

Université de Montréal

**Statistiques appliquées en chirurgie cardiaque adulte :
Analyses de survie et applications du “*propensity score*”**

Par

Louis-Mathieu Stevens

Sciences Biomédicales

Faculté de Médecine

Thèse présentée à la Faculté des Études Supérieures
en vue de l'obtention du grade de Philosophiæ Doctor (PhD)
en sciences biomédicales

Mai 2012

© Louis-Mathieu Stevens 2012

Université de Montréal
Faculté des Études Supérieures

Ce mémoire intitulé :

**Statistiques appliquées en chirurgie cardiaque adulte :
Analyses de survie et applications du “*propensity score*”**

Présenté par :

Louis-Mathieu Stevens

A été évalué par un jury composé des personnes suivantes :

Hung Li, MD, MSc

Président du comité/délégué du centre de recherche

Paul Khairy, MD, PhD

Directeur de recherche

Hugues Jeanmart, MD, MSc

Membre du jury

Stéphane Rinfret, MD, MSc

Évaluateur externe

Michel Carrier, MD, MSc

Représentant du doyen

RÉSUMÉ FRANÇAIS

L'objectif principal de ce travail est d'étudier en profondeur certaines techniques biostatistiques avancées en recherche évaluative en chirurgie cardiaque adulte. Les études ont été conçues pour intégrer les concepts d'analyse de survie, analyse de régression avec "*propensity score*", et analyse de coûts.

Le premier manuscrit évalue la survie après la réparation chirurgicale de la dissection aigüe de l'aorte ascendante. Les analyses statistiques utilisées comprennent : analyses de survie avec régression paramétrique des phases de risque et d'autres méthodes paramétriques (exponentielle, Weibull), semi-paramétriques (Cox) ou non-paramétriques (Kaplan-Meier); survie comparée à une cohorte appariée pour l'âge, le sexe et la race utilisant des tables de statistiques de survie gouvernementales ; modèles de régression avec "*bootstrapping*" et "*multinomial logit model*". L'étude a démontrée que la survie s'est améliorée sur 25 ans en lien avec des changements dans les techniques chirurgicales et d'imagerie diagnostique.

Le second manuscrit est axé sur les résultats des pontages coronariens isolés chez des patients ayant des antécédents d'intervention coronarienne percutanée. Les analyses statistiques utilisées comprennent : modèles de régression avec "*propensity score*"; algorithme complexe d'appariement (1:3); analyses statistiques appropriées pour les groupes appariés (différences standardisées, "*generalized estimating equations*", modèle de Cox stratifié). L'étude a démontrée que l'intervention coronarienne percutanée subie 14 jours ou plus avant la chirurgie de pontages coronariens n'est pas associée à des résultats négatifs à court ou long terme.

Le troisième manuscrit évalue les conséquences financières et les changements démographiques survenant pour un centre hospitalier universitaire suite à la mise en place d'un programme de chirurgie cardiaque satellite. Les analyses statistiques utilisées comprennent : modèles de régression multivariée “two-way” ANOVA (logistique, linéaire ou ordinale) ; “propensity score” ; analyses de coûts avec modèles paramétriques Log-Normal. Des modèles d’analyse de « survie » ont également été explorés, utilisant les «coûts» au lieu du « temps » comme variable dépendante, et ont menés à des conclusions similaires. L'étude a démontrée que, après la mise en place du programme satellite, moins de patients de faible complexité étaient référés de la région du programme satellite au centre hospitalier universitaire, avec une augmentation de la charge de travail infirmier et des coûts.

Nombre de mots (résumé) : 350

Mots-clés : analyse des coûts ; analyse de survie ; angioplastie coronarienne ; chirurgie cardiaque ; dissection aigüe de l’aorte de type A ; économie de la santé ; études de suivi ; méthodes épidémiologiques ; pontage, revascularisation ; statistique, analyse de régression.

RÉSUMÉ ANGLAIS

The main objective of this work is to study in depth advanced biostatistical techniques in adult cardiac surgery outcome research. The studies were designed to incorporate the concepts of survival analysis, regression analysis with propensity score, and cost analysis.

