appariement réussissent à équilibrer toutes les variables selon le critère de la différence des moyennes standardisées, il est toujours possible d'examiner en détail les écarts de ces différences. En général, nous devrions opter pour la méthode qui diminue l'écart sur le plus grand nombre de variables, représenté par les situations un, cinq et six du Tableau 9.

Tableau 10 : Analyse de l'équilibre des variables

		Différences (moyennes standardisées)				Réduction/Augmentation (%)		
		Brutes	Mahalanobis	Score	Génétique	Mahalanobis	Score	Génétique
Appariement 1:1 sans remise		N	69/181	69/69	69/69	69/69		
	Adversité	0,0318	-0,0225	-0,0818	-0,0219	29,16	-157,18	31,16
	Anxiété	0,3203	0,3023	0,0252	0,0168	5,65	92,14	94,76
	Prosocialité	0,1034	0,0032	-0,0482	-0,0450	96,89	53,37	56,48
	Bataille	-0,0303	0,0000	0,0376	-0,0094	100,00	-24,40	68,90
	Opposition	0,0972	0,1128	0,0150	-0,0075	-16,08	84,52	92,26
	Inattention	0,0454	-0,0401	-0,1130	-0,0193	11,64	-148,86	57,49
	Hyperactivité	0,1525	-0,0230	-0,0230	0,0014	84,92	84,92	99,10
	Doss. crim. - Père	-0,3189	0,0000	0,0000	0,0000	100,00	100,00	100,00
	Doss. crim. - Mère	-0,2862	0,0000	0,0000	0,0000	100,00	100,00	100,00
	Nbr var. déséq./ Nbr var. augment. déséq.	5	2	1	0	1	3	0
	Moy. de déséq./% moy. de réduct. déséq.	0,1540	0,0560	0,0382	0,0134	56,91	20,50	77,79
Appariement 1:1 avec remise		N	69/181	69/58	69/53	69/55		
	Adversité	0,0318	-0,0178	-0,2979	0,0053	43,93	-837,12	83,45
	Anxiété	0,3203	0,2771	0,0840	0,0084	13,51	73,79	97,38
	Prosocialité	0,1034	-0,0032	0,1093	-0,0129	96,89	-5,70	87,56
	Bataille	-0,0303	-0,0188	-0,0753	-0,0094	37,80	-148,80	68,90
	Opposition	0,0972	0,0978	-0,1128	0,0075	-0,60	-16,08	92,26
	Inattention	0,0454	-0,0089	-0,2380	0,0015	80,42	-423,99	96,65
	Hyperactivité	0,1525	-0,0596	-0,1570	-0,0108	60,94	-2,99	92,91
	Doss. crim. - Père	-0,3189	0,0000	0,0000	0,0000	100,00	100,00	100,00
	Doss. crim. - Mère	-0,2862	0,0000	0,0000	0,0000	100,00	100,00	100,00
	Nbr var. déséq./ Nbr var. augment. déséq.	5	1	5	0	1	6	0
	Moy. de déséq./% moy. de réduct. déséq.	0,1540	0,0537	0,1194	0,0062	59,21	-128,99	91,01
Appariement 2:1 sans remise		N	69/181	69/138	69/138	69/138		
	Adversité	0,0318	0,0231	0,0167	-0,0135	27,19	47,54	57,49
	Anxiété	0,3203	0,2519	0,1469	0,2099	21,37	54,13	34,48
	Prosocialité	0,1034	0,0064	0,0112	0,0370	93,78	89,12	64,25
	Bataille	-0,0303	0,0047	0,0612	-0,0094	84,45	-102,15	68,90
	Opposition	0,0972	0,0752	0,0639	0,0226	22,62	34,22	76,78
	Inattention	0,0454	-0,0643	-0,0453	-0,0297	-41,49	0,17	34,57
	Hyperactivité	0,1525	0,0867	0,0305	0,0501	43,17	80,01	67,14
	Doss. crim. - Père	-0,3189	0,0000	0,0000	0,0000	100,00	100,00	100,00
	Doss. crim. - Mère	-0,2862	0,0429	0,0000	0,0429	85,02	100,00	85,02
	Nbr var. déséq./ Nbr var. augment. déséq.	5	1	1	1	1	1	0
	Moy. de déséq./% moy. de réduct. déséq.	0,1540	0,0617	0,0417	0,0461	48,46	44,78	65,40
Appariement 2:1 avec remise		N	69/181	69/87	69/86	69/89		
	Adversité	0,0318	0,0119	-0,2080	-0,0023	62,50	-554,41	92,61
	Anxiété	0,3203	0,2351	0,0630	0,0210	26,61	80,34	93,45
	Prosocialité	0,1034	-0,0353	0,0787	-0,0048	65,80	23,83	95,34
	Bataille	-0,0303	-0,0518	-0,0188	-0,0141	-71,05	37,80	53,35
	Opposition	0,0972	0,0414	-0,1090	0,0188	57,44	-12,21	80,65
	Inattention	0,0454	-0,0974	-0,1443	-0,0297	-114,46	-217,64	34,57
	Hyperactivité	0,1525	-0,1022	-0,0244	-0,0108	32,97	84,03	92,91
	Doss. crim. - Père	-0,3189	0,0000	0,0000	0,0429	100,00	100,00	86,55
	Doss. crim. - Mère	-0,2862	0,0429	0,0429	0,0000	85,02	85,02	100,00
	Nbr var. déséq./ Nbr var. augment. déséq.	5	2	3	0	2	3	0
	Moy. de déséq./% moy. de réduct. déséq.	0,1540	0,0687	0,0766	0,0160	27,20	-41,47	81,05

Nous pouvons maintenant nous poser une question semblable à la précédente. Qu'arriverait-il si deux ou plusieurs méthodes d'appariement réussissaient non seulement à équilibrer toutes les variables (zéros sur l'avant-dernière ligne du premier sous-tableau), mais réussissaient aussi à diminuer l'écart sur toutes les variables (zéros sur l'avant-dernière ligne du sous-tableau de droite). Autrement dit, si les situations un et six du Tableau 9 se produisaient pour deux méthodes d'appariement, laquelle choisirait-on? La dernière ligne de chaque tableau fournit d'autres informations pour évaluer l'équilibre des variables. Cette dernière ligne du premier sous-tableau, *Différences (moyennes standardisées)*, indique la moyenne des valeurs absolues de chaque colonne. Ainsi, nous constatons que globalement, au départ la moyenne était de 0,1540 (colonne « Brutes »), alors qu'après les appariements nous obtenons 0,0560 pour la distance de Mahalanobis, 0,0382 pour le score de propension et 0,0134 pour l'appariement génétique. Pour le deuxième sous-tableau, la dernière ligne indique la moyenne directe (sans valeur absolue) de chaque colonne. Nous constatons que l'appariement génétique amène une diminution globale de 77,8%, ce qui constitue la plus forte diminution, alors que la distance de Mahalanobis permet une diminution de 56,9% et le score de propension de 20,5%. En général, nous devrions opter pour la méthode qui produit les plus petites moyennes standardisées globales et la plus grande diminution globale.

Les trois autres schèmes d'appariement peuvent être lus de la même façon. Nous constatons que l'appariement génétique surpasse les deux autres méthodes pour les quatre schèmes. Cette méthode d'appariement semble optimale avec un schème d'appariement 1:1 avec remise, du moins avec nos données (nous avons grisé les colonnes *Génétique* de ce deuxième tableau, Appariement 1:1 avec remise). Toutes les variables sont équilibrées selon notre critère (quatrième colonne, avant dernière ligne, la valeur est égale à zéro). L'appariement génétique n'a produit aucune augmentation du déséquilibre sur l'une ou l'autre des variables (dernière colonne, avant dernière ligne, la valeur est égale à zéro). L'appariement génétique obtient la meilleure moyenne globale des différences de moyennes standardisées (quatrième colonne, dernière ligne, la valeur est égale à 0,0062) ainsi que la meilleure moyenne globale du pourcentage de réduction de l'écart (dernière colonne, dernière ligne, la valeur est égale à 91,0%). Finalement, il est à noter que

l'appariement génétique a performé de façon exceptionnelle quant à l'équilibrage de la variable la plus problématique, c'est-à-dire la variable la plus déséquilibrée, *Anxiété*. Avant l'appariement, la différence entre les deux groupes est de 0,3203 alors qu'après l'appariement génétique cette différence est de 0,0084, ce qui revient à une réduction de l'écart de 97,4%. Il est aussi à noter que les trois méthodes sous les trois schèmes ont très bien performé en ce qui concerne les variables *Dossier criminel – Mère* et *Dossier criminel – Père*. Autrement dit, la proportion de cas dont les parents ont un dossier criminel est identique dans presque toutes les situations après les appariements.

Une remarque s'impose en ce qui concerne le nombre de cas sélectionnés par chaque schème d'appariement. Les 69 cas-interventions sont identiques pour les douze ensembles, mais les cas-contrôles peuvent varier de deux façons. Si nous prenons les schèmes 1:1 sans remise et 2:1 sans remise (premier et troisième tableau), le nombre de cas-contrôle sera égal en ce qui concerne les trois méthodes d'appariement, 69 cas-contrôles pour le premier schème et 138 pour le second schème. Cependant, il faut noter que ces cas-contrôles ne seront pas les mêmes en général d'une méthode d'appariement à l'autre. C'est pour cette raison d'ailleurs qu'on compare les trois méthodes d'appariement, parce qu'elles ne sélectionnent pas les mêmes cas-contrôles à appairer aux cas-interventions. En ce qui concerne les schèmes d'appariement 1:1 avec remise et 2:1 avec remise (deuxième et quatrième tableau), nous constatons non seulement que le nombre de cas-contrôles varie d'une méthode à l'autre (58, 53 et 55, puis 87, 86 et 89), mais ce nombre n'égale pas 69 pour le premier schème et n'égale pas 138 pour le second. En fait, il y a bel et bien 69 cas-contrôle pour le premier schème et 138 cas-contrôle pour le deuxième schème, mais puisqu'il s'agit d'une sélection avec remise, quelque cas-contrôles ont été choisis à plusieurs reprises. Ainsi, pour la méthode d'appariement par la distance de Mahalanobis sous le schème d'appariement 1:1 avec remise, 58 cas-contrôle parmi les 138 disponibles ont été sélectionnés et plusieurs de ces cas-contrôles ont été appariés plus d'une fois afin d'atteindre le nombre de 69.

Les résultats obtenus ne sont pas surprenants pour deux raisons. En premier lieu, nous nous attendions à ce que la méthode d'appariement génétique soit la plus puissante compte tenu des considérations théoriques, empiriques et de simulation examinées au moment de la présentation de cette méthode. En second lieu, le schème d'appariement 1:1 avec remise a tendance à sélectionner un petit nombre de cas-contrôles, ce qui augmente les chances d'obtenir un groupe-contrôle très semblable au groupe-intervention. Remarquons cependant que l'appariement par le score de propension n'est pas à son optimum sous ce schème et produit même de moins bons résultats que la distance de Mahalanobis pour les schèmes avec remise et des résultats similaires pour les schèmes sans remise. Il se peut que le score de propension lui-même ne soit pas à blâmer, mais plutôt la méthode d'appariement proprement dite. On se rappellera que le score de propension sert à estimer la similitude entre deux cas, c'est-à-dire qu'il devient une mesure de distance entre deux cas. Une fois estimé, nous avons utilisé l'appariement prédateur (*greedy matching*) pour appairer les cas. Cette technique d'appariement a le défaut de ne pas être optimale comme nous l'expliquions lors de notre discussion sur l'appariement pas la distance de Mahalanobis.

À la défense de l'appariement par la distance de Mahalanobis et de l'appariement par le score de propension, nous pouvons dire qu'il est possible d'améliorer leurs résultats par un procédé itératif. L'appariement itératif peut être utilisé avec les trois méthodes d'appariement examinées dans ce mémoire, bien que l'appariement génétique intègre déjà une forme d'itération. La technique consiste à utiliser comme variables d'appariement, en plus des variables originales, une première série de termes d'interaction deux à deux sur l'ensemble des variables, y compris les variables originales élevées au carré. Si un terme d'interaction est mal équilibré, c'est-à-dire s'il se situe au-delà d'un critère, 0,10 dans notre cas, alors on l'inclut dans l'ensemble des variables d'appariement. On poursuit ainsi avec les nouveaux termes d'interactions formées par les variables originales et les termes d'interaction nouvellement inclus. On cherche ainsi à coller aux données (*model fitting*). En un certain sens, il s'agit d'imiter l'appariement génétique, ce qui laisse dire à Sekhon :

The literature on matching [...] makes clear the need to find algorithms which produce matched datasets with high levels of covariate balance. The fact that so many talented researchers over several years failed to produce a propensity score model which had a high degree of covariate balance is a cautionary tale. In situations like these, machine learning can come to the rescue. There is little reason for a human to try all of the multitude of models possible to achieve balance when a computer can do this more systematically and much faster. Even talented and well trained researchers need aid. (Sekhon, à paraître, p.8)

Le diagnostic de l'équilibre des variables révèle que les méthodes d'appariement produisent des résultats passablement différents. L'examen de nos critères nous incite à favoriser l'appariement génétique, ce qui concorde avec les résultats théoriques, empiriques et de simulation examinés auparavant. L'appariement génétique est moins sensible à l'absence de la propriété EPBR, donc plus robuste à différentes distributions des variables. De plus, on utilise l'appariement afin d'équilibrer les variables à l'aide des données en main et on cherche le plus grand équilibrage possible. Il s'agit d'une approche empirique ou encore en anglais de *model fitting*. Il n'y a donc aucune raison de se passer d'un algorithme automatique, tel que l'appariement génétique, si cet algorithme produit systématiquement de meilleurs résultats que les méthodes traditionnelles.

Bien qu'il nous paraisse évident que l'appariement génétique est la plus efficace des trois méthodes d'appariement examinées ici, nous procéderons à l'analyse causale sur les ensembles de données produits par les trois méthodes et les quatre schèmes afin de tenter d'en tirer des leçons.

5.2.3 L'effet causal

Une fois l'appariement effectué, le chercheur a le choix quant aux types d'analyses qu'il désire appliquer aux ensembles appariés. Il peut s'agir d'une simple différence des moyennes entre le groupe-intervention et le groupe-contrôle sur la variable d'intérêt, comme il peut s'agir d'une analyse plus sophistiquée telle une régression linéaire, une régression multiniveau, une analyse de survie, etc. En fait toutes les techniques d'analyse utilisées sur des données expérimentales peuvent être mises à contribution.

Puisque nous disposons de données longitudinales, nous procéderons à des analyses multiniveaux. Nous nous servons du temps dans la composante aléatoire du modèle avec les cas comme variable de regroupement. Les analyses ont été séparées en deux grandes étapes. Une première étape sert à déterminer le modèle de base. Le modèle de base doit être composé d'une variable temps et de la variable pour laquelle nous étudions l'effet, c'est-à-dire la variable qui identifie les deux groupes à l'étude. Afin de ne pas surcharger le texte, nous avons regroupé ces analyses préliminaires en annexe (V. l'annexe 0 Appendice 1: Les analyses multiniveaux préliminaires). Le modèle de base retenu est le suivant:

$$\text{Niveau 1 :} \quad \text{violence}_{ij} = \pi_{0i} + \pi_{1i}\text{temps}_{ij} + \pi_{2i}\text{temps}_{ij}^2 + \pi_{3i}\text{temps}_{ij}^3 + \varepsilon_{ij}$$

$$\text{Niveau 2 :} \quad \pi_{0i} = \gamma_{00} + \gamma_{01}\text{groupe}_i + \zeta_{0i}$$

$$\pi_{1i} = \gamma_{10} + \zeta_{1i}$$

$$\pi_{2i} = \gamma_{20} + \zeta_{2i}$$

$$\pi_{3i} = \gamma_{30}$$

Ces analyses préliminaires ont montré la forme sinusoïdale du parcours des individus tant pour le groupe-contrôle que pour le groupe-intervention, ce qui justifie la modélisation du temps en élevant cette variable au carré puis au cube. De plus, les interactions entre les termes de temps et la variable 'groupe' ne sont pas significatives (V. la section 0), ce qui revient à dire que l'intervention n'a pas eu d'impact sur l'évolution de la violence dans le temps.

La deuxième série d'analyse multiniveaux sert essentiellement à extraire le coefficient de la variable qui identifie les deux groupes, laquelle est assimilée à l'effet causal. Il n'est pas nécessaire ici de présenter une typologie de ces modèles. Nous n'allons pas choisir un modèle sur la base d'un examen des critères d'information tels que la déviance, l'AIC ou le BIC. En fait, le modèle a déjà été sélectionné. Rappelons-nous que l'appariement génétique doté d'un schème 1:1 avec remise a produit le meilleur équilibre entre les deux groupes. En principe, nous pourrions présenter les résultats de l'analyse causale de cet appariement seulement. Dans le cadre de ce mémoire et afin de mieux

comprendre les conséquences d'utiliser différentes méthodes d'appariement, nous présentons quatorze résultats d'analyses causales. Treize de ces quatorze analyses utilisent le même modèle multiniveaux présenté ci-haut. La différence entre ces treize analyses étant l'ensemble de données utilisé. Un premier modèle utilise les 250 cas originaux, tandis que les douze autres modèles utilisent les cas sélectionnés par les différentes méthodes sous différents schèmes. Le quatorzième modèle utilise les 250 cas originaux, mais le modèle multiniveaux de base est augmenté des variables d'appariement. Ces résultats sont présentés au Tableau 11.

Tableau 11 : Effets de l'intervention

	Coefficient	Err. Stand.	T	p
Avant les appariements				
sans covariables	-0,0428	0,0302	-1,4172	0,1578
avec covariables	-0,0238	0,0311	-0,7672	0,4438
Appariement 1:1 sans remise				
Mahalanobis	-0,0443	0,0387	-1,1442	0,2548
Score	0,0039	0,0357	0,1102	0,9125
Génétique	-0,0549	0,0372	-1,4757	0,1425
Appariement 1:1 avec remise				
Mahalanobis	-0,0065	0,0379	-0,1728	0,8632
Score	-0,0149	0,0396	-0,3770	0,7069
Génétique	-0,0354	0,0412	-0,8582	0,3926
Appariement 2:1 sans remise				
Mahalanobis	-0,0398	0,0315	-1,2621	0,2085
Score	-0,0392	0,0315	-1,2443	0,2149
Génétique	-0,0444	0,0324	-1,3687	0,1727
Appariement 2:1 avec remise				
Mahalanobis	-0,0414	0,0365	-1,1330	0,2592
Score	-0,0405	0,0357	-1,1339	0,2588
Génétique	-0,0412	0,0361	-1,1416	0,2555

La première analyse porte sur l'ensemble de données originales (ligne *sans covariables*). La deuxième analyse inclut les variables dont nous nous sommes servis pour produire les appariements. Autrement dit, au lieu d'apparier les cas, nous nous servons des variables directement dans un modèle de régression. Il s'agit de la façon classique de contrôler l'effet de certaines variables dans un cadre de régression. Ces deux analyses utilisent donc l'ensemble des cas à notre disposition, soit 250 cas répartis en 69 cas-interventions et 181 cas-contrôles. Les douze autres analyses portent respectivement sur les ensembles de données produits par les trois méthodes d'appariement sous les quatre

schèmes retenus. Ces douze analyses utilisent donc les cas sélectionnés à l'étape de l'appariement. Nous avons grisé la ligne des résultats de l'appariement génétique sous le schème 1:1 avec remise puisqu'il s'agit de la méthode qui a produit le meilleur équilibrage des variables. (V. Tableau 10). En principe, il s'agit du résultat que nous devrions présenter si nos analyses faisaient l'objet d'une publication.