The first manuscript assessed survival, and cardiovascular specific mortality, following surgical repair of acute ascending aortic dissection. The statistical analyses included survival analyses with multiphase parametric hazard regression and other parametric (exponential, Weibull), semi-parametric (Cox) or non-parametric models (Kaplan Meier), comparison with the survival of a matched cohort for age, gender and race using State lifetables, and modelization with bootstrapping and multinomial logit models. The study showed that the early and late survival following surgical repair has improved progressively over 25 years in association with noticeable changes in surgical techniques and preoperative diagnostic testing.

The second manuscript focused on outcomes following isolated coronary artery bypass grafting in patients with a history of percutaneous coronary intervention. The statistical analyses included multivariable regression models with propensity score, complex matching algorithm (1:3) and appropriate statistical analyses for matched groups (standardized differences, generalized estimating equations, and survival analyses with stratified proportional hazards models). The study showed that remote prior percutaneous coronary intervention more than 14

days before coronary artery bypass grafting surgery was not associated with adverse outcomes at short or long-term follow-up.

The third manuscript evaluated the financial consequences and the changes in case-mix that occurred at an academic medical center subsequent to the implementation of a satellite cardiac surgery program. The statistical analyses included two-way ANOVA multivariable regression models (logistic, linear or ordinal), propensity score, and cost analyses using Log-Normal parametric models. “Survival” analyses models were also explored, using “cost” instead of “time” as the outcome of interest, and led to similar conclusions. The study showed that, after implementation of the satellite cardiac surgery program, fewer patients of lower complexity came to the academic medical center from the satellite program area, with a significant increase in nursing workload and costs.

Word count (abstract) : 327

Keywords : Acute Type A Aortic Dissection ; Bypass, revascularization ; Cardiac surgery ; Coronary angioplasty ; Cost analysis ; Epidemiologic methods ; Follow-up studies ; Health economics ; Statistics, regression analysis ; Survival Analysis.

TABLE DES MATIÈRES

Page de titre.....	i
Identification du jury.....	ii
Résumé et mots clés en français	iii
Résumé et mots clés en anglais.....	v
Table des matières.....	vii
Liste des tableaux.....	x
Liste des figures	xiii
Liste des abréviations.....	xviii
Dédicace.....	xxiii
Remerciements.....	xxiv
Avant-propos.....	xxvi
Chapitre I : Introduction.....	1
Utilisation des statistiques en chirurgie.....	2
Motivation	7
Dissection aigüe de l'aorte ascendante	9

Pontage coronariens après intervention coronarienne percutanée	12
Analyse de coûts en chirurgie cardiaque.....	17
Régression logistique et linéaire	20
<i>Propensity score</i>	26
Analyse de survie	42
Décomposition dans le temps du niveau de risque (<i>hazard</i>) en phases	69
<i>Bootstrapping</i>	73
Chapitre II : Hypothèses et buts.....	77
Objectifs généraux	78
Dissection aigüe de l’aorte ascendante	79
Pontage coronariens après intervention coronarienne percutanée	80
Analyse de coûts en chirurgie cardiaque.....	81
Chapitre III : Corps de la thèse – Articles originaux	82
<i>Evolution in surgical management and outcome for acute ascending aortic dissection</i>	83
<i>Coronary artery bypass grafting after recent or remote percutaneous coronary intervention in the Commonwealth of Massachusetts</i>	123

<i>Financial consequences of implementing a partner in care in cardiac surgery</i>	160
Chapitre IV : Discussion des résultats	194
Biostatistiques appliquées en chirurgie cardiaque	195
Dissection aigüe de l'aorte ascendante	196
Pontage coronariens après intervention coronarienne percutanée	214
Analyse de coûts en chirurgie cardiaque.....	225
Chapitre V : Conclusion et perspectives	235
Sources documentaires.....	244
Appendices	
Appendice A : Formulaire de dérogation.....	279
Appendice B : Autorisation de soumettre la thèse par articles	281
Appendice C : Formulaires de consentement des co-auteurs.....	283
Appendice D : Formulaires de permission des Journaux.....	295
Appendice E : Curriculum vitae.....	299

LISTE DES TABLEAUX

Chapitre 1 : Introduction

Table 1. Exemple simple de “*propensity score*” pour deux groupes fictifs.

Article : *Evolution in surgical management and outcome for acute ascending aortic dissection*

Table 1 Patient characteristics, operative data and hospital mortality.

Table 2 Independent predictors of hospital death.

Table 3 Independent predictors of postoperative stroke.

Table 4 Incremental risk factors for all-cause and cardiovascular death.