On constate qu'aucune des analyses ne détecte une différence significative entre les deux groupes (dernière colonne), bien que l'estimation du coefficient puisse varier considérablement d'un schème à l'autre (première colonne). Ainsi, en valeur absolue, nous passons de 0,0039 à 0,0444. Le plus étonnant est le fait que les variations sont énormes à l'intérieur des schèmes d'appariement. Ainsi, sous le schème 1:1 avec remise, nous obtenons des valeurs du coefficient de -0,0065, -0,0149 et -0,0354. Si les techniques d'appariement choisissaient sensiblement les mêmes cas-contrôles à appairier aux cas-interventions, alors nous aurions des résultats semblables à l'intérieur de chaque schème. Puisque ce n'est pas le cas, puisque la notion de similitude diffère d'une technique d'appariement à l'autre, le choix d'une technique doit reposer sur des critères externes aux techniques. C'est pour cette raison que l'examen de l'équilibre des variables est utilisé. Peu importe la technique ou le schème utilisés, nous pouvons toujours examiner l'équilibre des variables après un appariement. Bref, l'estimation du coefficient de l'effet causal est sensible à la technique et au schème d'appariement utilisé.

En ce qui concerne les erreurs standards, elles sont similaires sous chaque schème d'appariement. Elles sont aussi assez élevées, notamment sur les données appariées. C'est une situation assez courante lorsqu'on réduit la taille d'un échantillon. Dans le cadre d'une analyse causale, cette situation n'est pas considérée problématique puisqu'on cherche plus à réduire le biais de l'estimateur que l'efficacité de cet estimateur. L'efficacité d'un estimateur est la variation de la variable autour de sa moyenne. Plus cette variation est petite, plus l'estimateur est efficace. Cependant, un estimateur efficace, mais qui cause systématiquement un biais n'est pas d'une grande utilité pour évaluer l'impact d'un programme ou d'une intervention. Il est possible de réduire l'erreur standard en adoptant un

schème d'appariement 2:1, avec ou sans remise, c'est-à-dire en appariant deux cas-contrôles à un cas-intervention. C'est ce que nous constatons au Tableau 11 si nous comparons les erreurs standards du schème 1:1 sans remise (0,0387; 0,0357; 0,0372) aux erreurs standards du schème 2:1 sans remise (0,0315; 0,0315; 0,0324) et si nous comparons les erreurs standards du schème 1:1 avec remise (0,0379; 0,0396; 0,0412) aux erreurs standards du schème 2:1 avec remise (0,0365; 0,0357; 0,0361).

Finalement, la valeur p varie énormément d'une situation à l'autre, ce qui reflète bien entendu les variations du coefficient et dans une moindre mesure celles de l'erreur-type. Ainsi, nous passons d'une valeur p de 0,1425 à 0,9125. Puisque le coefficient et par conséquent la valeur p peuvent varier autant, avec des ensembles de données différents des nôtres, nous pouvons imaginer que certains schèmes produiraient une différence significative alors que d'autres schèmes n'en produiraient pas. De là l'importance du diagnostic des appariements. Nous pouvons nous fier seulement à la méthode et au schème qui a produit le meilleur équilibre.

Lors du diagnostic des résultats des appariements, nous avons déclaré que les résultats de l'appariement génétique étaient les plus fiables selon les critères retenus pour évaluer l'équilibre des variables après un appariement. Les résultats de l'analyse causale effectuée à partir de l'ensemble de données produit par l'appariement génétique peuvent donc servir d'étalon. Dans ces conditions, nous pouvons affirmer que toutes les analyses produisent un biais soit positif soit négatif. En effet, si nous comparons les coefficients à celui de l'appariement génétique sous le schème 1:1 avec remise (ligne grisée), nous constatons que les méthodes surestiment ou sous-estiment l'effet causal. En particulier, l'analyse sans co-variables surestime l'effet causal tandis que l'analyse avec co-variables sous-estime l'effet causal. Bien que dans le cas particulier présenté ici l'effet causal n'est pas significatif pour aucune des analyses, nous pourrions nous retrouver dans une situation plus ambiguë où une analyse produit un effet significatif et une autre n'en produit pas. Il est alors tentant de choisir l'analyse qui produit le résultat désiré. C'est pour cette raison qu'il

faut se doter de critères objectifs qui permettent de choisir la technique d'appariement qui rend les deux groupes aussi similaires que possibles.

Ces résultats nous invitent à être prudents. Au départ, nous étions préoccupés par un déséquilibre sur certaines variables, notamment la variable *Anxiété*. Nous aurions pu pour ainsi dire appairer à l'aveugle et utiliser n'importe quelle méthode d'appariement sous n'importe quel schème. Nous avons utilisé pour guide les résultats de l'appariement génétique puisque les recherches théoriques, empiriques et de simulation démontrent la supériorité de cette technique. Le plus important cependant réside dans le fait que c'est cette technique qui a le mieux équilibré nos variables. Au risque de se répéter, il est important de réaliser que c'est l'équilibre des variables qui détermine lequel des ensembles de données appariées devra être utilisé lors de l'analyse causale. Nos données furent mieux équilibrées avec l'appariement génétique sous le schème 1:1 avec remise (colonne grisée du Tableau 10). Il se pourrait cependant que des données différentes des nôtres soient mieux équilibrées avec une technique d'appariement distincte et un schème d'appariement distinct.

Notons, de façon plus spéculative, un dernier avantage de l'appariement génétique. Dans la mesure où les variables observées sont corrélées avec certaines variables inobservées, ces dernières subiront un rééquilibrage ou un déséquilibre selon le cas (Stuart, 2010). L'appariement génétique, en équilibrant fortement les neuf variables, a aussi équilibré partiellement toutes les variables corrélées à celles-ci. Pour ce qui est de l'appariement par la distance de Mahalanobis et l'appariement par le score de propension, non seulement est-il rare qu'ils équilibrent fortement des variables (une réduction de l'écart qui tendrait vers 100%), mais ils ont surtout tendance à fortement déséquilibrer certaines variables (les pourcentages négatifs du Tableau 10, deuxième sous-tableau). Il y a ici une asymétrie entre l'équilibre et le déséquilibre après un appariement. La réduction de l'écart en pourcentage ne peut être plus élevée que 100%. C'est ce qui s'est produit avec les variables *Doss. crim. – Père* et *Doss. crim. – Mère* sous la plupart des schèmes d'appariement. Les deux groupes, après appariement, ont la même moyenne standardisée

sur ces variables. Par contre, le déséquilibre produit par une méthode d'appariement peut être infini en théorie. Par exemple, l'appariement par le score de propension a produit une augmentation de l'écart de 554,4% sur la variable *Adversité* sous le schème 2:1 avec remise. En permettant à certaines variables, équilibrées à l'origine, de se déséquilibrer, ce nouveau déséquilibre se répercute sur toutes les variables corrélées à celles-ci, et ce d'autant plus si le nouveau déséquilibre est majeur. Ce phénomène ne semble jamais se produire avec l'appariement génétique puisque celui-ci ne produit presque jamais un déséquilibre.

6 Conclusion

Les résultats indiquent que le choix d'un schème d'appariement influence l'estimation de l'effet d'une intervention bien que, dans le cas qui nous concerne, les douze schèmes mènent à des conclusions similaires sur l'effet de l'intervention. L'ajustement des variables n'a pas permis de détecter de différences significatives entre les deux groupes en ce qui concerne la variable d'intérêt, *Violence*.

La leçon demeure cependant claire. La variation de la valeur de l'estimation du coefficient devrait nous inciter à être prudents lorsqu'on désire ajuster des données afin de rendre des groupes comparables. Il est impératif de se doter dès le début de critères clairs pour évaluer l'équilibre des variables. Ce sont ces critères qui détermineront quelle méthode et quel schème d'appariement sera utilisé pour évaluer l'effet causal.

Nous avons utilisé quatre critères pour évaluer les résultats de l'appariement. Le premier est la différence des moyennes standardisées, le critère de base sur lequel les autres critères reposent. Nous devons choisir arbitrairement une valeur au-delà de laquelle nous considérons que les deux groupes ne sont pas comparables sur une caractéristique particulière. Avec des données observationnelles, il n'est pas rare de choisir la valeur de 0,25. Nous avons opté pour la valeur plus restrictive de 0,10 puisque nos données proviennent d'une assignation aléatoire. Après un appariement, ce que nous souhaitons, c'est qu'il y ait moins de variables déséquilibrées qu'il en avait avant l'appariement, idéalement qu'il y en ait aucune.

Si ce premier critère ne parvient pas à distinguer les différents schèmes, un deuxième critère consiste alors à examiner le pourcentage de diminution ou d'augmentation de la différence des moyennes standardisées. Il se peut que cette différence augmente sur certaines variables. Il se peut que cette augmentation de la différence ne crée pas de déséquilibre selon notre premier critère. Cependant, si un des schèmes d'appariement ne produit que des diminutions de la différence standardisée, alors nous devons privilégier ce schème.

Si les deux premiers critères ne parviennent pas à distinguer les différents schèmes, nous pouvons faire appel à un troisième critère, la moyenne des différences des moyennes standardisées. Il suffit alors d'opter pour la moyenne la plus petite. Si nous avons encore de la difficulté à choisir un schème, nous pouvons finalement examiner la moyenne des pourcentages de réduction et de diminution des différences des moyennes standardisées. Pour ce dernier critère, on choisit la valeur positive la plus élevée.

Une fois sélectionné un schème d'appariement, selon les critères énumérés ci-dessus, l'analyse de l'effet causal se fait à l'aide de l'ensemble de données formé par ce schème. Tous les résultats obtenus à l'aide des autres schèmes ne peuvent être pris en considération, même si certains de ces résultats pourraient plaire au chercheur.

Deux limites à cette étude nous semblent importantes. La sélection des variables en vu de constituer un réseau causal formel n'a pas effectuée. On s'est contenté de reprendre les variables d'une étude antérieure (Paquin, 2010) afin de permettre des comparaisons. Cette question de la constitution d'un réseau causal à partir duquel il est possible de sélectionner les variables nécessaires à l'appariement est étudiée par Judea Pearl (Pearl, 2009). L'annexe 0 résume les résultats les plus importants obtenus par Pearl.

Étroitement associée à la constitution d'un réseau causale est l'examen de biais potentiels provoqués par l'existence d'une variable qui devrait se trouver dans le réseau causal mais qui n'a pas été observée ou mesurée. Ce type de biais est connu sous le nom de biais de sélection. Il existe des méthodes formelles pour évaluer l'impact potentiel d'un tel biais sur les résultats. Nous présentons à l'annexe 0 un résumé d'une de ces méthodes.

Les recherches se poursuivent sur l'ensemble du domaine de l'analyse causale, notamment pour l'élaboration de nouvelles techniques d'appariement. Nous présentons l'une d'elles à l'annexe 0.

Il va sans dire que tous ces résultats et l'ensemble de notre discussion mériteraient d'être validés par des travaux formels et par des simulations de Monte Carlo. Notamment, il serait intéressant de produire des simulations de Monte Carlo afin d'évaluer les différents

schèmes sur des données longitudinales dont le parcours des individus est de forme sinusoïdale. Les simulations de Monte Carlo permettent d'évaluer les biais d'estimation directement, sans passer par l'étape d'examen de l'équilibre de variable après les appariements. Est-ce que le schème qui produirait une estimation sans biais selon les simulations serait le même que nous avons sélectionné selon nos critères d'évaluation de l'équilibre des variables, c'est-à-dire l'appariement génétique sous un schème 2:1?

Bibliographie

- Abadie, Alberto et Guido W. Imbens (2006), « Large Sample Properties of Matching Estimators for Average Treatment Effects », *Econometrica*, vol. 74, n° 1, p. 235-67.
- Bacon, Francis (1986), *Novum organum* (Traduit par Michel Malherbe et Jean-Marie Pousseur), 1^e éd., Paris: Presses universitaires de France, Épipiméthée, 349 p.
- Barker, E. D., F. Vitaro, É Lacourse, N. M. G. Fontaine, R. Carbonneau et R. E. Tremblay (2010), « Testing the developmental distinctiveness of male proactive and reactive aggression with a nested longitudinal experimental intervention », *Aggressive Behavior*, vol. 36, n° 2, p. 127-40.
- Bernert, Christopher (1983), « The Career of Causal Analysis in American Sociology », *The British Journal of Sociology*, vol. 34, n° 2, p. 230-54.
- Blalock, Hubert M. (1964), *Causal inferences in nonexperimental research*, Chapel Hill: University of North Carolina Press, 200 p.
- Boisjoli, R., F. Vitaro, É Lacourse, E. D. Barker et R. E. Tremblay (2007), « Impact and clinical significance of a preventive intervention for disruptive boys: 15-year follow-up », *The British Journal of Psychiatry*, vol. 191, n° 5, p. 415-19.
- Bunge, Mario Augusto (2006), *Chasing reality : strife over realism*, Toronto: University of Toronto Press, Toronto studies in philosophy, 342 p.
- Campbell, Donald Thomas, Julian Cecil Stanley et Nathaniel Lees Gage (1966), *Experimental and quasi-experimental designs for research*, Chicago: R. McNally, 84 p.
- Cook, Thomas D. et Donald Thomas Campbell (1979), *Quasi-experimentation : design and analysis issues for field settings*, Chicago, Ill.: Rand McNally, 405 p.
- Dehejia, Rajeev H. et Sadek Wahba (1998), « Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs », *National Bureau of Economic Research Working Paper Series*, n° 6586.
- Diamond, Alexis et J. S. Sekhon (2005), « Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies », *Review of Economics and Statistics*, vol. 94, n° 3.
- Duncan, Otis Dudley (1966), « Path Analysis: Sociological Examples », *American Journal of Sociology*, vol. 72, n° 1, p. 1-16.
- Durkheim, Émile (1981), *Le suicide : étude de sociologie*, Paris: Presses universitaires de France, Quadrige, 463 p.

- Fox, William (1999), *Statistiques sociales*, Bruxelles; Sainte-Foy, Québec: De Boeck Université ; Presses de l'Université Laval, 374 p.
- Guo, S. et M. W. Fraser (2010), *Propensity score analysis: Statistical methods and applications*, Thousand Oaks: Sage.
- Harder, Valerie S., Elizabeth A. Stuart et James C. Anthony (2010), « Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research », *Psychological methods*, vol. 15, n° 3, p. 234-49.
- Hernan, Miguel A. et J. M. Robins (en préparation), *Causal Inference*.
- Ho, D. E., K. Imai, G. King et E. A. Stuart (2007), « Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference », *Political Analysis*, vol. 15, n° 3, p. 199-236.
- Holland, P. W. (1986), « Statistics and causal inference », *Journal of the American Statistical Association*, vol. 81, n° 396, p. 945-60.
- Hong, G. et S. W. Raudenbush (2004), *Causal inference for multi-level observational data with application to kindergarten retention study*: University of Michigan.
- Iacus, Stefano M., Gary King et Giuseppe Porro (2011a), « Causal Inference without Balance Checking: Coarsened Exact Matching », *Political Analysis*.
- (2011b), « Multivariate Matching Methods That Are Monotonic Imbalance Bounding », *Journal of the American Statistical Association*, vol. 106, n° 493, p. 345-61.
- Imai, K., G. King, E. A. Stuart et D. E. Ho (2011), « MatchIt: Nonparametric Preprocessing for Parametric Causal Inference », *Journal of Statistical Software*, vol. 42, n° 8, p. 1-43.
- Imbens, Guido W. et Donald B. Rubin (2008), « Rubin causal model », dans *The New Palgrave Dictionary of Economics*, Steven N. Durlauf et Lawrence E. Blume (dir.), Basingstoke: Palgrave Macmillan.
- Jöreskog, Karl G (1977), « Structural equation models in the social sciences: Specification, estimation and testing », dans *Applications of Statistics*, Paruchuri R. Krishnaiah (dir.), Amsterdam: North-Holland Publishing Company, p. 265-87.
- Keele, L., « An overview of rbounds: An R package for Rosenbaum bounds sensitivity analysis with matched data », n° Journal, Electronic, 2010, consulté le.
- Kendall, Patricia L et Paul F Lazarsfeld (1950), « Problems of survey analysis », dans *Continuities in social research*, R. K. Merton et P. F. Lazarsfeld (dir.), Glencoe: Free Press, p. 133-96.

- Lacourse, É, S. Coté, D. S. Nagin, F. Vitaro, M. Brendgen et R. E. Tremblay (2002), « A longitudinal-experimental approach to testing theories of antisocial behavior development », *Development and psychopathology*, vol. 14, n° 4, p. 909-24.
- LaLonde, Robert J. (1986), « Evaluating the Econometric Evaluations of Training Programs with Experimental Data », *The American Economic Review*, vol. 76, n° 4, p. 604-20.
- Mahalanobis, P. C. (1927), « Analysis of race-mixture in Bengal », *Journal of Asiatic Society of Bengal*, vol. 23, p. 301-33.
- Marchand, A. (2000), « L'analyse multi-niveaux avec MlwiN. Note de cours. », *GRASP/École de relations industrielles/Université de Montréal*.
- Mill, John Stuart (1970), *A system of logic, ratiocinative and inductive, being a connected view of the principles of evicence and the methods of scientific investigation*, London: Longman, 622 p.
- Morgan, S. L. et C. Winship (2007), *Counterfactuals and causal inference: Methods and principles for social research*, New York NY: Cambridge University Press.
- Nagin, D. S. et R. E. Tremblay (2001), « Parental and early childhood predictors of persistent physical aggression in boys from kindergarten to high school », *Archives of General Psychiatry*, vol. 58, n° 4, p. 389-94.
- Neyman, Jerzy Splawa (1990), « On the application of probability theory to agricultural experiments. Essay on principles. Section 9 », *Statistical Science*, vol. 5, n° 4, p. 465-80.
- Paquin, Stéphane (à venir), « Assessing the Effect of a Preventive Intervention on Violence Trajectories: Sensitivity to Missing Data Assumptions ».
- , « Comparaison de quatre méthodes pour le traitement des données manquantes au sein d'un modèle multiniveau paramétrique visant l'estimation de l'effet d'un programme de prévention », Université de Montréal, 2010.
- Pearl, J. (2009), *Causality: Models, Reasoning, and Inference*, New York NY: Cambridge University Press.
- (2010), « An introduction to causal inference », *Int J Biostat*, vol. 6, n° 2, p. Article 7.
- Rosenbaum, P. R. (2010), *Design of observational studies*, New York NY: Springer.
- (2002), *Observational studies*, New York NY: Springer.
- (2005), « Sensitivity Analysis in Observational Studies », dans *Encyclopedia of Statistics in Behavioral Science*(dir.): John Wiley & Sons, Ltd.
- Rosenbaum, P. R. et D. B. Rubin (1983), « The Central Role of the Propensity Score in Observational Studies for Causal Effects », *Biometrika*, vol. 70, n° 1, p. 41-55.