Article : *Coronary artery bypass grafting after recent or remote percutaneous coronary intervention in the Commonwealth of Massachusetts*

Table 1 Patient Characteristics in CABG Patients With Remote Prior PCI (≥14 days) or With No Prior PCI.

Table 2 *Operative Data and Postoperative Course in CABG Patients With Remote Prior PCI (≥ 14 days) or With No Prior PCI (matched pairs).*

Tables supplémentaires :

Table 1 *Independent Predictors of Long-Term Mortality.*

Table 2 *Patient Characteristics in CABG Patients With Recent Prior PCI (< 14 days) or With No Prior PCI.*

Article : *Financial consequences of implementing a partner in care in cardiac surgery*

Table 1 *Characteristics of AMC patients from SAT or other referral areas, before (01/2002-03/2003) or after (08/2003-12/2005) SAT implementation, and characteristics of SAT patients treated at SAT hospital (08/2003-12/2005).*

Table 2 *Operative data of AMC patients from SAT or other referral areas, before (01/2002-03/2003) or after (08/2003-12/2005) SAT implementation, and operative data of SAT patients treated at SAT hospital (08/2003-12/2005).*

Table 3 Postoperative course and insurance status of AMC patients from SAT or other referral areas, before (01/2002-03/2003) or after (08/2003-12/2005) SAT implementation.

Table 4 Adjusted mean costs, revenue, hospital margin (in thousand\$) for AMC patients from SAT or other referral areas, before (01/2002-03/2003) or after (08/2003-12/2005) SAT implementation.

Chapitre IV : Discussion des résultats

Table 1. Facteurs de risque de décès au suivi, selon le modèle de Cox multivarié.

Table 2. Facteurs de risque de décès lorsque le suivi maximal est limité à 12 mois, selon le modèle de Cox multivarié.

Table 3. Effet des ratios d'appariement sur le nombre de patients dans chaque groupe et sur les rapports de cote (*odds ratio* ; OR) du décès à 30 jours et de l'utilisation de l'artère mammaire interne gauche (AMIG).

Table 4. Résultats obtenus pour les rapports de risque entre les groupes en utilisant diverses techniques d'application du PSc.

Table 5. Comparaison des distributions Normale et Log-Normale.

LISTE DES FIGURES

Chapitre 1 : Introduction

- Figure 1. Vue antérieure (frontale) du cœur ouvert. Les flèches blanches indiquent le flux sanguin normal.
- Figure 2. Schéma d'une dissection aortique.
- Figure 3. Classifications de la dissection de l'aorte ascendante.
- Figure 4. Intervention coronarienne percutanée avec tutorisation (*stenting*).
- Figure 5. Pontages coronariens.
- Figure 6. Courbe typique de coût marginal avec revenu marginal superposé.
- Figure 7. Distribution Normale.
- Figure 8. Exemple de nuage de point (*scatter-plot*) et de régression linéaire.
- Figure 9. Exemples de distributions du PSc.
- Figure 10. Distribution du "*propensity score*" et effet de l'appariement 1:3.
- Figure 11. Exemple de distribution du "*propensity score*" et effet de la stratification.
- Figure 12. Censure à droite (*right-censoring*).

Figure 13. Exemples d'estimés de survie avec censure en intervalle.

Figure 14. Exemple courbes de survie et de risque de Cox ajustées.

Article : *Evolution in surgical management and outcome for acute ascending aortic dissection*

Figure 1 Survival (A) and freedom from cardiovascular death (B) after surgery for all patients.

Figure 2 Multivariable-adjusted survival for patient aged 35, 50, 65 and 80 years (A) and multivariable-adjusted 1, 5, 10 and 15-year survival by year of surgery (B).

Figure 3 Changing patterns in management of acute ascending aortic dissection.

Figure E1 Survival after surgery for all hospital survivors compared to Massachusetts' general population matched for age, gender and race.

Figure E2 Freedom from aortic reoperation on aorta or aortic valve for all hospital survivors.

Article : *Coronary artery bypass grafting after recent or remote percutaneous coronary intervention in the Commonwealth of Massachusetts*

Figure 1 Time interval between PCI and CABG for patients with remote prior PCI (≥ 14 days).

Figure 2 Survival after CABG surgery for (A) patients with remote (≥ 14 days) prior PCI vs. patients with no prior PCI, and (B) propensity score matched patients with remote prior PCI vs. patients with no prior PCI.