- Rubin, D. B. (2007), « The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials », *Statistics in medicine*, vol. 26, n° 1, p. 20-36.
- (1980), « Discussion of 'Randomization Analysis of Experimental Data: The Fisher Randomization Test' by D. Basu », *Journal of the American Statistical Association*, vol. 75, n° 371, p. 591-93.
- Rubin, D. B. et Neal Thomas (1992), « Characterizing the effect of matching using linear propensity score methods with normal distributions », *Biometrika*, vol. 79, n° 4, p. 797-809.
- Rubin, Donald B. (1976), « Multivariate Matching Methods That are Equal Percent Bias Reducing I: Some Examples », *Biometrics*, vol. 32, n° 1, p. 109-20.
- Rubin, Donald B. et Neal Thomas (1992), « Affinely Invariant Matching Methods with Ellipsoidal Distributions », *The Annals of Statistics*, vol. 20, n° 2, p. 1079-93.
- Sekhon, J. S. (à paraître), « Alternative Balance Metrics for Bias Reduction in Matching Methods for Causal Inference », *The Review of Economics and Statistics*.
- (2008), « The Neyman-Rubin model of causal inference and estimation via matching methods », dans *The Oxford Hand book of Political Methodology*, J. M. Box-Steffensmeier, H. Brady et D. Collier (dir.), Oxford: Oxford University Press, p. 271-99.
- Shipley, Bill (2000), *Cause and correlation in biology : a user's guide to path analysis, structural equations, and causal inference*, Cambridge, RU ; New York, NY, ÉU: Cambridge University Press, 317 p.
- Singer, Judith D (1998), « Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models », *Journal of educational and behavioral statistics*, vol. 23, n° 4, p. 323-55.
- Singer, Judith D. et John B. Willett (2003), *Applied longitudinal data analysis : modeling change and event occurrence*, Oxford ; Toronto: Oxford University Press, xx, 644 p.
- Sobel, Michael E. (1996), « An Introduction to Causal Inference », *Sociological Methods & Research*, vol. 24, n° 3, p. 353-79.
- Stuart, E. A. (2010), « Matching Methods for Causal Inference: A Review and a Look Forward », *Statistical Science*, vol. 25, n° 1, p. 1-21.
- Tremblay, R. E. (dir.) (2004), *Étude longitudinale et expérimentale de Montréal : guide d'intervention et effets à long terme*, Montréal: Groupe de recherche sur l'inadaptation psychosociale de l'enfant, 372 p.
- Tremblay, R. E., R. Loeber, C. Gagnon, P. Charlebois, S. Larivée et M. LeBlanc (1991), « Disruptive boys with stable and unstable high fighting behavior patterns during

junior elementary school », *Journal of Abnormal Child Psychology*, vol. 19, n° 3, p. 285-300.

Vitaro, F., M. Brendgen et R. E. Tremblay (2001), « Preventive Intervention: Assessing its Effects on the Trajectories of Delinquency and Testing for Mediational Processes », *Applied Developmental Science*, vol. 5, n° 4, p. 201-13.

Appendices

Appendice 1: Les analyses multiniveaux préliminaires

Ces analyses préliminaires ont été réalisées à l'aide du logiciel SAS 9.1.

Analyses multiniveaux pour études longitudinales

D'emblée, lorsqu'on dispose de données longitudinales, on se trouve en quelque sorte avec des données multiniveaux. Si l'étude porte sur des individus, alors l'individu est l'équivalent des groupes de deuxième niveau d'une étude transversale. Les mesures répétées dans le temps sur le même individu sont équivalentes aux mesures répétées sur des individus à l'intérieur d'un groupe dans une étude transversale.

Mise à part les conditions d'application des algorithmes multiniveaux pour modèles linéaires, que nous traiterons lors du diagnostic de notre modèle, nous devons disposer d'au moins trois vagues de mesure et d'au moins 30 à 40 individus. Ces conditions sont rencontrées pour notre étude. Remarquons, en ce qui concerne le nombre de vagues de mesure, certains individus peuvent ne pas avoir été mesurés à toutes les vagues. C'est souvent le cas dans une étude longitudinale lorsque des participants quittent l'étude à chaque vague : il s'agit alors d'un problème d'attrition. Nous en reparlerons lorsque nous traiterons des données manquantes.

L'analyse multiniveaux offre plusieurs avantages. Elle permet de tenir compte de l'interdépendance des mesures répétées. Elle permet aussi, et c'est sûrement ce qui la caractérise le mieux, d'attribuer une partie de la variation de la variable dépendante au niveau 1 et une autre partie au niveau 2, i.e. que la variable d'intérêt dépend à la fois de facteurs qui distinguent les individus entre eux (niveau 2) et de facteurs qui distinguent un même individu dans le temps (niveau 1). On dit aussi que les variables de niveau 2 représentent des caractéristiques de l'individu qui ne varient pas dans le temps et celles de niveau 1 représentent des caractéristiques qui varient dans le temps. D'ailleurs, certaines de ces variables de niveau 1 peuvent servir de variables temporelles si ceci fait du sens pour la question à l'étude, ce qui offre beaucoup de flexibilité dans la modélisation du temps

(Singer et Willett, 2003, chap. 5 et 6). En termes de régression, l'introduction du niveau 2 permet de tenir compte de la variation qu'on retrouve entre les individus concernant l'ordonnée à l'origine et les pentes.

L'approche multiniveaux permet d'éviter deux erreurs : l'imputation des variables de groupes aux individus et l'agrégation de variables individuelles au niveau des groupes (Marchand, 2000). Dans le premier cas, on injecte de l'information redondante, ce qui a pour effet de réduire les écarts-types et de découvrir des relations significatives là où il n'y en a peut-être pas. Dans le deuxième cas, il y a perte d'information sur la variation intra-groupe et le chercheur risque de commettre l'erreur écologique en croyant analyser des individus alors qu'il analyse des groupes.

L'intégration des scores de propension aux analyses multiniveaux

Une analyse causale de données observationnelles tente de reproduire les conditions d'une analyse causale de données expérimentales. Les données expérimentales sont recueillies selon un plan expérimental. Plusieurs types de plans expérimentaux sont utilisés et selon le plan expérimental choisi, le chercheur aura recours à une analyse de variance particulière.

Lorsqu'on désire procéder à une analyse causale de données observationnelles à l'aide de scores de propension, l'estimation de ces scores diffèrera selon le plan expérimental qu'on tente de simuler. Des chercheurs ont démontré ce lien entre plan expérimental et la façon d'estimer les scores de propension (Hong et Raudenbush, 2004). Ils exposent trois plans expérimentaux et leurs équivalents observationnels. Pour ces derniers, ils démontrent, sous certaines conditions, la nature des scores de propension nécessaires à rendre chaque étude observationnelle équivalente à sa contrepartie expérimentale.

En ce qui nous concerne, notre étude est de type observationnelle par grappe (*cluster observational study*), ce qui fait qu'on doit simuler un plan expérimental par grappe (*cluster-randomized design*). Dans ce type d'étude, le traitement est au niveau de la grappe

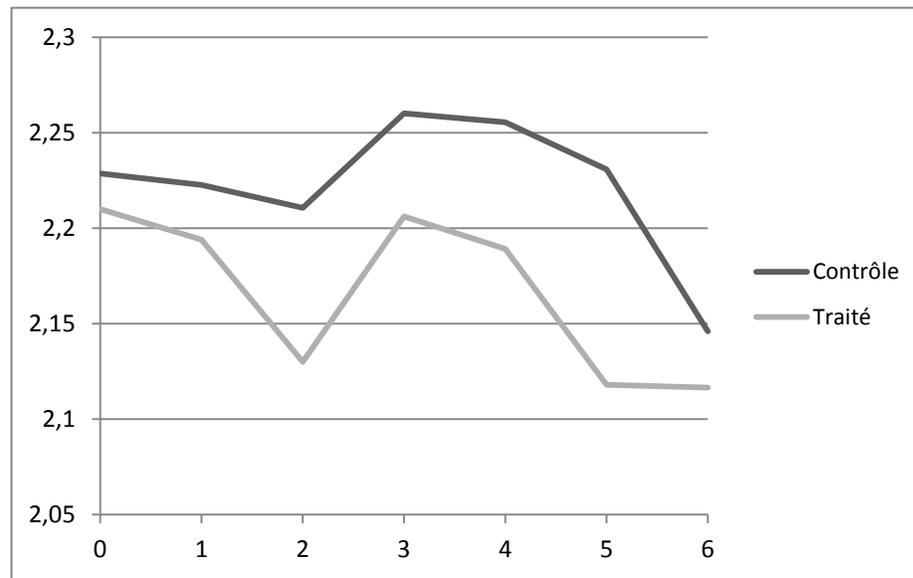
et les individus d'une grappe reçoivent tous le même traitement, ce qui revient à dire que les individus sont des mesures répétées pour une grappe. Dans notre cas, puisque nous disposons de données longitudinales, les individus sont les grappes et les mesures à travers le temps sont les mesures répétées à l'intérieur de la grappe.

Les mêmes auteurs (Hong et Raudenbush, 2004, p.1853) démontre alors que sous ce type d'étude les scores de propension sont des probabilités prédites d'appartenir au groupe traité, étant données les covariables au niveau de la grappe (individu dans notre cas). Ainsi, les covariables au niveau de l'individu (les variables qui varient dans le temps dans notre cas) ne servent pas à l'estimation des scores de propension. Autrement dit, on se sert avec ce genre d'étude que des caractéristiques mesurées au début d'une étude longitudinale.

Les analyses graphiques de données longitudinales

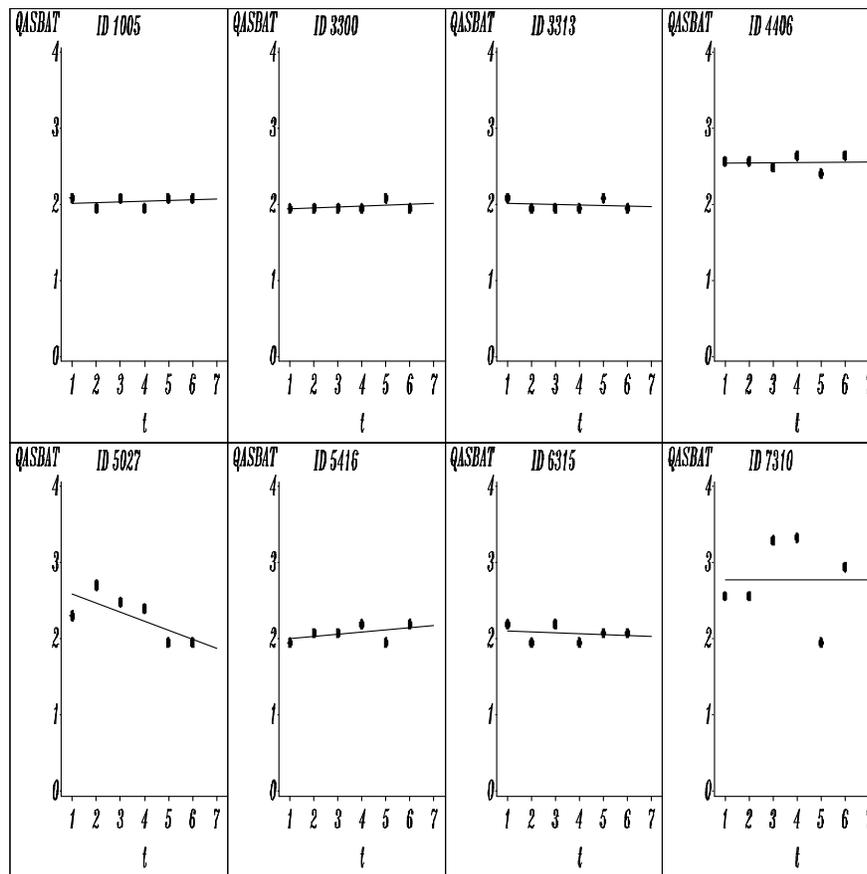
Nous présentons ici quelques résultats graphiques de nos données longitudinales. Nous débutons par une représentation dans le temps des moyennes brutes pour les deux groupes. Nous remarquons d'emblée la nature sinusoïdale de ces courbes. Notre modèle multiniveaux devra donc comporter des termes polynomiaux jusqu'au degré 3. Nous constatons que les deux groupes suivent des trajectoires pratiquement identiques. S'il y a une différence entre les deux groupes, ce ne peut être dans la forme de la trajectoire mais plutôt dans l'écart entre les deux trajectoires. Nous remarquons aussi une tendance à la baisse des courbes.

Figure 2 : Moyennes brutes du log de Violence



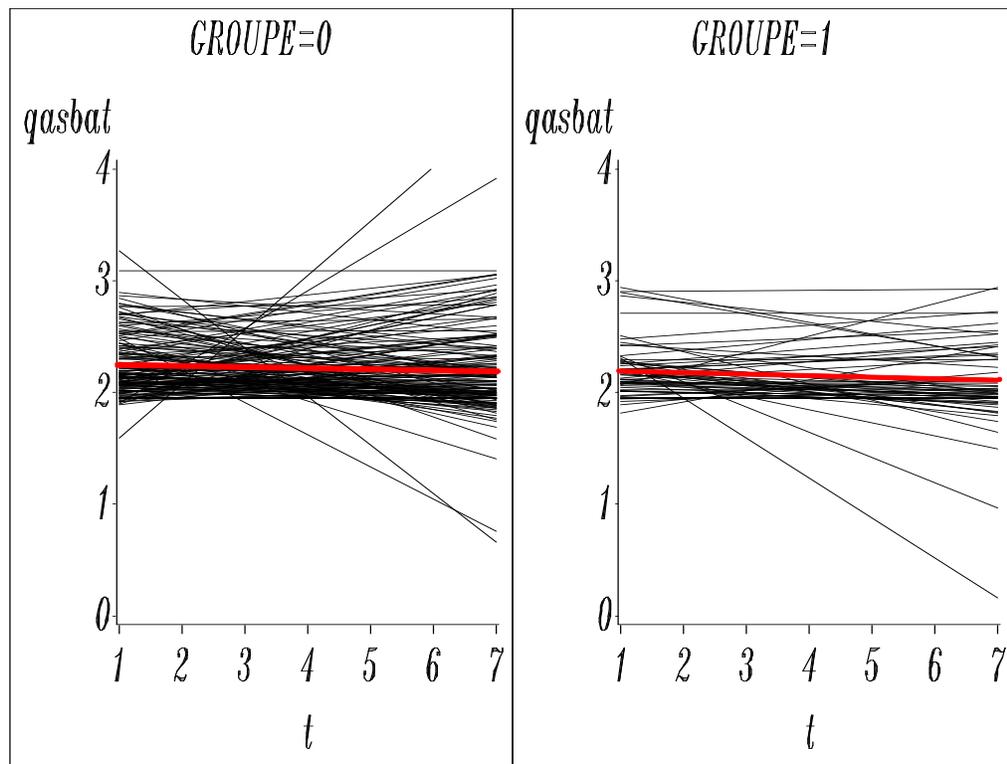
Examinons maintenant des graphiques dits empiriques, i.e. construits à l'aide des données brutes de tous les cas ou d'un échantillon de cas (Singer et Willett, 2003, Chap.4). Débutons par l'examen d'un sous-échantillon constitué de 8 cas sélectionnés de façon aléatoire sur les cas dont nous disposons de toutes les mesures (Figure 3). Les points représentent les données brutes de Violence et les courbes sont estimées par la méthode des moindres carrés ordinaires (MCO). Nous constatons que la majorité des cas ont possiblement une tendance linéaire nulle, bien que les points soient situés légèrement au-dessus ou en-dessous de la droite, ce qui confirmerait la nature sinusoïdale des données. En fait, peu importe la tendance linéaire, tous les cas affichent une trajectoire sinusoïdale. Remarquons que le cas 7310 se démarque des autres cas par l'ampleur des variations.

Figure 3 : Graphique de croissance empirique avec trajectoires MCO



L'examen de la variation des ordonnées à l'origine et de la variation des pentes se fait en disposant sur un graphique les courbes MCO de tous les cas et ce pour chaque groupe (Figure 4). Nous constatons une assez grande variation des ordonnées, peut-être avec un peu plus d'ampleur pour le groupe-contrôle (groupe=0) que pour le groupe traité (groupe=1). Mais il faut se rappeler que le groupe-contrôle est constitué de 181 cas et le groupe traité par 59 cas. Bien que ce ne soit pas très apparent ici, il semble se dégager au moins deux sous-groupes pour chacun des groupes. Un sous-groupe plutôt stable autour de la trajectoire moyenne (en rouge) et un sous-groupe dont la tendance est à l'augmentation de Violence (qasbat dans le graphique).

Figure 4 : Trajectoires MCO pour tous les cas et pour chaque groupe



Dans les deux groupes, nous retrouvons quelques cas exceptionnels, avec une tendance marquée soit à une augmentation de la violence soit à une diminution. Les trajectoires moyennes (ligne grisée au centre) sont pratiquement identiques dans les deux groupes. Ainsi, en général, les deux groupes ne semblent pas se distinguer quand à l'expression de la violence. Pour un même individu, le schéma comportemental le plus courant semble être une même expression de la violence à tous les âges, avec peut-être une légère variation sinusoïdale telle que dégagée par l'examen des moyennes brutes (Figure 2) et des graphiques de croissance empirique (Figure 3).

Les modèles multiniveaux de base

Nous avons élaboré deux séries de modèles. La première série concerne la modélisation du temps et la seconde a trait à l'ajout de la variable indiquant l'appartenance au groupe et l'examen de cette variable avec le temps. Ces modèles sont les suivants :

Modélisation 1 – Modélisation du temps

- Modèle B : Modèle de croissance inconditionnelle du temps linéaire
- Modèle C : Modèle de croissance inconditionnelle du temps quadratique
- Modèle E : Modèle de croissance inconditionnelle non aléatoire pour le terme cubique

En ce qui concerne la première série de modèles, nous constatons que tous les termes temporels sont significatifs pour les trois modèles (Tableau 12). De plus, nous constatons que les critères d'information décroissent substantiellement à chaque ajout d'un terme temporel. Testons un peu plus formellement ces modèles afin de la comparer. Puisque nos modèles sont imbriqués l'un dans l'autre et puisque nous utilisons les mêmes observations à chaque modèle, nous utiliserons la différence des déviations, laquelle suit asymptotiquement une distribution de χ^2 (Tableau 12). Le modèle M1E s'ajuste mieux que les deux autres modèles selon ce critère. Nous retenons le modèle M1E. Sous forme d'équations, le modèle se présente ainsi :

Niveau 1 :
$$qasbat_{ij} = \pi_{0i} + \pi_{1i} \text{temps}_{ij} + \pi_{2i} \text{temps}_{ij}^2 + \pi_{3i} \text{temps}_{ij}^3 + \varepsilon_{ij}$$

Niveau 2 :

$$\pi_{0i} = \gamma_{00} + \zeta_{0i}$$

$$\pi_{1i} = \gamma_{10} + \zeta_{1i}$$

$$\pi_{2i} = \gamma_{20} + \zeta_{2i}$$

$$\pi_{3i} = \gamma_{30}$$

Tableau 12 : Typologie des modèles de la modélisation du temps

		Modèle B	Modèle C	Modèle E
Effets fixes				
Statut initial	π_{0i} Ordonnée	γ_{00} 2.2279***	2.2076***	2.2260***
Taux de changement	π_{1i} Temps	γ_{10} 0.0073*	0.0186~	0.0464*
	π_{3i} Temps ²	γ_{20}	0.0044**	0.0255***
	π_{3i} Temps ³	γ_{30}		0.0034***
Composante de la variance				
Niveau 1	Intra-sujet, ε_{ij}	σ_ε 0.0330***	0.0294***	0.0288***
Niveau 2	Statut initial, ζ_{0i}	σ_0 0.0420***	0.0393***	0.0399***
Terme linéaire	Variance, ζ_{1i}	σ_1 0.0013***	0.0108***	0.0111***
	Covariance, ζ_{0i} et ζ_{1i}	σ_{01} 0.0021*	0.0033	0.0036
Terme quadratique	Variance, ζ_{2i}	σ_2	0.0002***	0.0002***
	Covariance, ζ_{0i} et ζ_{2i}	σ_{02}	0.0000	0.0000
	Covariance, ζ_{1i} et ζ_{2i}	σ_{12}	0.0014***	0.0014***
Corrélation intra-classe				
Statistiques d'ajustement	Déviante	-178.86	-226.25	-242.41
	AIC	-166.86	-206.25	-220.41
	BIC	-145.73	-171.03	-181.67
	χ^2		47.39(4df)***	16.16(1df)***
~p<.10; *p<.05; **p<.01; ***p<.001				

Que le modèle M1E s'ajuste mieux que les deux autres modèles n'a rien d'étonnant si on se souvient de la forme sinusoïdale des courbes des moyennes brutes dans le temps (Figure 2).

Modélisation 2 – Modélisation de l'effet de traitement

- Modèle A : Modélisation de l'interaction de l'appartenance au groupe avec trois termes d'interaction
- Modèle B : Modélisation de l'interaction de l'appartenance au groupe avec deux termes d'interaction
- Modèle C : Modélisation de l'interaction de l'appartenance au groupe avec un terme d'interaction
- Modèle D : Modélisation du temps et de l'appartenance au groupe sans interaction

La deuxième série de modèles modélisent le temps et l'appartenance au groupe sans pondération par les scores de propension. Notre point de départ est le modèle M1E ci-dessus auquel on ajoute des termes d'interaction entre la variable représentant les groupes et chacun des termes temporels. Ceci nous conduit au modèle M2A. Tous les termes d'interaction de ce modèle sont non significatifs. En retirant un à un les termes d'interaction, nous obtenons le modèle M2D. À chaque étape les critères d'information ce

sont améliorés, notamment le BIC. Cependant, la déviance ne varie pas sensiblement. Nous retenons quand même le modèle M2D par sa parcimonie. À ce stade-ci, nous ne détectons pas de différence significative entre les deux groupes. Sous forme d'équations, le modèle retenu se présente ainsi :

$$\text{Niveau 1 :} \quad qasbat_{ij} = \pi_{0i} + \pi_{1i} \text{temps}_{ij} + \pi_{2i} \text{temps}_{ij}^2 + \pi_{3i} \text{temps}_{ij}^3 + \varepsilon_{ij}$$

$$\text{Niveau 2 :} \quad \pi_{0i} = \gamma_{00} + \gamma_{01} \text{groupe}_i + \zeta_{0i}$$

$$\pi_{1i} = \gamma_{10} + \zeta_{1i}$$

$$\pi_{2i} = \gamma_{20} + \zeta_{2i}$$

$$\pi_{3i} = \gamma_{30}$$

Pour les fins de l'exercice, comparons le modèle M2D au modèle M1E, i.e. deux modèles dont la seule variable qui les distingue est Groupe. Les critères d'information sont sensiblement les mêmes : $AIC_{M1E} = -220.41$ et $AIC_{M2D} = -220.40$; $BIC_{M1E} = -181.67$ et $BIC_{M2D} = -178.14$. Le test de la déviance nous donne un $\chi^2 = 1.99$, $df=1$, ce qui est non significatif même à $p < .10$. L'ajout du groupe ne change donc rien au modèle temporel de degré 3. Nous gardons tout de même le groupe dans le modèle puisque nous voulons examiner si un ajustement par les techniques d'appariement fera une différence.

Tableau 13: Typologie des modèles - Temps et groupe

			Modèle A	Modèle B	Modèle C	Modèle D
Effets fixes						
Statut initial	π_{0i}	Ordonnée	γ_{00} 2.2317***	2.2308***	2.2348***	2.2380***
		Groupe	γ_{01} 0.0203	0.0172	0.0316	0.0429
Taux de changement	π_{1i}	Temps	γ_{10} 0.0396~	0.0365~	0.0448*	0.0462*
		Temps*Groupe	γ_{11} 0.0233	0.0343	0.0048	
	π_{3i}	Temps ²	γ_{20} 0.0256**	0.0241**	0.0254***	0.0255***
		Temps ² *Groupe	γ_{21} 0.0003	0.0048		
	π_{3i}	Temps ³	γ_{30} 0.0035***	0.0034***	0.0034***	0.0034***
	Temps ³ *Groupe	γ_{30} 0.0006				
Composante de la variance						
Niveau 1		Intra-sujet, ε_{ij}	σ_ε 0.0288***	0.0288***	0.0288***	0.0288***
Niveau 2		Statut initial, ζ_{0i}	σ_0 0.0398***	0.0398***	0.0399***	0.0400***
Terme linéaire		Variance, ζ_{1i}	σ_1 0.0109***	0.0109***	0.0110***	0.0111***
		Covariance, ζ_{0i} et ζ_{1i}	σ_{01} 0.0037	0.0037	0.0038	0.0039
Terme quadratique		Variance, ζ_{2i}	σ_2 0.0002***	0.0002***	0.0002***	0.0002***
		Covariance, ζ_{0i} et ζ_{2i}	σ_{02} 0.0000	0.0000	0.0000	0.0000
		Covariance, ζ_{1i} et ζ_{2i}	σ_{12} 0.0014***	0.0014***	0.0014***	0.0014***
Statistiques d'ajustement		Déviante	-246.51	-246.42	-244.78	-244.40
		AIC	-216.51	-218.42	-218.78	-220.40
		BIC	-163.69	-169.12	-173.00	-178.14
~ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$						

Modélisation de la structure de variance-covariance intra-sujet

Il n'est pas réaliste de supposer que les résidus intra-sujet ne sont pas corrélés à travers le temps. En examinant différentes matrices de variance-covariance, on tente de trouver la structure qui rend le mieux compte de cette corrélation. Nous nous limitons à l'analyse de la structure de covariance intra-sujet, tel que démontrée dans (Singer, 1998). Nous supposons alors que les résidus inter-sujets sont indépendants.

Nous débutons avec un modèle de croissance inconditionnel dont le statut initial et le taux de changement sont considérés fixes :

Niveau 1 : $qasbat_{ij} = \pi_{0i} + \pi_{1i}TEMPS_{ij} + \varepsilon_{ij}$

Niveau 2 : $\pi_{0i} = \gamma_{00}$

$\pi_{1i} = \gamma_{10}$

Composite : $qasbat_{ij} = \gamma_{00} + \gamma_{10}TEMPS_{ij} + [\varepsilon_{ij}]$

où $\varepsilon_{ij} \sim N(0, \Sigma)$ et Σ représente une matrice de variance-covariance qui tient compte de la corrélation intra-sujet et qui reste à spécifier.

Nous modélisons ce modèle trois fois de suite, en utilisant les structures *un* (sans structure), *ar(1)* (auto-régressif) et *cs* (symétrie composée), les structures les plus utilisées. Il faut alors comparer les critères d'information et choisir la structure qui produit la valeur la moins élevée pour ces critères (Tableau 14).

Tableau 14 : Comparaison de trois structures de variance-covariance

Statistiques d'ajustement	Sans structure	Symétrie composée	Auto-régressif
-2 Log Likelihood	-225.80	-38.40	-74.82
AIC (smaller is better)	-161.80	-26.40	-62.82
AICC (smaller is better)	-160.20	-26.34	-62.76
BIC (smaller is better)	-49.11	-5.27	-41.69

Les quatre critères nous indiquent que le modèle standard avec structure « sans structure » est le meilleur choix. Nous n'avons pas à modéliser une structure de corrélation intra-sujets. On peut donc supposer que les résidus des sujets à travers le temps sont indépendants.

Diagnostic

Aucune analyse n'est complète si les conditions d'application d'une technique d'estimation ne sont pas vérifiées. Nous diagnostiquons donc notre modèle final en suivant les conseils de (Singer et Willett, 2003, p.127). Nous examinons les conditions de linéarité, de normalité et d'homoscédasticité de notre modèle. Pour chacune des ces conditions, le diagnostic doit se faire tant au niveau 1 qu'au niveau 2. Autrement dit, nous devons examiner les résidus intra-sujets et les résidus inter-sujets. Nous reproduisons ici notre modèle afin de faciliter la discussion.

$$\text{Niveau 1 : } \quad qasbat_{ij} = \pi_{0i} + \pi_{1i}\text{temps}_{ij} + \pi_{2i}\text{temps}_{ij}^2 + \pi_{3i}\text{temps}_{ij}^3 + \varepsilon_{ij}$$

$$\text{Niveau 2 : } \quad \pi_{0i} = \gamma_{00} + \gamma_{01}\text{groupe}_i + \zeta_{0i}$$

$$\pi_{1i} = \gamma_{10} + \zeta_{1i}$$

$$\pi_{2i} = \gamma_{20} + \zeta_{2i}$$

$$\pi_{3i} = \gamma_{30}$$

Linéarité

Au niveau 1, l'examen de la linéarité peut se faire à l'aide des graphiques de croissance empirique que nous avons présentés précédemment (Figure 3). Puisque nous observons des trajectoires sinusoïdales, trajectoires non linéaires en principe, celles-ci peuvent être linéarisées en incluant des termes temporels quadratiques et cubiques.

Au niveau 2, nous avons modélisé quatre paramètres, l'un lié à l'ordonnée à l'origine et les trois autres liés aux trois termes temporels du niveau 1. Les trois paramètres liés aux termes temporels (π_{1i} , π_{2i} , π_{3i}) n'ont pas à être examinés puisque nous n'avons inclus aucune covariable dans la modélisation de ces paramètres. Le paramètre π_{0i} n'a pas à être examiné non plus puisque la variable Groupe est une variable dichotomique et qu'une variable dichotomique est liée automatiquement de façon linéaire à une variable d'échelle, puisqu'avec deux modalités seule une courbe linéaire peut être ajustée.

Normalité

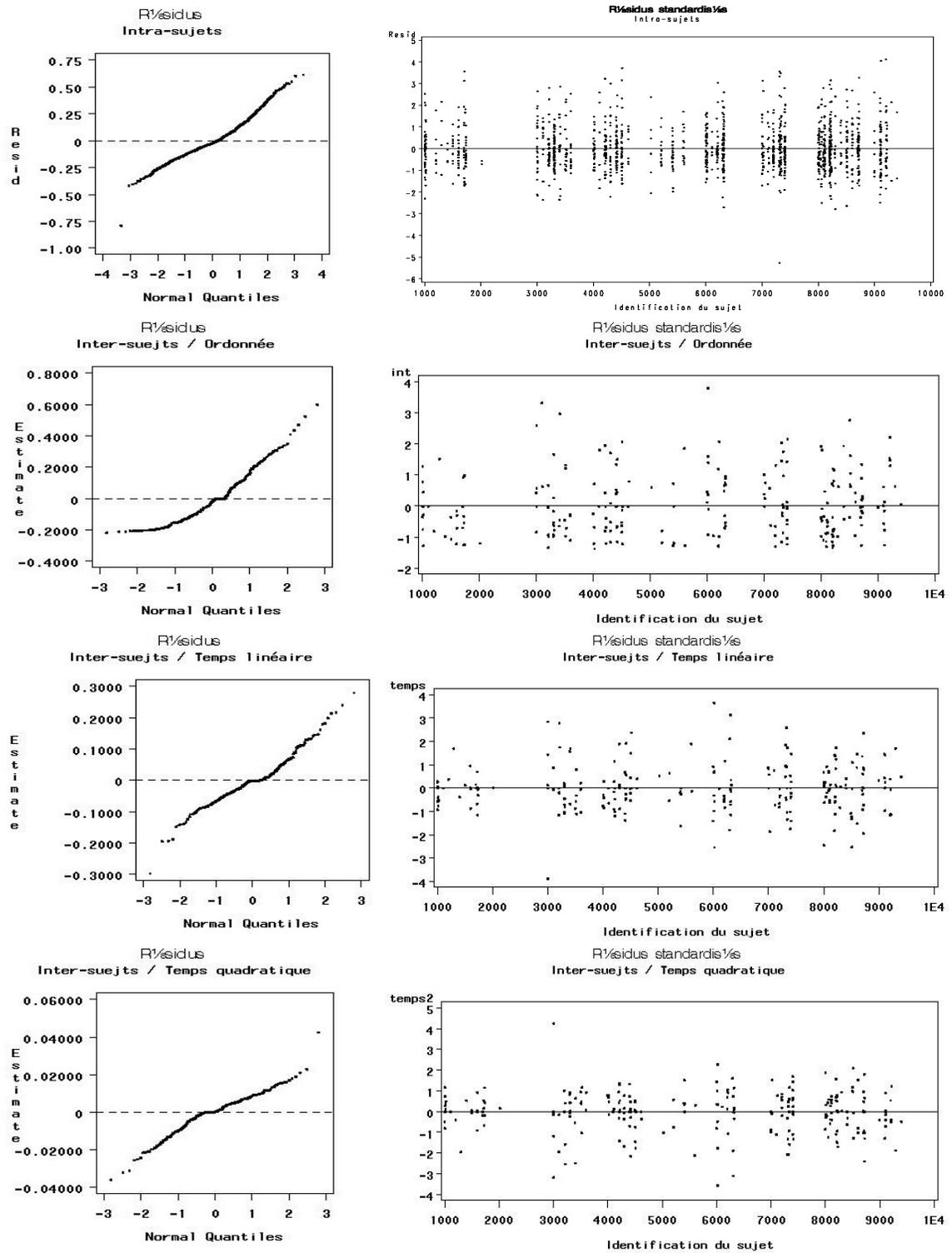
Nous devons examiner la normalité des résidus intra-sujets et des résidus inter-sujets pour les paramètres π_{0i} , π_{1i} et π_{2i} . Nous n'avons pas de résidus pour le paramètre π_{3i} puisque nous n'avons pas modélisé un effet aléatoire. La Figure 5 présente deux séries de graphiques pour l'examen de la normalité des résidus. La colonne de gauche présente la distribution des résidus par rapport à une distribution normale à l'aide de graphiques de probabilité normale (*Normal probability plot; QQPLOT*) et la colonne de droite présente la distribution des résidus standardisés des sujets.

Les résidus intra-sujets sont bien distribués (les deux graphiques du haut). Nous remarquons un cas aberrant dans le coin inférieur gauche du graphique de probabilité normale. Pour les résidus standardisés, seuls quelques sujets se trouvent au-dessus de la valeur absolue de trois déviations standards, y compris notre cas aberrant identifié tantôt, qui se trouve au-dessus de la valeur absolue de cinq déviations standards.

Par contre, il y a un problème avec la distribution des résidus inter-sujets (les trois paires de graphiques représentent les résidus inter-sujets pour l'ordonnée, le temps linéaire et le temps quadratique). Les résidus ne semblent pas suivre une distribution normale (graphique de gauche), notamment pour les résidus associée à l'ordonnée. De ces mêmes graphiques, on remarque quelques valeurs extrêmes parmi les résidus. Les graphiques des résidus standardisés (graphiques de droite) nous montrent peu de valeurs au-dessus de la valeur absolue de trois déviations standards. Quelques cas extrêmes apparaissent encore ici.

Nous ne croyons pas que les défauts qui sont apparus dans la discussion précédente soient critiques. Mais il demeure qu'un examen approfondi devra être entrepris dans le cours de nos recherches ultérieures.

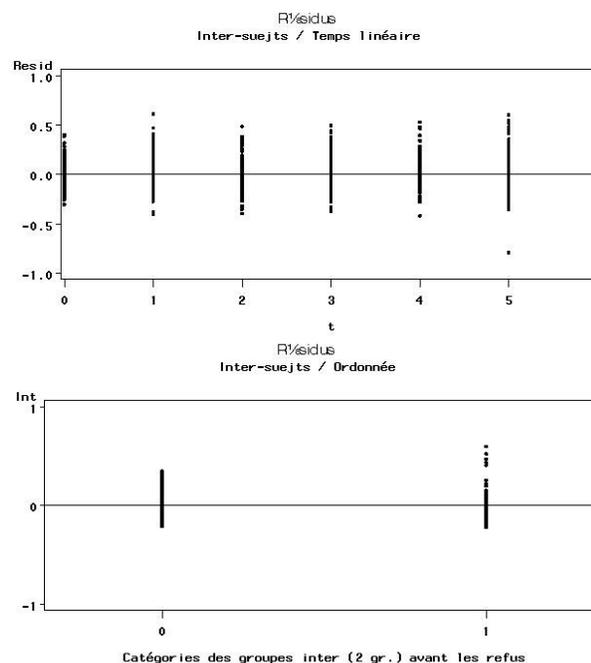
Figure 5 : Examen de la normalité des résidus



Homoscédasticité

En ce qui concerne l'homoscédasticité, nous devons examiner la distribution des résidus pour chaque niveau d'une variable prédictive. Ceci nous conduit à deux examens : examiner les résidus de niveau 1 à chaque niveau du temps linéaire et examiner les résidus de niveau 2 pour l'ordonnée à chaque niveau de Groupe. L'examen des résidus de niveau 2 pour les paramètres π_{1i} , π_{2i} et π_{3i} non pas à avoir lieu puisque nous n'avons pas introduit de variables prédictives dans l'estimation des ces paramètres. De plus, l'examen des résidus de niveau 1 pour le temps quadratique et le temps cubique donne exactement les mêmes résultats que l'examen du temps linéaire seulement. Notons que les résidus présentés sont les résidus brutes tout comme pour ceux de la colonne de gauche du la Figure 5. Ainsi, de la Figure 6, nous constatons une assez bonne homoscédasticité des résidus, tant au niveau 1 qu'au niveau 2. Le graphique du haut nous montre encore une fois notre cas aberrant au sixième temps de mesure (t=5). Il y a peut-être un léger problème pour les résidus inter-sujet, graphique du bas, où certains résidus semblent s'éloigner vers le haut pour le groupe traité (Groupe=1).

Figure 6 : Examen de l'homoscédasticité



Appendice 2: Graphes causaux et sélection des variables

Lorsqu'un chercheur désire étudier l'effet d'un traitement, d'une intervention, d'un programme gouvernemental ou, de façon générale, l'effet d'une cause, il est placé devant un choix impressionnant d'approches. Plusieurs de ces approches ont été unifiées dans le cadre de la théorie de la causalité de Judea Pearl, informaticien et philosophe, professeur à l'Université de Californie à Los Angeles. La théorie de la causalité de Pearl repose sur une interprétation causale des graphes orientés acycliques de la théorie mathématique des graphes. Un problème classique en statistique est celui du choix des variables à considérer comme « contrôles » afin d'obtenir une estimation sans biais de l'effet d'une cause. En fait, il n'y a jamais eu de consensus sur ce sujet. L'utilisation des graphes orientés acycliques permet à Pearl de proposer des critères rigoureux quand au choix des variables à contrôler.