Figure 3 Survival after CABG surgery for (A) patients with recent (< 14 days) prior PCI vs. patients with no prior PCI, and (B) propensity score matched patients with recent prior PCI vs. patients with no prior PCI.

Article : *Financial consequences of implementing a partner in care in cardiac surgery*

Figure 1 Annualized cardiac surgery rate non-parametrically smoothed with a one-year moving window for all cardiac surgeries (A) or isolated CABG surgeries (B).

Figure 2 Median contribution margin expressed as revenue minus variable costs (patient-related) for AMC patients referred or not from the SAT area, before and after SAT implementation.

Chapitre IV : Discussion des résultats

- Figure 1. Transformation de variables continues et score logit.
- Figure 2. Modèle multinomial pour les modalités d'imagerie diagnostique.
- Figure 3. Fonction de risque cumulative.
- Figure 4. Importance relative des phases précoce et tardive au suivi.
- Figure 5. Survie et niveau de risque instantané de décès après la chirurgie.
- Figure 6. Courbe de survie ajustée selon l'âge du patient.
- Figure 7. Courbe de survie des patients survivant à l'hospitalisation comparée à la population générale du Massachusetts appariée selon l'âge, le sexe et la race.
- Figure 8. Devis d'étude pour l'analyse des résultats des patients PAC ayant ou non un antécédent d'ICP.
- Figure 9. Distribution du "propensity score" et effet de divers ratio d'appariement.

Figure 10. Distribution des coûts totaux d'hospitalisation pour la cohorte entière.

Figure 11. Analyse des coûts, revenus et ratio profits/pertes selon le modèle de Kaplan-Meier.

LISTE DES ABRÉVIATIONS

AAAD :	dissection aigüe de l'aorte ascendante (<i>acute ascending aortic dissection</i>)
AAR :	remplacement de l'aorte ascendante (<i>ascending aorta replacement</i>)
ACE inhibitors :	inhibiteurs de l'enzyme de conversion de l'angiotensine (<i>angiotensin-converting enzyme inhibitors</i>)
AMC :	centre médical universitaire (<i>academic medical center</i>)
ANOVA :	analyse de variance (<i>analyse of variance</i>)
CA :	arrêt circulatoire (<i>circulatory arrest</i>)
CABG :	pontage aorto-coronarien (<i>coronary artery bypass grafting</i>)
CHF :	insuffisance cardiaque (<i>congestive heart failure</i>)
CI :	intervalle de confiance (<i>confidence interval</i>)
CHU :	centre hospitalier universitaire (<i>academic medical center</i>)
COPD :	maladie pulmonaire obstructive chronique (<i>chronic obstructive pulmonary disease</i>)

CPB :	circularion extracorporelle coeur-poumon (<i>cardiopulmonary bypass</i>)
CT (-scan) :	tomographie assistée par ordinateur (<i>computed tomography</i>)
CT :	Connecticut
DAAA :	dissection aigüe de l'aorte ascendante (<i>acute ascending aortic dissection</i>)
echo :	échocardiogramme (<i>echocardiogram</i>)
EKG :	électrocardiogramme (<i>electrocardiogram</i>)
FEV1 :	volume maximal expiré en une seconde (<i>forced expiratory volume in one second</i>)
GEE :	<i>generalized estimating equations</i>
HMO :	<i>health maintenance organization</i>
HR :	rapport de risque (<i>hazard ratio</i>)
Hx :	histoire de, antécédent de
IC :	intervalle de confiance (<i>confidence interval</i>)
ICD-9-CM :	<i>International Classification of Diseases, 9th Revision, Clinical Modification</i>

ICP :	intervention coronarienne percutanée, incluant dilatation et tuteurs (<i>percutaneous coronary intervention</i>)
ICU :	soins intensifs (<i>intensive care unit</i>)
IPTW :	<i>inverse-probability-of-treatment</i>
IRAD :	base de données internationale sur la dissection aigüe de l'aorte ascendante (<i>International Registry of Acute Aortic Dissection</i>)
IV :	intraveineux (<i>intravenous</i>)
K\$:	mille dollars (<i>thousand dollars</i>)
LAD :	artère descendante antérieure (<i>left anterior descending artery</i>)
LMS :	Louis-Mathieu Stevens
LOS :	durée de séjour (<i>length of stay</i>)
LVEF :	fraction d'éjection du ventricule gauche (<i>left ventricular ejection fraction</i>)
MA :	Massachusetts
MASSDAC :	base de données en chirurgie cardiaque adulte du Massachusetts (<i>Massachusetts Adult Cardiac Surgery Database</i>)
ME :	Maine