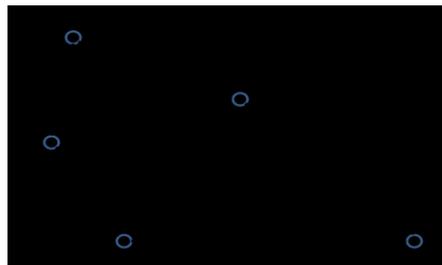
Les techniques que nous présentons n'ont pas pour but d'identifier un réseau causal. Aucun algorithme ne peut construire un réseau causal à l'exception de l'élaboration théorique d'un chercheur. Un réseau causal se met en place sur le long terme. À l'aide de multiples études et recoupements, des chercheurs peuvent s'entendre sur un réseau causal ou à tous les moins sur un petit ensemble de réseaux causaux. Les études expérimentales, lorsque possibles, peuvent être mises à contribution pour élaborer certaines sections du réseau causal. Remarquons cependant qu'il existe des algorithmes pour identifier des *familles* de réseaux causaux sur la base des relations statistiques (corrélation/covariance) qui existent entre les variables en cause. Ces algorithmes sont des outils puissants à la portée des chercheurs, mais ils n'identifient jamais un réseau causal unique (Shiple, 2000). Ainsi, les techniques qui suivent servent à identifier les ensembles de variables suffisants à l'estimation d'un effet causal. Nous verrons que l'identification de tels ensembles a des conséquences pratiques très intéressantes.

Graphe orienté acyclique

En théorie des graphes, un graphe orienté acyclique est un graphe dont les flèches d'un point à l'autre ne possèdent qu'une seule direction et dont les chemins ne forment pas de boucles. Il est donc impossible de partir d'un point puis d'y revenir. La Figure 7

représente un tel graphe orienté acyclique. Toujours à la Figure 7, chaque sommet du graphe représente une variable. Un petit cercle plein représente une variable observée tandis qu'un petit cercle creux représente une variable inobservée, c'est-à-dire une variable pour laquelle nous ne disposons pas de mesure. Les flèches d'un sommet à l'autre représentent les relations causales et leur direction. Il est important de noter que l'absence de flèches constitue aussi des postulats causaux. À la Figure 7, l'absence de flèche de Y vers D affirme que Y n'influence pas D . Un graphe orienté acyclique interprété de cette façon peut être qualifié de graphe causal. L'utilisation d'un graphe orienté acyclique restreint les types de réseaux causaux qu'il est possible d'étudier. Le fait que le graphe soit orienté et acyclique interdit d'étudier les cas de rétroaction directe entre deux variables ou encore les cas de boucles de rétroaction. Malgré cette limitation, les graphes orientés acycliques peuvent représenter les trois types de relations causales de base. Une première relation est la relation de médiation lorsque deux variables sont liées par une troisième variable qui se situe entre les deux premières dans la chaîne causale. Une deuxième relation est la relation de causalité commune lorsque deux variables sont influencées par une même cause. Finalement, il y a la relation de causalité simultanée lorsque deux variables sont la cause d'une troisième variable (Pearl 2009, Chap. 3; Morgan and Winship 2007, Chap. 3). On parle alors de variables de confusion.

Figure 7 : Graphe causal général



Peu importe le type de relation causale étudié, il est nécessaire de contrôler certaines variables afin d'obtenir une estimation sans biais de l'effet. Remarquons que l'ensemble S n'est qu'une composante de ce réseau causal. Que devons-nous faire ? Contrôler toutes les variables de ce réseau alors même que la variable V n'est pas mesurée?

Contrôle des variables

Il n'y a pas consensus parmi les chercheurs quant au choix des variables à contrôler lors d'une étude de l'effet d'une cause. Dans ces circonstances, les chercheurs ont tendance à contrôler pour un grand nombre de variables, voir pour toutes les variables à leur disposition. La théorie de la causalité de Pearl nous apporte deux enseignements précieux. Premièrement, il n'est pas nécessaire de contrôler pour toutes les causes connues et mesurées. Pour un réseau causal le moins complexe, il existe plusieurs combinaisons de causes. Le contrôle des variables de certaines combinaisons est suffisant pour estimer l'effet causal. Le graphe causal devient alors un outil puissant pour identifier les combinaisons de causes suffisantes à l'estimation sans biais de l'effet de la cause d'intérêt. Pearl propose deux critères pour identifier ces combinaisons selon la position des causes dans la chaîne causale. Le deuxième enseignement concerne certaines variables dites de collision (*collider variable*), lesquelles nécessitent un traitement spécial.

Le critère de la porte arrière (*backdoor criteria*)

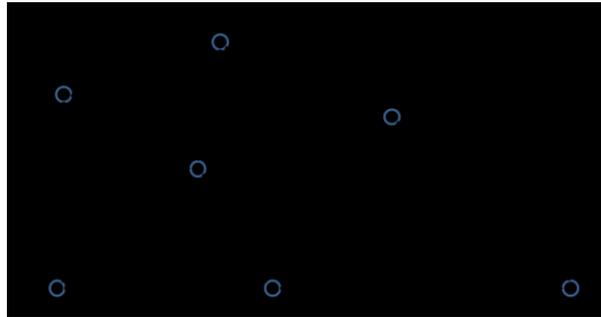
Lorsque les causes à contrôler précèdent la cause d'intérêt, comme c'est le cas pour S à la Figure 7, nous pouvons faire appel au critère de la porte arrière (*backdoor criteria*) pour identifier les combinaisons de causes suffisantes pour l'estimation de l'effet (Pearl 2009, p. 79, 338; Morgan and Winship 2007, p. 61). À la Figure 7, D est la cause principale pour laquelle nous voulons estimer l'effet sur Y . La flèche orientée de D vers Y indique la relation causale. Cependant, D et Y sont aussi liés statistiquement l'un à l'autre par le fait même qu'ils font partie du même réseau causal. Les chercheurs parlent alors de relations fallacieuses. Cette perturbation statistique doit être neutralisée afin d'estimer l'effet de la seule relation causale. Il faut donc contrôler certaines variables du réseau causal, mais lesquelles? Le critère de la porte arrière nous dit qu'il faut bloquer tous les chemins arrière qui lient D et Y . À la Figure 7, nous trouvons un seul chemin arrière de D à Y , soit (D, S, V, F, Y) . Dans ce cas-ci, nous avons trois choix pour le contrôle des variables. Nous pouvons contrôler seulement S , contrôler seulement F ou contrôler S et F . De cette façon on neutralise le bruit statistique et seul l'effet causal demeure.

Remarquons, bien que V fasse parti du réseau causal, Pearl a démontré qu'il n'est pas nécessaire de contrôler pour l'effet de cette variable. Le chercheur sera heureux de l'apprendre puisqu'il ne dispose pas de données sur V (cercle vide). Il s'agit d'un résultat des plus extraordinaires pour la pratique. Si nous disposons d'un réseau causal au début d'une étude, disons au moment de la mise au point du protocole de recherche, il est alors possible de choisir les variables qui seront mesurées selon différents critères. Par exemple, si V est une variable difficile ou coûteuse à mesurer, nous pouvons décider de l'exclure de la cueillette des données.

Le cas des variables de collision (*collider variables*)

La Figure 8 représente un réseau causal pour lequel S est causé simultanément par U et V . Cette fois-ci, il existe deux chemins arrières de D à Y , soit (D, S, V, F, Y) , comme à la Figure 7, et (D, B, U, S, V, F, Y) . Tout comme pour la Figure 7, nous pouvons contrôler seulement F et ainsi bloquer les deux chemins arrière du réseau causal. Mais contrairement à la Figure 7, il n'est pas possible de contrôler seulement S afin de bloquer les deux chemins arrière. Si nous contrôlons seulement S , nous bloquons seulement un des deux chemins, soit (D, S, V, F, Y) . En fait, il a été démontré que le deuxième chemin est déjà bloqué par cette situation de causalité simultanée. Il a été aussi démontré qu'en contrôlant seulement pour S nous débloquons alors le deuxième chemin, soit (D, B, U, S, V, F, Y) . Dans ce cas-ci, si nous tenons à contrôler S , nous devons aussi contrôler B afin que le deuxième chemin soit bloqué (Hernan et Robins, en préparation, chap. 6-8; Morgan et Winship, 2007, p.64-73; Pearl, 2009, p.79-80).

Figure 8 : Graphe causal avec causes simultanées

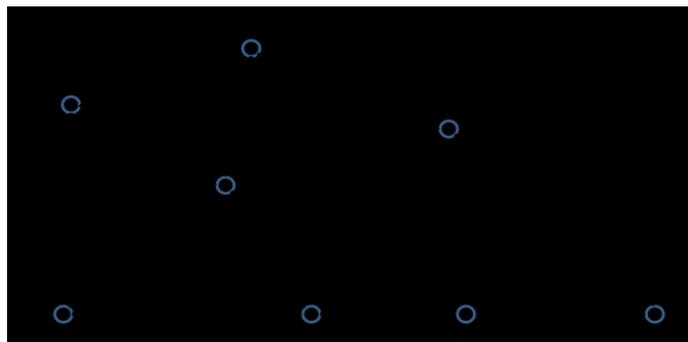


Nous voyons ici l'importance de ne pas contrôler pour toutes les variables à notre disposition et d'être certain que les chemins ouverts par le contrôle des variables qui sont le siège d'un nexus de causes soient fermés par d'autres variables du réseau.

Le critère de la porte d'entrée (*frontdoor criteria*)

Supposons maintenant que les causes B et F ne sont pas mesurées. Ceci est représenté à la Figure 9 par des cercles creux. Il est alors impossible d'appliquer le critère de la porte arrière afin de bloquer tous chemins arrière. Sur les chemins arrière, seul S est mesurée et nous avons vu que S bloque qu'un seul des deux chemins. Cependant, si nous disposons d'une variable qui se situe entre la cause d'intérêt et l'effet, alors nous pouvons faire appel au critère de la porte d'entrée (*frontdoor criteria*) pour identifier les combinaisons de causes suffisantes pour l'estimation de l'effet. À la Figure 9, cette cause est représentée par M . Dans la littérature, M est nommé le mécanisme de D , c'est-à-dire la façon dont l'effet de D se propage jusqu'à Y .

Figure 9 : Graphe causale avec un mécanisme



Le critère de la porte d'entrée est en fait une application répétée du critère de la porte arrière. L'idée est d'estimer dans un premier temps l'effet causal de D sur M , puis l'effet causal de M sur Y , pour finalement combiner les deux effets afin d'obtenir l'effet causal de D sur Y . Si on tente d'appliquer le critère de la porte arrière à l'effet causal de D sur M , on s'aperçoit qu'il n'y a pas de chemins arrière de D à M . Il est alors possible d'estimer cet effet causal sans aucun contrôle statistique. Par contre, en ce qui concerne l'effet causal de M sur Y , il existe deux chemins arrière, soit (D, S, V, F, Y) et (D, B, U, S, V, F, Y) . Nous constatons cependant que D bloque ces deux chemins arrière. Il suffit alors de contrôler D lorsqu'on estime l'effet de M sur Y . Finalement, on multiplie les deux effets afin d'obtenir l'effet de D sur Y (Pearl 2009, p. 81; Morgan and Winship 2007, p. 182-183).

La théorie causale structurale de Pearl est vaste. À l'instar de certains auteurs, nous nous sommes limités qu'à une petite portion de cette théorie, celle qui permet de représenter un réseau causal puis identifier, à l'aide de critères formels les variables qu'il est nécessaire de contrôler afin d'obtenir une estimation sans biais de l'effet causal. Tout comme pour le modèle causal contrefactuel, nous voyons encore une fois, dans cette exposition des graphes causaux de la théorie de Pearl, l'avantage de la théorisation et de la formalisation pour la pratique.

L'identification des variables à contrôler est une première étape de l'analyse de l'effet causal. Une fois identifier ces variables, nous devons choisir les techniques que nous utiliserons pour estimer l'effet causal. Une pratique largement répandue consiste à utiliser l'appariement afin d'équilibrer les deux groupes sur les variables à contrôler, puis à utiliser un modèle de régression adapté à la nature des données. Par exemple, dans notre cas, une fois équilibrés les deux groupes sur les variables identifiées, nous utiliserons un modèle de régression multiniveau pour évaluer l'effet causal parce que nos données sont de nature longitudinale. Les avantages de cette approche sont nombreux : elle est facile à comprendre et à mettre en application; les chercheurs n'ont pas à mettre de côté l'expérience qu'ils ont développée pour la construction de modèles de régression; les variables utilisées lors de

l'appariement n'ont pas à être introduites dans un modèle de régression, ce qui réduit l'impact des problèmes de colinéarité et de forme fonctionnelle du modèle (Ho *et coll.*, 2007; Imai *et coll.*, 2011; Morgan et Winship, 2007, p.90 et 158).

Appendice 3: L'analyse de sensibilité

Les études ne bénéficiant pas de l'assignation aléatoire souffrent d'un grave défaut, même celles qui ont été mené dans des conditions idéales pour ce type d'étude. L'assignation aléatoire garantit que toutes les caractéristiques des individus sont équilibrées en moyenne entre les deux groupes. Un chercheur qui mène une étude observationnelle doit espérer qu'il a mesuré toutes les caractéristiques essentielles à l'analyse de la variable d'intérêt parce que même si ces variables sont équilibrées après un appariement, il n'est pas garanti que les caractéristiques inobservées seront équilibrées, et si ces caractéristiques s'avèrent essentielles à l'analyse, on aura alors affaire à un biais de sélection.

Il existe plusieurs moyens de faire face au biais de sélection. Une première façon est de justifier théoriquement et empiriquement que nous disposons de toutes les variables essentielles à l'analyse. Si nous savons que nous ne disposons pas de toutes ces variables, alors on peut utiliser l'approche des économètres, celle des variables instrumentales. Le défaut de cette approche est qu'elle réintroduit le problème de la modélisation et donc de la forme de la relation entre la variables d'intérêt et les variables-contrôles. Une troisième approche consiste à mener des simulations de Monte-Carlo. Il est possible de reproduire approximativement les caractéristiques d'un ensemble de données, incluant les défauts de celle-ci, puis de simuler la présence de variables inobservées hypothétiques afin de voir comment réagit différents modèles d'analyse de l'effet causal. L'approche ou le modèle qui produit le biais le plus petit est alors utilisé pour analyser les données réelles en notre possession. La plupart des logiciels d'analyse statistique propose des procédures pour générer des variables selon les distributions les plus courantes, ce qui rend relativement aisé la simulation de situations idéales. Ce qui est beaucoup plus difficile est de simuler les défauts de variables réelles, ce qui rend cette approche difficile d'accès. Une quatrième approche est celle des graphes orientés acyclique. Si nous disposons d'une représentation d'un réseau causal, obtenue soit théoriquement et/ou par des études empiriques, il est possible d'analyser ce graphe de telle sorte qu'on détermine l'ensemble minimal de variables nécessaires à l'analyse. De cette façon, même si notre représentation du réseau causal contient une variable dont nous n'avons aucune mesure, si cette variable ne fait pas

partie de l'ensemble minimale, alors elle n'est pas nécessaire à l'analyse (Pearl, 2009; Pearl, 2010, pour un exposé plus accessible)⁵. Une cinquième approche est celle d'estimer la sensibilité d'un résultat à l'existence possible de variables essentielles à l'analyse mais inobservées. C'est cette dernière approche que nous examinons dans la prochaine section.

Il existe plusieurs façons de procéder à une analyse de sensibilité des résultats d'une étude. Une façon simple est d'utiliser plusieurs techniques d'analyse qui divergent quant à leurs postulats d'application et d'examiner l'estimation de l'effet causal ainsi que l'écart-type associé à ces estimations. Si aucune divergence notable n'existe, nous savons alors qu'au moins les données sont robustes à l'absence de certaines conditions d'application. L'utilisation des trois méthodes d'appariement présentées ci-dessus relève de cette approche. Cette approche a des mérites seulement si les techniques examinées n'ont pas tous les mêmes postulats en commun, ce qui est rarement le cas puisque chaque technique est conçue pour répondre à certaines conditions d'application. Par exemple, bien que l'appariement par la distance de Mahalanobis et l'appariement par le score de propension partagent la condition de la présence de la propriété EPBR, la seconde méthode crée automatiquement un biais lorsqu'on utilise plus d'une variable continue lors de l'appariement. Ainsi, si les deux méthodes produisent des résultats semblables, c'est que les données sont robustes à la présence de plusieurs variables continues. Seulement, que faire si les trois méthodes d'appariement partagent un même postulat invérifiable ? Peu importe que l'estimation de l'effet causal soit semblable ou différente entre les trois méthodes, nos conclusions demeurerons incertaines puisque nous ne connaissons pas le comportement de chaque méthode à l'absence de la condition postulée.

Nous avons mentionné que les trois méthodes reposent sur le postulat de l'absence de biais de sélection sur les inobservées. Nous ne savons pas si ce postulat est respecté. Nous pouvons cependant faire l'hypothèse que le postulat n'est pas respecté, c'est-à-dire qu'il existe des variables influentes inobservées, et estimer qu'elle devrait être l'ampleur de

⁵ Un autre résultat critique obtenu par Pearl est l'existence de variables de collision (*collider variables*), lesquelles ne doivent en aucun cas faire partie de l'analyse de l'effet causal même si elles font partie du réseau causal postulé. Ces variables créent automatiquement un biais dans l'analyse.

l'influence de ces variables pour changer la conclusion de nos résultats. Plus l'ampleur nécessaire pour affecter l'analyse est élevée, mais il est probable qu'il existe des variables inobservées pouvant causer un tel effet. C'est l'approche développée par Rosenbaum (Keele, 2010; Rosenbaum, 2010; Rosenbaum, 2002; Rosenbaum, 2005). L'approche de Rosenbaum est spécifiquement conçue pour analyser des données après appariement.

Rosenbaum part de l'idée d'une étude expérimentale idéale. L'assignation aléatoire des cas garantit que les deux groupes sont identiques en moyenne sur l'ensemble de leurs caractéristiques, que celles-ci soient observées ou non. Dans le cas d'une étude observationnelle, nous pouvons l'assimiler à une étude expérimentale seulement si nous avons apparié les cas à l'aide toutes les variables pertinentes. Dans ce cas idéal, les techniques d'inférence basées sur l'assignation aléatoire deviennent valides pour des données observationnelles. Mais il est rare que nous puissions aussi facilement assimiler une étude observationnelle à une étude expérimentale. Rosenbaum définit alors une mesure de l'influence de variables inobservées sur le résultat de l'analyse de l'effet causal de telle manière que celle-ci est égale à un dans le cas d'une étude expérimentale idéale. Toute valeur supérieure à un indique la présence et quantifie l'ampleur de l'influence de variables inobservées.

Le concept central ici est celui d'interchangeabilité (*exchangeability*) et la technique adéquate pour le calcul du nombre d'échanges possibles est la combinatoire, plus précisément le calcul des combinaison sans répétition dans le cas des données assignées aléatoirement et le calcul des arrangement avec répétition dans le cas des données appariées (Keele, 2010, p.3)⁶. L'assignation aléatoire simple nous assure que les cas sont interchangeables, c'est-à-dire que chaque cas a la même probabilité de se retrouver dans le groupe-intervention, ici une probabilité de 0,5. Dans le cas de données appariées, ces probabilités sont estimées par une mesure de distance (Mahalanobis, scores de propension, poids génétiques) et chaque cas a une probabilité estimée différente. Mais plusieurs cas ont une probabilité assez semblable pour être appariés. Ainsi, deux cas qui ont une probabilité

⁶ Keele utilise le terme générique de permutation.

approximative de 0,4 seront appariés et deux cas qui ont une probabilité de 0,8 seront appariés. Autrement dit, dans le cas de l'assignation aléatoire, tous les cas sont interchangeables tandis que dans le cas d'un appariement seuls les cas appariés sont interchangeables. C'est pour cette raison que le calcul combinatoire ne peut être le même dans les deux situations.

Prenons l'exemple suivant. Nous disposons de quatre cas et nous voulons en assigner deux à un groupe-intervention et deux à un groupe-contrôle : [A, B, C, D]. Sous assignation aléatoire, toutes les combinaisons possibles de deux cas recevant le traitement sont au nombre de six : (A, B), (A, C), (A, D), (B, C), (B, D), (C, D). Maintenant, nous disposons de quatre cas, mais cette fois ils ont été appariés : [(A, B), (C, D)]. Tous les arrangements possibles de deux cas recevant le traitement mais n'appartenant pas à la même paire sont au nombre de quatre : (A, C), (A, D), (B, C), (B, D). L'idée de Rosenbaum est d'exploiter le fait que les cas sont interchangeables à l'intérieur de chaque paire. Si l'hypothèse nulle d'aucun effet du traitement est vraie, alors l'estimation de l'effet causal devrait être le même peu importe qu'on utilise l'un ou l'autre des sous-ensembles formé par la combinatoire. Si l'effet causal varie d'un sous-ensemble à l'autre, alors nous devons disposer d'un moyen pour évaluer si cette variation est aléatoire ou causal, autrement dit, il s'agit de savoir si l'effet de traitement est significatif. Nous avons donc besoin d'un test statistique. Prenons un exemple tiré de Keele (Keele, 2010, p.4).

Nous disposons de dix observations appariées. Ceci nous fait cinq paires d'observations. Il y a donc trente-deux sous-ensembles possibles de cinq cas recevant le traitement, mais n'appartenant pas à la même paire. Keele utilise le test des rangs signés de Wilcoxon. Il suffit d'obtenir la différence entre chaque cas-intervention et chaque cas-contrôle (sous l'hypothèse nulle ces différences sont égales à zéro), d'ordonner ces différences en ordre croissant, leur assigner un rang et additionner seulement les rangs dont la différence est positive (l'hypothèse alternative affirme que l'effet de traitement est plus grand que zéro). Dans le cas de Keele, il obtient un test statistique observé sur l'échantillon apparié égal à quinze. Nous calculons ensuite le test statistique pour chacun des trente-deux

sous-ensembles. Ceci constitue alors la distribution de la statistique sous l'hypothèse nulle. Nous comptons finalement le nombre de fois que le test statistique est égal ou supérieur à quinze. Dans l'exemple de Keele, ceci ne se produit qu'une fois et il s'agit de la distribution donnée par l'appariement et qui a servi à calculer le test statistique observé. Nous pouvons maintenant évaluer notre valeur-p exacte et déterminer si nous rejetons ou non l'hypothèse nulle. La valeur-p exacte est le nombre de sous-ensembles qui ont produit un test statistique égal ou supérieur à quinze sur le nombre totale de sous-ensembles. Dans cet exemple, nous obtenons une valeur-p égal à 0,03125 (1/32). Nous pouvons donc rejeter l'hypothèse nulle est affirmer que la différence observée entre les deux groupes n'est probablement pas dû au hasard mai bien plutôt à l'effet du traitement.

Nous venons de procéder à une inférence aléatoire qui suppose que les cas ont été assignés de façon aléatoire dans chaque groupe. Ce type d'inférence, c'est-à-dire la validité de la valeur-p obtenue, est aussi valable si les cas ont été assignés par appariement, mais cette fois-ci on suppose qu'il n'existe aucune variable inobservée pertinente à l'appariement. C'est la sensibilité à la présence possible de variables inobservées qu'évalue la mesure de sensibilité de Rosenbaum.

L'analyse de sensibilité de Rosenbaum est donc fondée sur le raisonnement aléatoire. Rosenbaum définit un paramètre, Γ , comme la mesure du degré de divergence du résultat d'une étude à un résultat qu'on aurait obtenu avec une étude expérimentale. Plus techniquement, Γ est le ratio des rapports de cote des observations :

$$\frac{1}{\Gamma} \leq \frac{\pi_j / (1 - \pi_j)}{\pi_k / (1 - \pi_k)} \leq \Gamma \quad (31)$$

Une assignation aléatoire nous assure que ce ratio est égal à un puisque la probabilité de recevoir le traitement est le même pour tous. Ceci implique alors que Γ est égal à un. Lors d'une étude appariée, ce ratio ne sera égal à un qui si nous avons appariés sur toutes les variables pertinentes, c'est-à-dire qu'aucune variable inobservée n'existe. Ainsi, à $\Gamma = 1$ est associé une valeur-p valide, que ce soit pour une étude expérimentale idéale ou une étude observationnelle idéale, c'est-à-dire pour des études sans défaut,

notamment sans la présence de variables inobservées dans le cas d'études observationnelles. L'analyse de sensibilité consiste alors à choisir des valeurs pour Γ supérieures à un et recalculer les valeurs-p associées. On cherche alors la plus petite valeur de Γ qui changera la conclusion de notre étude. Plus cette valeur est élevée, moins est sensible notre conclusion à la présence d'inobservées. Par exemple, s'il faut $\Gamma = 4$ pour passer d'un effet significatif à un effet non significatif, cela s'interprète de la façon suivante : afin d'attribuer l'effet à une variable inobservée plutôt qu'au traitement, il faudrait que cette variable inobservée provoque une augmentation d'ordre quatre de la propension de recevoir le traitement plutôt que pas. On peut soit conclure qu'il existe une telle variable, mais qu'elle a échappée aux observateurs pendant des décennies, on encore on peut conclure qu'il y ait peu de chance qu'une variable dotée d'une telle force causale puisse exister et dans ce cas on déclare notre étude insensible à la présence d'inobservées et nous retenons nos conclusions préliminaires.

Prenons un autre exemple de Keele (Keele, 2010, p.11-14) pour illustrer ces considérations théoriques. Nous reproduisons dans le Tableau 15 les résultats d'un test de sensibilité. La première colonne contient les valeurs de Γ spécifiées par le chercheur. La deuxième et troisième colonne contiennent la borne inférieure et la borne supérieure des valeur-p calculée pour valeur de Γ . Les valeurs de la borne inférieure sont toujours décroissantes et les valeurs de la borne supérieures sont toujours croissantes. Si le résultat original est non significatif, nous devons alors examiner les bornes inférieures afin de déterminer à quelle valeur de gamma le test devient significatif, ce qui changerait notre conclusion. De même, si le résultat original est significatif, nous devons alors examiner les bornes supérieures afin de déterminer à quelle valeur de gamma le test devient non significatif, ce qui changerait aussi notre conclusion.

**Tableau 15 : Test de sensibilité de Rosenbaum
valeur-p des rangs signés de Wilcoxon**

<i>Gamma</i>	<i>Borne inférieure</i>	<i>Borne supérieure</i>
1.0	0,0346	0,0346
1.1	0,0062	0,1271
1.2	0,0009	0,3000
1.3	0,0001	0,5164
1.4	0,0000	0,7139
1.5	0,0000	0,8539

La première ligne d'un tableau de sensibilité s'interprète toujours de la même façon. Si les données proviennent d'un appariement parfait, alors il n'y a pas d'inobservées et nous obtenons la valeur-p indiquée au tableau. Puisque cette étude a produit un résultat significatif, alors nous examinons les bornes supérieures. Notre conclusion change dès la deuxième ligne, alors que la valeur-p passe de 0,0346 à 0,1271, et bien que Γ n'est augmenté que de dix pourcent. Ceci revient à dire qu'une petite différence entre les deux groupes dans la distribution d'une variable inobservée provoque un changement de conclusion (Keele, 2010, p.14). Il est donc probable qu'une variable importante n'est pas été retenue lors de l'appariement.

L'analyse de sensibilité devrait faire partie intégrante de toute étude observationnelle. Elle fait parti d'une panoplie d'outils à la disposition du chercheur pour l'aider à évaluer la conformité de l'étude à tout un ensemble de postulats. L'analyse de sensibilité de Rosenbaum est spécialement conçue pour évaluer la sensibilité des résultats à la présence de variables inobservées lorsque l'analyse est effectuée sur des données appariées.

Appendice 4: Recherches récentes sur l'appariement

À l'aide d'une recherche sommaire dans Google Scholar, effectuée le 27 août 2013, nous trouvons en 2013 1990 entrées pour « causal analysis », 119 entrées pour « counterfactual model » ou « Rubin model », 695 entrées pour « matching analysis », 2 790 entrées pour « propensity score », 29 entrées pour « mahalanobis matching », 95 entrées pour « genetic matching ». De toute évidence, l'appariement par le score de propension demeure la méthode la plus populaire. Un chercheur très actif en ce domaine est Gary King, Professeur à l'Université de Harvard, concepteur du module MatchIt sous R et de plusieurs autres modules, notamment le module CEM. CEM est l'acronyme pour « Coarsened Exact Matching », un nouvel algorithme d'appariement mis au point par King et ses collaborateurs (Iacus, King et Porro, 2011a; Iacus, King et Porro, 2011b). Nous trouvons 97 entrées pour « coarsened exact matching » en 2013, donc plus que pour l'appariement par la distance de Mahalanobis, autant que l'appariement génétique, mais bien en deçà de l'appariement par le score de propension. Les résultats théoriques, les simulations de Monte Carlo et l'application de l'algorithme à des ensembles de données réels semblent démontrer que cet algorithme serait supérieur aux trois méthodes examinées dans ce mémoire. Appliqué à nos données, l'algorithme ne produit pas de résultats aussi concluants selon les critères retenus. Nous comparons, au **Erreur ! Source du renvoi introuvable.**, les résultats de la méthode CEM avec ceux de l'appariement génétique sous le schème 1:1 avec remise, le schème sous lequel l'appariement génétique a le mieux performé. Ces résultats préliminaires nous confirment la puissance de l'algorithme génétique.

Tableau 16 : Comparaison des méthodes CEM et génétique

	Différences standardisées			Réduction/Augmentation	
	Brut	CEM	Génétique	CEM	Génétique
N	69/181	60/116	69/55		
Adversité	0,0318	0,0338	0,0053	-290,28	83,45
Anxiété	0,3203	0,0392	0,0084	92,91	97,38
Prosocialité	0,1034	-0,1384	-0,0129	70,30	87,56
Bataille	-0,0303	-0,1118	-0,0094	-139,94	68,90
Opposition	0,0972	0,0722	0,0075	61,46	92,26
Inattention	0,0454	-0,1128	0,0015	-78,47	96,65
Hyperactivité	0,1525	0,0174	-0,0108	90,41	92,91
Doss. crim. - Père	-0,3189	0,0000	0,0000	100,00	100,00
Doss. crim. - Mère	-0,2862	0,0000	0,0000	100,00	100,00
Nbr var. déséq./ Nbr var. augment. déséq.	5	3	0	3	0
Moy. de déséq./% moy. de réduct. déséq.	0,1540	0,0584	0,0062	71,08	91,01

Appendice 5: Syntaxes de programmation

```

# François Maurice, B. Sc., A. Stat.
# Candidat à la maîtrise
# Département de sociologie
# Université de Montréal

# 2011-2013

## CHARGER LES MODULES
# CHARGER LE MODULE MATCHIT POUR LES APPARIEMENTS
library(MatchIt)
library(cem)

# CHARGER LE MODULE MMLE POUR L'ANALYSE CAUSALE MULTINIVEAUX
library(nlme)
library(lme4)

# CHARGER LES MODULES MITOOLS ET AMELIA POUR L'IMPUTATION MULTIPLE
library(Amelia)
library(mitools)

# CHARGER LE MODULE HMISC ET FOREIGN POUR IMPORTATION/EXPORTATION DE FICHIERS
library(Hmisc)

# CHARGER LE MODULE RESHAPE POUR LA RETRUCTURATION DES FICHIERS
library(reshape)

## SPÉCIFIER LES CHEMINS D'ACCÈS
# Dossiers des données construites
chemin.donnees.construites <-
paste("D:\\dossiers\\SCV\\Client76072_Francois_Maurice\\scvadmin76072\\Depot77474_Recherches
\\Racine\\0001_Memoire\\Programmation\\DonneesConstruites\\")

# Dossier des résultats
chemin.resultats <-
paste("D:\\dossiers\\SCV\\Client76072_Francois_Maurice\\scvadmin76072\\Depot77474_Recherches
\\Racine\\0001_Memoire\\Programmation\\Resultats\\")

# Dossier des résultats graphiques
chemin.resultats.graphiques <-
paste("D:\\dossiers\\SCV\\Client76072_Francois_Maurice\\scvadmin76072\\Depot77474_Recherches
\\Racine\\0001_Memoire\\Programmation\\Resultats\\Figures\\")

## OUVERTURE DU FICHIER .POR DE SPSS
# Ouverture

```

```

data.from.spss <- spss.get(file = paste(chemin.donnees.construites,
"DEGEPPaquinOlacours_R_Imput_Covar.sav", sep=""), use.value.labels=FALSE)
# Transformation des nom de variables en majuscules
names(data.from.spss) <- toupper(names(data.from.spss))

## SÉLECTION DES VARIABLES
var.tx <- c("CINTER2G")
var.id <- c("ID")
var.effet <- c("QASBAT2", "QASBAT3", "QASBAT4", "QASBAT5", "QASBAT6", "QASBAT7", "QASBAT8")
var.appariement <- c("ADVERS84", "QCP84AX3", "QCP84PS2", "QCP84BAT", "QCP84OP", "QCP84IN2",
"QCP84HYP", "DOSSP", "DOSSM")
var.nominales.imput <- c("DOSSP", "DOSSM")
# Création d'une chaine de caractères contenant les noms des variables qui apparaîtront dans
les fichiers d'exportation
var.appariement.nom <- c("Adversité", "Anxiété", "Pro-socialité", "Bataille", "Opposition",
"Inattention", "Hyperactivité", "Doss. crim. - Père", "Doss. crim. - Mère", "Nbr var.
déséquil. / Nbr var. augment. déséquil.", "Moy. stand. de déséquil. / % moy. de réduct.
déséquil.")

## CRÉATION DES DIFFÉRENTS MODÈLES D'ANALYSE
# Modèle pour les appariements
formula.matching <- as.formula ("CINTER2G ~ QCP84AX3 + QCP84PS2 + QCP84BAT + QCP84OP +
QCP84IN2 + QCP84HYP + ADVERS84 + DOSSM + DOSSP")

# Modèles pour les analyses causales (modèles multiniveaux qui utilisent lme())
formula.fixed <- as.formula ("effet ~ CINTER2G + T + I(T*T) + I(T*T*T)")
formula.random <- as.formula ("~T + I(T*T)|ID.factor")
formula.fixed.covar <- as.formula ("effet ~ CINTER2G + QCP84AX3 + QCP84PS2 + QCP84BAT +
QCP84OP + QCP84IN2 + QCP84HYP + ADVERS84 + DOSSM + DOSSP + T + I(T*T) + I(T*T*T)")

## CRÉATION DE DEUX TABLES DE DONNÉES
# Extraire les variables servant à l'appariement
data.for.matching <- data.from.spss[c(var.id, var.tx, var.appariement)]

# Extraire les effets
# Nécessaire pour fusionner les outcomes aux tables de données générées par les
appariements afin d'obtenir des tables de données pour les analyses causales
data.outcome <- data.from.spss[c(var.id, var.effet)]

## CONTRÔLE DE L'IMPUTATION
# Activation de l'imputation (1=oui, 0=non)
imput.active <- 0
# Nombre d'imputations (Pour un maximum de puissance, m <- 10)
m <- 10

## CONFIGURATION DE L'APPARIEMENT GÉNÉTIQUE
# Pour un maximum de puissance, popsize <- 1000, maxgeneration <- 100, waitgenerations <-10
popsize <- 1000
maxgenerations <- 100
waitgenerations <- 4

## CRÉATION DES FICHIERS .CVS POUR LA COLLECTE DES RÉSULTATS

```

```

write.table(c("", "Analyse de l'équilibrage des variables - Tous les résultats"), file =
paste(chemin.resultats,"balance.csv", sep=""), sep = ";", dec="," , na="", col.names = FALSE,
row.names= FALSE, qmethod = "double")
write.table(c("", "Analyse de la réduction du déséquilibrage - Tous les résultats"), file =
paste(chemin.resultats,"reduction.csv", sep=""), sep = ";", dec="," , na="", col.names =
FALSE, row.names= FALSE, qmethod = "double")
write.table(c("", "Analyse causale - Tous les résultats"), file =
paste(chemin.resultats,"causal.csv", sep=""), sep = ";", dec="," , na="", col.names = FALSE,
row.names= FALSE, qmethod = "double")
write.table(c("", "Analyse causale avec imputation - Tous les résultats"), file =
paste(chemin.resultats,"causal_imput.csv", sep=""), sep = ";", dec="," , na="", col.names =
FALSE, row.names= FALSE, qmethod = "double")

write.table(c("", "Analyse de l'équilibrage des variables et de la réduction du
déséquilibrage"), file = paste(chemin.resultats,"results_bal.csv", sep=""), sep = ";",
dec="," , na="", col.names = FALSE, row.names= FALSE, qmethod = "double")
write.table(c("", "Analyse causale"), file = paste(chemin.resultats,"results_causal.csv",
sep=""), sep = ";", dec="," , na="", col.names = FALSE, row.names= FALSE, qmethod = "double")
write.table(c("", "Analyse causale avec imputation"), file =
paste(chemin.resultats,"results_causal_imput.csv", sep=""), sep = ";", dec="," , na="",
col.names = FALSE, row.names= FALSE, qmethod = "double")

## CRÉATION DE 4 BOUCLES DE PROGRAMMATION, UNE BOUCLE PAR SCHEMA D'APPARIEMENT
for (i in 1:4){

  # 1:1 SANS REMISE
  if (i==1) ratio <- 1
  if (i==1) replace <- FALSE

  # 1:1 AVEC REMISE
  if (i==2) ratio <- 1
  if (i==2) replace <- TRUE

  # 2:1 SANS REMISE
  if (i==3) ratio <- 2
  if (i==3) replace <- FALSE

  # 2:1 AVEC REMISE
  if (i==4) ratio <- 2
  if (i==4) replace <- TRUE

#####
## APPARIEMENT ##
#####

# Distance de Mahalanobis
match.mahalanobis <- matchit(formula = formula.matching, data = data.for.matching,
distance="mahalanobis", ratio=ratio, replace=replace)

# Score de propension