MI :	infarctus du myocarde (<i>myocardial infarction</i>)
NC :	Caroline du Nord (<i>North Carolina</i>)
NDI :	base de données nationale des décès aux États-Unis (<i>National Death Index</i>)
NH :	New Hampshire
NYHA :	New York Heart Association
OR :	rapport de côtes (<i>odds ratio</i>)
P :	valeur de <i>P</i> (<i>p-value</i>)
PAC :	pontage aorto-coronarien (<i>coronary artery bypass grafting</i>)
PCI :	intervention coronarienne percutanée, incluant dilatation et tuteurs (<i>percutaneous coronary intervention</i>)
PMC :	pontage mammo-coronarien (<i>internal thoracic artery graft</i>)
postop :	postopératoire (<i>postoperative</i>)
preop :	préopératoire (<i>preoperative</i>)
PSc :	<i>propensity score</i>
PTCA :	angioplastie (i.e. dilatation) coronarienne percutanée (<i>percutaneous coronary angioplasty</i>)

PVD :	maladie vasculaire périphérique (<i>peripheral vascular disease</i>)
r :	corrélation de Pearson r (<i>Pearson correlation r</i>)
ref. :	valeur de référence (<i>reference</i>)
SAT :	programme de chirurgie cardiaque satellite (<i>satellite cardiac surgery program</i>)
SD :	déviatiion standard, ou différence standardisée ^a (<i>standard deviation, or standardized difference</i>)
SMR :	<i>standardized mortality/morbidity ratio</i>
STS :	<i>Society of Thoracic Surgeons</i>
TSI :	<i>Transition System Incorporated</i>
US (A) :	États-Unis d'Amérique (<i>United States (of America)</i>)
VAD :	système d'assistance ventriculaire (<i>ventricular assistant devices</i>)
y :	année (<i>year</i>)

^a Seulement pour le second article: *Coronary artery bypass grafting after recent or remote percutaneous coronary intervention in the Commonwealth of Massachusetts.*

DÉDICACE

À ma tendre épouse, Geneviève, et mes enfants, Éliane, Thomas, Alexandre et Samuel, qui sont ma raison d'être et sans qui tout ceci n'aurait pu être possible.

REMERCIEMENTS

Merci à mes mentors en recherche clinique qui m'ont supporté tout au long de mon doctorat. Je pense tout particulièrement au Dr. Paul Khairy, mon directeur de thèse, un clinicien-chercheur tout aussi dynamique que disponible dont j'ai suivi les traces à l'Université Harvard. Je pense aussi au Dr. Arvind K. Agnihotri qui a eu confiance en moi et m'a supporté tout au long de ma formation postdoctorale en recherche clinique au *Massachusetts General Hospital*. Et enfin, les Drs. David F. Torchiana, un clinicien d'expérience qui m'a enseigné à voir au-delà des chiffres, et Sharon-Lise T. Normand, biostatisticienne à l'Université Harvard et directrice de la base de données de l'État du Massachusetts (MASSDAC), pour ses précieux conseils et enseignements.

Merci à mes professeurs de Maîtrise en épidémiologie clinique de la *Harvard School of Public Health*, et plus particulièrement les Drs. E. John Orav, James H. Ware, Milton Weinstein, Tobias Kurth, E. Francis Cook, Murray Mittleman et Miguel Hernán, qui m'ont donné les outils nécessaires pour réaliser mon ambitieux programme de doctorat. Merci aussi à mon grand ami, le Dr. Pallav Garg, avec qui j'ai complété mon programme de maîtrise.

Merci aux organismes subventionnaires, et plus particulièrement les Instituts en Recherche en Santé du Canada (IRSC/CIHR), le Fonds de la Recherche en Santé du Québec (FRSQ), le Département de Chirurgie (Dr. Luc Valiquette) et la Faculté des Études Supérieures de l'Université de Montréal, qui m'ont supporté financièrement durant ma formation postdoctorale aux États-Unis.

Merci à mes professeurs en résidence en chirurgie cardiaque à l'Université de Montréal, et tout particulièrement les Drs. Michel Carrier et Louis P. Perrault, qui m'ont enseigné la rigueur scientifique dans mes travaux de recherche.

Merci à mon chef de service en chirurgie cardiaque, Dr. Ignacio Prieto, et mes collègues chirurgiens, les Drs. Nicolas Noiseux, Fadi Basile et Pierre Ghosn, qui m'ont accueilli à l'Hôtel-Dieu de Montréal, me supportent tant au plan clinique que de la recherche et m'ont donné le temps nécessaire pour compléter ma thèse de doctorat à mon retour de stage.

Merci encore à mon épouse, Geneviève, et mes enfants, Éliane, Thomas, Alexandre et Samuel, qui m'ont accompagné, en personne ou en pensée, tout au long de ce périple aux États-Unis. Ils sont ma Raison d'Être, mon Havre de Paix.

Merci à ma sœur et mon frère, Claudie-Anne et Pierre-Philippe, que j'aime tant... et qui ont eu à supporter un grand frère perfectionniste.

Et surtout, surtout, merci à mes parents, Martine et Claude, qui ont créé une famille unie dans l'Amour et le Respect, m'ont encouragé dès les premiers instants à me dépasser sans cesse et m'ont donné toutes les chances de réaliser mes rêves...

AVANT-PROPOS

En tant que premier auteur de chacun des trois manuscrits, j'ai conçu et développé le devis de chacun de ces essais cliniques, procédé à la collecte de donnée et/ou procédé aux démarches administratives pour obtenir ces données dans le cadre de ma recherche, effectué la gestion et la fusion de diverses bases de données ensemble, décidé des analyses statistiques appropriées, procédé à toutes les analyses statistiques avec le logiciel d'analyses statistiques SAS, interprété les analyses statistiques, rédigé les manuscrits avec l'apport critique des co-auteurs, publié les manuscrits dans des revues révisées par les pairs de haut calibre, et présenté ces données à des congrès nationaux et internationaux.

Les lignes directrices du Guide de présentation et d'évaluation des thèses de doctorat de l'Université de Montréal stipulent que « *les programmes de doctorat visent à développer chez l'étudiant la capacité de réaliser, de façon autonome, des travaux originaux de recherche (ou de création) qui seront présentés dans une thèse et diffusés de façon appropriée* »¹. Bien que j'ai tenu à conserver de ce fait une grande autonomie pour l'ensemble de mon œuvre, je tiens à souligner les nombreuses collaborations aux plans épidémiologiques et biostatistiques que j'ai entretenu pour mes travaux avec mes mentors, professeurs et collaborateurs : Dr. Paul Khairy, Dr. Arvind K. Agnihotri, Dr. David F. Torchiana, Dr. Sharon-Lise T. Normand, Robert E. Wolf, Ann Lovett, Katya Zelevinsky, Dr. E. John Orav, Dr. Tobias Kurth, Dr. E. Francis Cook, Dr. Murray Mittleman et Dr. Miguel Hernán.

Notons qu'aucun de ces manuscrits n'a servi pour ma maîtrise en épidémiologie clinique à la *Harvard School of Public Health* effectuée à temps partiel de juillet 2006 à novembre 2008. J'ai obtenu en avril 2006 une dérogation de la part du vice-doyen de la Faculté des Études Supérieures de l'Université de Montréal, Dr. André Ferron, pour être inscrit en même temps à la maîtrise en épidémiologie clinique de la *Harvard School of Public Health* et au doctorat en sciences biomédicales à l'Université de Montréal (cf. Appendice A). Mon choix de cours à la *Harvard School of Public Health* correspond exactement au cursus de cours suivis par les étudiants inscrits au programme de doctorat en épidémiologie de la *Harvard School of Public Health*. Ces cours m'ont permis d'obtenir des bases solides pour mener à bien tous mes projets de recherche clinique effectués en parallèle au *Massachusetts General Hospital*.