```

```

match.score <- matchit(formula = formula.matching, data = data.for.matching,
distance="logit", method="nearest", ratio=ratio, replace=replace)

# Génétique
match.genetic <- matchit(formula = formula.matching, data = data.for.matching,
method="genetic", ratio=ratio, replace=replace, pop.size = popsize,
max.generations=maxgenerations, wait.generations=waitgenerations, int.seed = 3818, unif.seed
= 3527)

#####
### ANALYSES CAUSALES ###
#####

## SANS APPARIEMENT, SANS COVARIABLE
# FUSIONNIER EN UN SEUL FICHIER TOUTES LES VARIABLES (OUTCOMES ET COVARAIBLES)
# Remarquons que le fichier de données data.for.matching est le fichiers pré-appariement,
i.e. qu'il contient les 250 cas
data.sans.appariement <- merge(data.for.matching, data.outcome, by=var.id)

# RESTRUCTURATION EN FORMAT LONG POUR L'ANALYSE MULTINIVEAU
matched.sans.appariement.long <- melt(data.sans.appariement, id=c(var.id, var.tx,
var.appariement))

# RENOMMER ET CHANGER LE TYPE DE CERTAINES VARIABLES
matched.sans.appariement.long[["TIME"]] <- matched.sans.appariement.long[["variable"]]
matched.sans.appariement.long[["T"]] <- as.numeric(matched.sans.appariement.long[["TIME"]])
matched.sans.appariement.long[["effet"]] <- matched.sans.appariement.long[["value"]]
matched.sans.appariement.long[["ID.factor"]] <-
as.factor(matched.sans.appariement.long[[var.id]])
# RÉORDONNER LES VARIABLES
matched.sans.appariement.long <- matched.sans.appariement.long[c(var.id, "ID.factor",
"TIME", "T", "effet", var.tx, var.appariement)]
# TRIER LES OBSERVATIONS
order.index <- order(matched.sans.appariement.long[[var.id]],
matched.sans.appariement.long[["T"]])
matched.sans.appariement.long <- matched.sans.appariement.long[order.index,]

# Estimation de l'effet causal
lm.out <- summary(lme(fixed=formula.fixed, random=formula.random,
data=matched.sans.appariement.long, na.action=na.omit, method="REML",
control=list(opt="optim")))
resultat.sans.appariement <- as.data.frame(lm.out[["tTable"]])
resultat.sans.appariement <- resultat.sans.appariement[c("Value", "Std.Error")]
names(resultat.sans.appariement) <- c("Coeff.", "E.S.")
resultat.sans.appariement <- transform(resultat.sans.appariement, B.I.=Coeff. - 1.96*E.S.,
B.S.=Coeff. + 1.96*E.S.)

# Estimation de l'effet causal avec imputation multiple
if (imput.active==1){
  b.out<-NULL
  se.out<-NULL
}

```

```

a.out <- amelia(matched.sans.appariement.long, m=m, ts="T", polytime=3, cs=var.id,
noms=var.nominales.imput, idvars=c("ID.factor", "TIME"))
for(j in 1:m) {
  lm.out <- lme(fixed=formula.fixed, random=formula.random, data =
a.out$imputations[[j]], na.action=na.omit, method="REML", control=list(opt="optim"))
  b.out <- rbind(se.out, summary(lm.out)[["tTable"]][,1])
  se.out <- rbind(se.out, summary(lm.out)[["tTable"]][,2])
}
resultats.Rubin <- mi.meld(q = b.out, se = se.out)
resultat.sans.appariement.imput <- cbind(resultats.Rubin[["q.mi"]][1,],
resultats.Rubin[["se.mi"]][1,])
resultat.sans.appariement.imput <- as.data.frame(resultat.sans.appariement.imput)
names(resultat.sans.appariement.imput) <- c("Coeff.", "E.S.")
resultat.sans.appariement.imput <- transform(resultat.sans.appariement.imput,
B.I.=Coeff. - 1.96*E.S., B.S.=Coeff. + 1.96*E.S.)
}

## SANS APPARIEMENT, AVEC COVARIABLES
# FUSIONNIER EN UN SEUL FICHER TOUTES LES VARIABLES (OUTCOMES ET COVARAIBLES)
# Remarquons que le fichier de données data.for.matching est le fichiers pré-appariement,
i.e. qu'il contient les 250 cas
data.avec.covariables <- merge(data.for.matching, data.outcome, by=var.id)

# RESTRUCTURATION EN FORMAT LONG POUR L'ANALYSE MULTINIVEAU
matched.avec.covariables.long <- melt(data.avec.covariables, id=c(var.id, var.tx,
var.appariement))

# RENOMMER ET CHANGER LE TYPE DE CERTAINES VARIABLES
matched.avec.covariables.long[["TIME"]] <- matched.avec.covariables.long[["variable"]]
matched.avec.covariables.long[["T"]] <- as.numeric(matched.avec.covariables.long[["TIME"]])
matched.avec.covariables.long[["effet"]] <- matched.avec.covariables.long[["value"]]
matched.avec.covariables.long[["ID.factor"]] <-
as.factor(matched.avec.covariables.long[[var.id]])
# RÉORDONNER LES VARIABLES
matched.avec.covariables.long <- matched.avec.covariables.long[c(var.id, "ID.factor",
"TIME", "T", "effet", var.tx, var.appariement)]
# TRIER LES OBSERVATIONS
order.index <- order(matched.avec.covariables.long[[var.id]],
matched.avec.covariables.long[["T"]])
matched.avec.covariables.long <- matched.avec.covariables.long[order.index,]

# Estimation de l'effet causal
lm.out <- summary(lme(fixed=formula.fixed.covar, random=formula.random,
data=matched.avec.covariables.long, na.action=na.omit, method="REML",
control=list(opt="optim")))
resultat.avec.covariables <- as.data.frame(lm.out[["tTable"]])
resultat.avec.covariables <- resultat.avec.covariables[c("Value", "Std.Error")]
names(resultat.avec.covariables) <- c("Coeff.", "E.S.")
resultat.avec.covariables <- transform(resultat.avec.covariables, B.I.=Coeff. - 1.96*E.S.,
B.S.=Coeff. + 1.96*E.S.)

# Estimation de l'effet causal avec imputation multiple
if (imput.active==1){

```

```

b.out<-NULL
se.out<-NULL
a.out <- amelia(matched.avec.covariables.long, m=m, ts="T", polytime=3, cs=var.id,
noms=var.nominales.imput, idvars=c("ID.factor", "TIME"))
for(j in 1:m) {
  lm.out <- lme(fixed=formula.fixed.covar, random=formula.random, data =
a.out$imputations[[j]], na.action=na.omit, method="REML", control=list(opt="optim"))
  b.out <- rbind(se.out, summary(lm.out)[["tTable"]][,1])
  se.out <- rbind(se.out, summary(lm.out)[["tTable"]][,2])
}
resultats.Rubin <- mi.meld(q = b.out, se = se.out)
resultat.avec.covariables.imput <- cbind(resultats.Rubin[["q.mi"]][1,],
resultats.Rubin[["se.mi"]][1,])
resultat.avec.covariables.imput <- as.data.frame(resultat.avec.covariables.imput)
names(resultat.avec.covariables.imput) <- c("Coeff.", "E.S.")
resultat.avec.covariables.imput <- transform(resultat.avec.covariables.imput,
B.I.=Coeff. - 1.96*E.S., B.S.=Coeff. + 1.96*E.S.)
}

## MAHALANOBIS
# EXTRACTION DES DONNÉES DE L'OBJET matchit()
matched.mahalanobis <- match.data(match.mahalanobis)

# FUSIONNIER EN UN SEUL FICHIER TOUTES LES VARIABLES (OUTCOMES ET COVARAIBLES)
matched.mahalanobis <- merge(matched.mahalanobis, data.outcome, by=var.id)

# RESTRUCTURATION EN FORMAT LONG POUR L'ANALYSE MULTINIVEAU
matched.mahalanobis.long <- melt(matched.mahalanobis, id=c(var.id, var.tx, var.appariement,
"distance", "weights"))

# RENOMMER ET CHANGER LE TYPE DE CERTAINES VARIABLES
matched.mahalanobis.long[["TIME"]] <- matched.mahalanobis.long[["variable"]]
matched.mahalanobis.long[["T"]] <- as.numeric(matched.mahalanobis.long[["TIME"]])
matched.mahalanobis.long[["effet"]] <- matched.mahalanobis.long[["value"]]
matched.mahalanobis.long$ID.factor <- as.factor(matched.mahalanobis.long[[var.id]])
# RÉORDONNER LES VARIABLES
matched.mahalanobis.long <- matched.mahalanobis.long[c(var.id, "ID.factor", "TIME", "T",
"effet", var.tx, var.appariement, "weights")]
# TRIER LES OBSERVATIONS
order.index <- order(matched.mahalanobis.long[[var.id]], matched.mahalanobis.long[["T"]])
matched.mahalanobis.long <- matched.mahalanobis.long[order.index,]

# Estimation de l'effet causal
lm.out <- summary(lme(fixed=formula.fixed, random=formula.random,
data=matched.mahalanobis.long, na.action=na.omit, method="REML", control=list(opt="optim"),
weights=~weights))
resultat.mahalanobis <- as.data.frame(lm.out[["tTable"]])
resultat.mahalanobis <- resultat.mahalanobis[c("Value", "Std.Error")]
names(resultat.mahalanobis) <- c("Coeff.", "E.S.")
resultat.mahalanobis <- transform(resultat.mahalanobis, B.I.=Coeff. - 1.96*E.S., B.S.=Coeff.
+ 1.96*E.S.)

```

```

# Estimation de l'effet causal avec imputation multiple
if (imput.active==1){
  b.out<-NULL
  se.out<-NULL
  if (i==1 | i==3){
    a.out <- amelia(matched.mahalanobis.long, m=m, ts="T", polytime=3, cs=var.id,
noms=var.nominales.imput, idvars=c("ID.factor", "TIME", "weights"))
  }else
    a.out <- amelia(matched.mahalanobis.long, m=m, ts="T", polytime=3, cs=var.id,
noms=var.nominales.imput, idvars=c("ID.factor", "TIME"))

  for(j in 1:m) {
    lm.out <- lme(fixed=formula.fixed, random=formula.random, data=a.out$imputations[[j]],
na.action=na.omit, method="REML", control=list(opt="optim"), weights=~weights)
    b.out <- rbind(se.out, summary(lm.out)[["tTable"]][,1])
    se.out <- rbind(se.out, summary(lm.out)[["tTable"]][,2])
  }
  resultats.Rubin <- mi.meld(q = b.out, se = se.out)
  resultat.mahalanobis.imput <- cbind(resultats.Rubin[["q.mi"]][1,],
resultats.Rubin[["se.mi"]][1,])
  resultat.mahalanobis.imput <- as.data.frame(resultat.mahalanobis.imput)
  names(resultat.mahalanobis.imput) <- c("Coeff.", "E.S.")
  resultat.mahalanobis.imput <- transform(resultat.mahalanobis.imput, B.I.=Coeff. -
1.96*E.S., B.S.=Coeff. + 1.96*E.S.)
}

## SCORE
# EXTRACTRION DES DONNÉES DE L'OBJET matchit()
matched.score <- match.data(match.score)

# FUSIONNIER EN UN SEUL FICHER TOUTES LES VARIABLES (OUTCOMES ET COVARAIBLES)
matched.score <- merge(matched.score, data.outcome, by=var.id)

# RESTRUCTURATION EN FORMAT LONG POUR L'ANALYSE MULTINIVEAU
matched.score.long <- melt(matched.score, id=c(var.id, var.tx, var.appariement, "distance",
"weights"))

# RENOMMER ET CHANGER LE TYPE DE CERTAINES VARIABLES
matched.score.long[["TIME"]] <- matched.score.long[["variable"]]
matched.score.long[["T"]] <- as.numeric(matched.score.long[["TIME"]])
matched.score.long[["effet"]] <- matched.score.long[["value"]]
matched.score.long$ID.factor <- as.factor(matched.score.long[[var.id]])
# RÉORDONNER LES VARIABLES
matched.score.long <- matched.score.long[c(var.id, "ID.factor", "TIME", "T", "effet",
var.tx, var.appariement, "weights")]
# TRIER LES OBSERVATIONS
order.index <- order(matched.score.long[[var.id]], matched.score.long[["T"]])
matched.score.long <- matched.score.long[order.index,]

# Estimation de l'effet causal
lm.out <- summary(lme(fixed=formula.fixed, random=formula.random, data=matched.score.long,
na.action=na.omit, method="REML", control=list(opt="optim"), weights=~weights))

```

```

resultat.score <- as.data.frame(lm.out[["tTable"]])
resultat.score <- resultat.score[c("Value", "Std.Error")]
names(resultat.score) <- c("Coeff.", "E.S.")
resultat.score <- transform(resultat.score, B.I.=Coeff. - 1.96*E.S., B.S.=Coeff. +
1.96*E.S.)

# Estimation de l'effet causal avec imputation multiple
if (imput.active==1){
  b.out<-NULL
  se.out<-NULL
  if (i==1 | i==3){
    a.out <- amelia(matched.score.long, m=m, ts="T", polytime=3, cs=var.id,
noms=var.nominales.imput, idvars=c("ID.factor", "TIME", "weights"))
  }else
    a.out <- amelia(matched.score.long, m=m, ts="T", polytime=3, cs=var.id,
noms=var.nominales.imput, idvars=c("ID.factor", "TIME"))

  for(j in 1:m) {
    lm.out <- lme(fixed=formula.fixed, random=formula.random, data=a.out$imputations[[j]],
na.action=na.omit, method="REML", control=list(opt="optim"), weights=~weights)
    b.out <- rbind(se.out, summary(lm.out)[["tTable"]][,1])
    se.out <- rbind(se.out, summary(lm.out)[["tTable"]][,2])
  }
  resultats.Rubin <- mi.meld(q = b.out, se = se.out)
  resultat.score.imput <- cbind(resultats.Rubin[["q.mi"]][1,],
resultats.Rubin[["se.mi"]][1,])
  resultat.score.imput <- as.data.frame(resultat.score.imput)
  names(resultat.score.imput) <- c("Coeff.", "E.S.")
  resultat.score.imput <- transform(resultat.score.imput, B.I.=Coeff. - 1.96*E.S.,
B.S.=Coeff. + 1.96*E.S.)
}

## GENETIC
# EXTRACTRION DES DONNÉES DE L'OBJET matchit()
matched.genetic <- match.data(match.genetic)

# FUSIONNIER EN UN SEUL FICHIER TOUTES LES VARIABLES (OUTCOMES ET COVARAIBLES)
matched.genetic <- merge(matched.genetic, data.outcome, by=var.id)

# RESTRUCTURATION EN FORMAT LONG POUR L'ANALYSE MULTINIVEAU
matched.genetic.long <- melt(matched.genetic, id=c(var.id, var.tx, var.appariement,
"distance", "weights"))

# RENOMMER ET CHANGER LE TYPE DE CERTAINES VARIABLES
matched.genetic.long[["TIME"]] <- matched.genetic.long[["variable"]]
matched.genetic.long[["T"]] <- as.numeric(matched.genetic.long[["TIME"]])
matched.genetic.long[["effet"]] <- matched.genetic.long[["value"]]
matched.genetic.long$ID.factor <- as.factor(matched.genetic.long[[var.id]])
# RÉORDONNER LES VARIABLES
matched.genetic.long <- matched.genetic.long[c(var.id, "ID.factor", "TIME", "T", "effet",
var.tx, var.appariement, "weights")]
# TRIER LES OBSERVATIONS

```

```

order.index <- order(matched.genetic.long[[var.id]], matched.genetic.long[["T"]])
matched.genetic.long <- matched.genetic.long[order.index,]

# Estimation de l'effet causal
lm.out <- summary(lme(fixed=formula.fixed, random=formula.random, data=matched.genetic.long,
na.action=na.omit, method="REML", control=list(opt="optim"), weights=~weights))
resultat.genetic <- as.data.frame(lm.out[["tTable"]])
resultat.genetic <- resultat.genetic[c("Value", "Std.Error")]
names(resultat.genetic) <- c("Coeff.", "E.S.")
resultat.genetic <- transform(resultat.genetic, B.I.=Coeff. - 1.96*E.S., B.S.=Coeff. +
1.96*E.S.)

# Estimation de l'effet causal avec imputation multiple
if (imput.active==1){
  b.out<-NULL
  se.out<-NULL

  if (i==1 | i==3){
    a.out <- amelia(matched.genetic.long, m=m, ts="T", polytime=3, cs=var.id,
noms=var.nominales.imput, idvars=c("ID.factor", "TIME", "weights"))
  }else
    a.out <- amelia(matched.genetic.long, m=m, ts="T", polytime=3, cs=var.id,
noms=var.nominales.imput, idvars=c("ID.factor", "TIME"))

  for(j in 1:m) {
    lm.out <- lme(fixed=formula.fixed, random=formula.random, data=a.out$imputations[[j]],
na.action=na.omit, method="REML", control=list(opt="optim"), weights=~weights)
    b.out <- rbind(se.out, summary(lm.out)[["tTable"]][,1])
    se.out <- rbind(se.out, summary(lm.out)[["tTable"]][,2])
  }
  resultats.Rubin <- mi.meld(q = b.out, se = se.out)
  resultat.genetic.imput <- cbind(resultats.Rubin[["q.mi"]][1,],
resultats.Rubin[["se.mi"]][1,])
  resultat.genetic.imput <- as.data.frame(resultat.genetic.imput)
  names(resultat.genetic.imput) <- c("Coeff.", "E.S.")
  resultat.genetic.imput <- transform(resultat.genetic.imput, B.I.=Coeff. - 1.96*E.S.,
B.S.=Coeff. + 1.96*E.S.)
}

#####
####      EXPORTATION DES RÉSULTATS POUR LA CONSTRUCTION DES TABLEAUX      #####
#####

## MANIPULATIONS DES DONNÉES AVANT L'EXPORTATION

#EXTRAIRACON DES MOYENNES STANDARDISÉES
summary.score <- summary(object=match.score, standardize=TRUE)
summary.mahalanobis <- summary(object=match.mahalanobis, standardize=TRUE)
summary.genetic <- summary(object=match.genetic, standardize=TRUE)

# SUPPRIMER L'OBSERVATION 'DISTANCE' DE CHAQUE ENSEMBLE DE DONNÉES DANS CHAQUE OBJET
'SUMMARY'
summary.score[["sum.all"]] <- summary.score$sum.all[ var.appariement, ]

```

```
summary.mahalanobis[["sum.all"]] <- summary.mahalanobis$sum.all[ var.appariement, ]
summary.genetic[["sum.all"]] <- summary.genetic$sum.all[ var.appariement, ]

summary.score[["sum.matched"]] <- summary.score$sum.matched[ var.appariement, ]
summary.mahalanobis[["sum.matched"]] <- summary.mahalanobis$sum.matched[ var.appariement, ]
summary.genetic[["sum.matched"]] <- summary.genetic$sum.matched[ var.appariement, ]

summary.score[["reduction"]] <- summary.score$reduction[ var.appariement, ]
summary.mahalanobis[["reduction"]] <- summary.mahalanobis$reduction[ var.appariement, ]
summary.genetic[["reduction"]] <- summary.genetic$reduction[ var.appariement, ]
```

AGRÉGER LES DONNÉES

CRÉATION DE DEUX FONCTIONS POUR COMPTER LE NOMBRE DE VARIABLES

DONT LA VALEUR SE TROUVE AU-DESSUS DE NOTRE CRITÈRE ET

COMPTER LE NOMBRE DE VARIABLES DONT LE DÉSÉQUILIBRE A AUGMENTÉ

```
count.over <- function (x) sum(x > 0.10)
count.unbal <- function (x) sum(x < 0)
count.obs <- nrow(summary.mahalanobis[["sum.all"]])
```

MAHALANOBIS - COMPTER LES VARIABLES

```
summary.mahalanobis[["sum.all"]] <- rbind.data.frame(summary.mahalanobis[["sum.all"]],
sapply( abs(summary.mahalanobis[["sum.all"]]), count.over))
summary.mahalanobis[["sum.matched"]] <-
rbind.data.frame(summary.mahalanobis[["sum.matched"]], sapply(
abs(summary.mahalanobis[["sum.matched"]]), count.over))
summary.mahalanobis[["reduction"]] <- rbind.data.frame(summary.mahalanobis[["reduction"]],
sapply(summary.mahalanobis[["reduction"]], count.unbal))
```

MAHALANOBIS - OBTENIR LES MOYENNES GÉNÉRALES DES MOYENNES STANDARDISÉES ET DES POURCENTAGES DE RÉDUCTION

```
summary.mahalanobis[["sum.all"]] <- rbind.data.frame(summary.mahalanobis[["sum.all"]],
sapply( abs(summary.mahalanobis[["sum.all"]][1:count.obs,]), mean))
summary.mahalanobis[["sum.matched"]] <-
rbind.data.frame(summary.mahalanobis[["sum.matched"]], sapply(
abs(summary.mahalanobis[["sum.matched"]][1:count.obs,]), mean))
summary.mahalanobis[["reduction"]] <- rbind.data.frame(summary.mahalanobis[["reduction"]],
sapply( summary.mahalanobis[["reduction"]][1:count.obs,]), mean))
```

SCORE DE PROPENSION - COMPTER LES VARIABLES

```
summary.score[["sum.all"]] <- rbind.data.frame(summary.score[["sum.all"]], sapply(
abs(summary.score[["sum.all"]]), count.over))
summary.score[["sum.matched"]] <- rbind.data.frame(summary.score[["sum.matched"]], sapply(
abs(summary.score[["sum.matched"]]), count.over))
summary.score[["reduction"]] <- rbind.data.frame(summary.score[["reduction"]],
sapply(summary.score[["reduction"]], count.unbal))
```

SCORE DE PROPENSION - OBTENIR LES MOYENNES GÉNÉRALES DES MOYENNES STANDARDISÉES ET DES POURCENTAGES DE RÉDUCTION

```
summary.score[["sum.all"]] <- rbind.data.frame(summary.score[["sum.all"]], sapply(
abs(summary.score[["sum.all"]])[1:count.obs,], mean))
summary.score[["sum.matched"]] <- rbind.data.frame(summary.score[["sum.matched"]], sapply(
abs(summary.score[["sum.matched"]])[1:count.obs,], mean))
```

```

summary.score[["reduction"]] <- rbind.data.frame(summary.score[["reduction"]], sapply(
summary.score[["reduction"]][1:count.obs,], mean))

# GÉNÉTIQUE - COMPTER LES VARIABLES
summary.genetic[["sum.all"]] <- rbind.data.frame(summary.genetic[["sum.all"]], sapply(
abs(summary.genetic[["sum.all"]]), count.over))
summary.genetic[["sum.matched"]] <- rbind.data.frame(summary.genetic[["sum.matched"]],
sapply(abs(summary.genetic[["sum.matched"]]), count.over))
summary.genetic[["reduction"]] <- rbind.data.frame(summary.genetic[["reduction"]],
sapply(summary.genetic[["reduction"]], count.unbal))

# GÉNÉTIQUE - OBTENIR LES MOYENNES GÉNÉRALES DES MOYENNES STANDARDISÉES ET DES POURCENTAGES
DE RÉDUCTION
summary.genetic[["sum.all"]] <- rbind.data.frame(summary.genetic[["sum.all"]], sapply(
abs(summary.genetic[["sum.all"]])[1:count.obs,], mean))
summary.genetic[["sum.matched"]] <- rbind.data.frame(summary.genetic[["sum.matched"]],
sapply(abs(summary.genetic[["sum.matched"]][1:count.obs,]), mean))
summary.genetic[["reduction"]] <- rbind.data.frame(summary.genetic[["reduction"]], sapply(
summary.genetic[["reduction"]][1:count.obs,], mean))

## CRÉER DES NOMS POUR LES OBERVATION DE CHAQUE ENSEMBLE DE DONNÉES DANS CHAQUE OBJET
'SUMMARY'
row.names(summary.score[["sum.all"]]) <- var.appariement.nom
row.names(summary.mahalanobis[["sum.all"]]) <- var.appariement.nom
row.names(summary.genetic[["sum.all"]]) <- var.appariement.nom

row.names(summary.score[["sum.matched"]]) <- var.appariement.nom
row.names(summary.mahalanobis[["sum.matched"]]) <- var.appariement.nom
row.names(summary.genetic[["sum.matched"]]) <- var.appariement.nom

row.names(summary.score[["reduction"]]) <- var.appariement.nom
row.names(summary.mahalanobis[["reduction"]]) <- var.appariement.nom
row.names(summary.genetic[["reduction"]]) <- var.appariement.nom

## CRÉATION DES EN-TÊTES POUR CHAQUE SCÈME D'APPARIEMENT
if (i==1) matching.scheme <- c(" ", "Appariement 1:1 sans remise")
if (i==2) matching.scheme <- c(" ", "Appariement 1:1 avec remise")
if (i==3) matching.scheme <- c(" ", "Appariement 2:1 sans remise")
if (i==4) matching.scheme <- c(" ", "Appariement 2:1 avec remise")

write.table(matching.scheme, file = paste(chemin.resultats,"balance.csv", sep=""), sep =
";", dec=".", na="", col.names = FALSE, row.names = FALSE, qmethod = "double", append=TRUE)
write.table(matching.scheme, file = paste(chemin.resultats,"reduction.csv", sep=""), sep =
";", dec=".", na="", col.names = FALSE, row.names = FALSE, qmethod = "double", append=TRUE)
write.table(matching.scheme, file = paste(chemin.resultats,"causal.csv", sep=""), sep = ";",
dec=".", na="", col.names = FALSE, row.names = FALSE, qmethod = "double", append=TRUE)
write.table(matching.scheme, file = paste(chemin.resultats,"causal_imput.csv", sep=""), sep
= ";", dec=".", na="", col.names = FALSE, row.names = FALSE, qmethod = "double",
append=TRUE)

write.table(matching.scheme, file = paste(chemin.resultats,"results_bal.csv", sep=""), sep =
";", dec=".", na="", col.names = FALSE, row.names = FALSE, qmethod = "double", append=TRUE)

```

```

write.table(matching.scheme, file = paste(chemin.resultats,"results_causal.csv", sep=""),
sep = ";", dec="," , na="", col.names = FALSE, row.names = FALSE, qmethod = "double",
append=TRUE)
write.table(matching.scheme, file = paste(chemin.resultats,"results_causal_imput.csv",
sep=""), sep = ";", dec="," , na="", col.names = FALSE, row.names = FALSE, qmethod =
"double", append=TRUE)

## EXPORTATION DE TOUS LES RÉSULTATS EN FORMAT .CSV
# RÉSULTATS DES APPARIEMENTS
write.table(c("", "Données originales"), file = paste(chemin.resultats,"balance.csv",
sep=""), sep = ";", dec="," , na="", col.names = FALSE, row.names = FALSE, qmethod =
"double", append=TRUE)
write.table(summary.score[["sum.all"]], file = paste(chemin.resultats,"balance.csv",
sep=""), sep = ";", dec="," , na="", col.names = NA, qmethod = "double", append=TRUE)

write.table(c("", "Mahalanobis"), file = paste(chemin.resultats,"balance.csv", sep=""), sep
= ";", dec="," , na="", col.names = FALSE, row.names = FALSE, qmethod = "double",
append=TRUE)
write.table(summary.mahalanobis[["sum.matched"]], file =
paste(chemin.resultats,"balance.csv", sep=""), sep = ";", dec="," , na="", col.names = NA,
qmethod = "double", append=TRUE)

write.table(c("", "Score de propension"), file = paste(chemin.resultats,"balance.csv",
sep=""), sep = ";", dec="," , na="", col.names = FALSE, row.names = FALSE, qmethod =
"double", append=TRUE)
write.table(summary.score[["sum.matched"]], file = paste(chemin.resultats,"balance.csv",
sep=""), sep = ";", dec="," , na="", col.names = NA, qmethod = "double", append=TRUE)

write.table(c("", "Appariement génétique"), file = paste(chemin.resultats,"balance.csv",
sep=""), sep = ";", dec="," , na="", col.names = FALSE, row.names = FALSE, qmethod =
"double", append=TRUE)
write.table(summary.genetic[["sum.matched"]], file = paste(chemin.resultats,"balance.csv",
sep=""), sep = ";", dec="," , na="", col.names = NA, qmethod = "double", append=TRUE)

# RÉSULTATS DE LA RÉDUCTION DES ÉCARTS EN %
write.table(c("", "Mahalanobis"), file = paste(chemin.resultats,"reduction.csv", sep=""),
sep = ";", dec="," , na="", col.names = FALSE, row.names = FALSE, qmethod = "double",
append=TRUE)
write.table(summary.mahalanobis[["reduction"]], file =
paste(chemin.resultats,"reduction.csv", sep=""), sep = ";", dec="," , na="", col.names = NA,
qmethod = "double", append=TRUE)

write.table(c("", "Score de propension"), file = paste(chemin.resultats,"reduction.csv",
sep=""), sep = ";", dec="," , na="", col.names = FALSE, row.names = FALSE, qmethod =
"double", append=TRUE)
write.table(summary.score[["reduction"]], file = paste(chemin.resultats,"reduction.csv",
sep=""), sep = ";", dec="," , na="", col.names = NA, qmethod = "double", append=TRUE)

write.table(c("", "Appariement génétique"), file = paste(chemin.resultats,"reduction.csv",
sep=""), sep = ";", dec="," , na="", col.names = FALSE, row.names = FALSE, qmethod =
"double", append=TRUE)

```

```

write.table(summary.genetic[["reduction"]], file = paste(chemin.resultats,"reduction.csv",
sep=""), sep = ";", dec=",", na="", col.names = NA, qmethod = "double", append=TRUE)

# ANALYSE CAUSALE
write.table(c("", "Sans covariables"), file = paste(chemin.resultats,"causal.csv", sep=""),
sep = ";", dec=",", na="", col.names = FALSE, row.names = FALSE, qmethod = "double",
append=TRUE)
write.table(resultat.sans.appariement, file = paste(chemin.resultats,"causal.csv", sep=""),
sep = ";", dec=",", na="", col.names = NA, qmethod = "double", append=TRUE)

write.table(c("", "Avec covariables"), file = paste(chemin.resultats,"causal.csv", sep=""),
sep = ";", dec=",", na="", col.names = FALSE, row.names = FALSE, qmethod = "double",
append=TRUE)
write.table(resultat.avec.covariables, file = paste(chemin.resultats,"causal.csv", sep=""),
sep = ";", dec=",", na="", col.names = NA, qmethod = "double", append=TRUE)

write.table(c("", "Mahalanobis"), file = paste(chemin.resultats,"causal.csv", sep=""), sep =
";", dec=",", na="", col.names = FALSE, row.names = FALSE, qmethod = "double", append=TRUE)
write.table(resultat.mahalanobis, file = paste(chemin.resultats,"causal.csv", sep=""), sep =
";", dec=",", na="", col.names = NA, qmethod = "double", append=TRUE)

write.table(c("", "Score de propension"), file = paste(chemin.resultats,"causal.csv",
sep=""), sep = ";", dec=",", na="", col.names = FALSE, row.names = FALSE, qmethod =
"double", append=TRUE)
write.table(resultat.score, file = paste(chemin.resultats,"causal.csv", sep=""), sep = ";",
dec=",", na="", col.names = NA, qmethod = "double", append=TRUE)

write.table(c("", "Appariement génétique"), file = paste(chemin.resultats,"causal.csv",
sep=""), sep = ";", dec=",", na="", col.names = FALSE, row.names = FALSE, qmethod =
"double", append=TRUE)
write.table(resultat.genetic, file = paste(chemin.resultats,"causal.csv", sep=""), sep =
";", dec=",", na="", col.names = NA, qmethod = "double", append=TRUE)

# ANALYSE CAUSALE AVEC IMPUTATION
write.table(c("", "Sans covariables"), file = paste(chemin.resultats,"causal_imput.csv",
sep=""), sep = ";", dec=",", na="", col.names = FALSE, row.names = FALSE, qmethod =
"double", append=TRUE)
write.table(resultat.sans.appariement.imput, file =
paste(chemin.resultats,"causal_imput.csv", sep=""), sep = ";", dec=",", na="", col.names =
NA, qmethod = "double", append=TRUE)

write.table(c("", "Avec covariables"), file = paste(chemin.resultats,"causal_imput.csv",
sep=""), sep = ";", dec=",", na="", col.names = FALSE, row.names = FALSE, qmethod =
"double", append=TRUE)
write.table(resultat.avec.covariables.imput, file =
paste(chemin.resultats,"causal_imput.csv", sep=""), sep = ";", dec=",", na="", col.names =
NA, qmethod = "double", append=TRUE)

write.table(c("", "Mahalanobis"), file = paste(chemin.resultats,"causal_imput.csv", sep=""),
sep = ";", dec=",", na="", col.names = FALSE, row.names = FALSE, qmethod = "double",
append=TRUE)
write.table(resultat.mahalanobis.imput, file = paste(chemin.resultats,"causal_imput.csv",
sep=""), sep = ";", dec=",", na="", col.names = NA, qmethod = "double", append=TRUE)

```

```

write.table(c("", "Score de propension"), file = paste(chemin.resultats, "causal_imput.csv",
sep=""), sep = ";", dec=",", na="", col.names = FALSE, row.names = FALSE, qmethod =
"double", append=TRUE)
write.table(resultat.score.imput, file = paste(chemin.resultats, "causal_imput.csv", sep=""),
sep = ";", dec=",", na="", col.names = NA, qmethod = "double", append=TRUE)

write.table(c("", "Appariement génétique"), file =
paste(chemin.resultats, "causal_imput.csv", sep=""), sep = ";", dec=",", na="", col.names =
FALSE, row.names = FALSE, qmethod = "double", append=TRUE)
write.table(resultat.genetic.imput, file = paste(chemin.resultats, "causal_imput.csv",
sep=""), sep = ";", dec=",", na="", col.names = NA, qmethod = "double", append=TRUE)

## EXPORTATION DES RÉSULTATS LES PLUS IMPORTANTS EN FORMAT .CSV
# CONSTRUIRE UNE TABLE DE DONNÉES POUR COMBINER LES RÉSULTATS DES APPARIEMENTS ET DE LA
RÉDUCTION DES ÉCARTS
results.bal <- cbind.data.frame(summary.score[["sum.all"]][["Std. Mean Diff."]],
summary.mahalanobis[["sum.matched"]][["Std. Mean Diff."]],
summary.score[["sum.matched"]][["Std. Mean Diff."]],
summary.genetic[["sum.matched"]][["Std. Mean Diff."]],
summary.mahalanobis[["reduction"]][["Std. Mean Diff."]],
summary.score[["reduction"]][["Std. Mean Diff."]],
summary.genetic[["reduction"]][["Std. Mean Diff."]]
)

# NOMS DES COLONNES
names(results.bal) <- c("Brut", "Mahalanobis", "Score", "Génétique", "Mahalanobis ", "Score
", "Génétique ")
# NOMS DES LIGNES
row.names(results.bal) <- var.appariement.nom

# CONSTRUIRE UNE TABLE DE DONNÉES POUR COMBINER LES RÉSULTATS DES ANALYSES CAUSALES
results.causal <- rbind.data.frame("Brut"=resultat.sans.appariement[var.tx, ],
"Co-var"=resultat.avec.covariables[var.tx, ],
"Mahalanobis"=resultat.mahalanobis[var.tx, ],
"Score"=resultat.score[var.tx, ],
"Gététique"=resultat.genetic[var.tx, ]
)

# NOMS DES COLONNES
names(results.causal) <- c("Coefficients", "Err. Stand.", "B.I.", "B.S.")

# CONSTRUIRE UNE TABLE DE DONNÉES POUR COMBINER LES RÉSULTATS DES ANALYSES CAUSALES AVEC
IMPUTATION
results.causal.imput <- rbind.data.frame("Brut"=resultat.sans.appariement.imput[var.tx, ],
"Co-var"=resultat.avec.covariables.imput[var.tx, ],
"Mahalanobis"=resultat.mahalanobis.imput[var.tx, ],
"Score"=resultat.score.imput[var.tx, ],
"Gététique"=resultat.genetic.imput[var.tx, ]
)

# NOMS DES COLONNES
names(results.causal.imput) <- c("Coefficients", "Err. Stand.", "B.I.", "B.S.")

```

```

# EXPORTATIONS DES TABLES DE DONNÉES EN FORMATS .CSV
write.table(results.bal, file = paste(chemin.resultats,"results_bal.csv", sep=""), sep =
";", dec=".", na="", col.names = NA, qmethod = "double", append=TRUE)
write.table(results.causal, file = paste(chemin.resultats,"results_causal.csv", sep=""), sep
= ";", dec=".", na="", col.names = NA, qmethod = "double", append=TRUE)
write.table(results.causal.imput, file = paste(chemin.resultats,"results_causal_imput.csv",
sep=""), sep = ";", dec=".", na="", col.names = NA, qmethod = "double", append=TRUE)

#####
##          GRAPHIQUES          ##
#####

win.metafile( file = paste(chemin.resultats.graphiques, paste ("balance_scheme", i,
"graph%03d.wmf", sep = "_"), sep=""), width = 2, height = 2, pointsize = 6)
plot(summary.mahalanobis, interactive=FALSE)
title(main = "Distance de Mahalanobis")           # 1st plot goes to myPlot 1.wmf
plot(summary.score, interactive=FALSE, beside=TRUE)
title(main = "Score de propension")              # 2nd plot goes to myPlot 2.wmf
plot(summary.genetic, interactive=FALSE, beside=TRUE)
title(main = "Appariement génétique")           # 3rd plot goes to myPlot 3.wmf
dev.off()

#####
### OBTENIR LE N DE CHAQUE SCÈME ###
#####
if (i==1) summary.mahalanobis.N.sheme.1.1.sans.rep <-
as.data.frame(summary.mahalanobis[["nn"]])
if (i==2) summary.mahalanobis.N.sheme.1.1.avec.rep <-
as.data.frame(summary.mahalanobis[["nn"]])
if (i==3) summary.mahalanobis.N.sheme.2.1.sans.rep <-
as.data.frame(summary.mahalanobis[["nn"]])
if (i==4) summary.mahalanobis.N.sheme.2.1.avec.rep <-
as.data.frame(summary.mahalanobis[["nn"]])

if (i==1) summary.score.N.sheme.1.1.sans.rep <- as.data.frame(summary.score[["nn"]])
if (i==2) summary.score.N.sheme.1.1.avec.rep <- as.data.frame(summary.score[["nn"]])
if (i==3) summary.score.N.sheme.2.1.sans.rep <- as.data.frame(summary.score[["nn"]])
if (i==4) summary.score.N.sheme.2.1.avec.rep <- as.data.frame(summary.score[["nn"]])

if (i==1) summary.genetic.N.sheme.1.1.sans.rep <- as.data.frame(summary.genetic[["nn"]])
if (i==2) summary.genetic.N.sheme.1.1.avec.rep <- as.data.frame(summary.genetic[["nn"]])
if (i==3) summary.genetic.N.sheme.2.1.sans.rep <- as.data.frame(summary.genetic[["nn"]])
if (i==4) summary.genetic.N.sheme.2.1.avec.rep <- as.data.frame(summary.genetic[["nn"]])

} # Fin de la boucle des quatre schèmes d'appariement

```