De ce fait, les données cliniques de mes travaux sont tirées de bases de données américaines. Bien que les deux premiers travaux soient plus généralisables à un contexte canadien, les conclusions du troisième travail sur les analyses de coûts hospitaliers pour l'État du Massachusetts aux États-Unis ne sont pas directement transférables. Néanmoins certains principes restent : même dans un réseau public, on retrouve une certaine compétition entre les centres hospitaliers couvrant une même région (en raison des limites parfois floues des corridors de service, des références particulières, etc.) ; et les frais absorbés par l'hôpital pour le soin de patients plus malade nécessitent une compensation adéquate et ajustée de la part des payeurs, représentée principalement par un organisme gouvernemental au Québec, la Régie de l'Assurance Maladie du Québec.

CHAPITRE I

INTRODUCTION

UTILISATION DES STATISTIQUES EN CHIRURGIE

Les statistiques présentées dans les articles de recherche clinique en chirurgie sont en pleine évolution avec l'utilisation de méthodes de plus en plus complexes telles que les tests paramétrique et non paramétriques, divers types d'analyses de régression, et d'analyses de survie². Alors que 35% des articles publiés en 1985 dans *Annals of Surgery* et *Archives of Surgery* ne comportent aucune analyse statistique, on en retrouve moins de 10% en 2003². Une étude évaluant l'utilisation de méthodes statistiques dans le *New England Journal of Medicine* rapporte qu'un lecteur avec une connaissance limitée de quelques statistiques descriptive (pourcentage, moyenne, déviation standard) a plein accès à 27% des articles publiés en 1978-1978 contre seulement 12% en 2004³⁻⁵. Parmi les méthodes statistiques utilisées en 2004, 66% sont des analyses de survie, suivi par les statistiques épidémiologiques (risque relatif, rapport de cote, mesures d'association, sensibilité/spécificité ; 53%) et tables de contingence (test de Chi-carré, test exact de Fisher, test de McNemar ; 43%)^{3,5}.

Paradoxalement à cette utilisation plus fréquente des biostatistiques, on note que 27% d'un échantillon représentatif de 187 études publiées en 2003 dans cinq grands journaux chirurgicaux ont mal sélectionné ou rapporté leurs méthodes statistiques : test de t inapproprié (confusion entre test de t pour données indépendantes vs appariées, ou utilisé pour des proportions alors qu'il est indiqué pour des variables continues) ; test de Chi-carré inapproprié (par exemple, utilisé pour de petits échantillons alors que le test exact de Fisher aurait dû être utilisé) ; test paramétrique utilisé pour des données qui ne sont pas distribuées selon la loi Normale alors qu'un test non paramétrique aurait dû être utilisé et vice versa ; modèle de régression inapproprié (régression linéaire utilisé pour une variable dépendante binaire [au lieu de continue] ou régression logistique utilisé pour une variable dépendante continue [au lieu de binaire]) ; tests ne tenant pas compte des comparaisons multiples effectuées où il est théoriquement possible d'obtenir 5% du temps, par chance, un résultat « statistiquement significatif » pour l'échantillon de sujets étudié alors qu'en réalité il n'y a pas de différence entre les groupes de sujets étudiés pour la population à laquelle on souhaite que nos inférences s'appliquent².

Une autre analyse de 144 articles publiés en 2005 dans quatre grands journaux chirurgicaux rapporte que les méthodes de statistiques descriptives utilisées pour les variables continues sont inappropriées dans 50% des cas⁶. La qualité méthodologique de 653 articles publiés dans huit des journaux chirurgicaux ayant les meilleurs index de citation révèle que la qualité méthodologique des articles est faible (exemple : objectifs vague, devis d'études sous-optimaux ou non précisés, absence de critères d'éligibilité et de justification du nombre de patients étudiés)⁷.

L'évolution de la méthodologie statistique et leur pauvre qualité d'application dans les articles scientifique est également notée dans mon domaine d'expertise clinique en chirurgie cardiaque. Je le remarque plus particulièrement lors de mes lectures dans nos journaux en chirurgie et cardiologie, de nos réunions d'analyse d'articles scientifiques (« *Journal Club* »), lors d'évaluation de manuscrits pour des journaux avec comités de pairs ou encore lors de mes diverses collaborations avec mes collègues chirurgiens et cardiologues. Une série d'articles publiés ces dix dernières années dans le *Journal of Thoracic and Cardiovascular Surgery* (“*Statistics for the rest of us*”) et dans *Annals of Thoracic Surgery* (“*Statistician's page*”) ont pour but d'adresser certaines de ces lacunes en vulgarisant des concepts statistiques classiques, introduisant des techniques biostatistiques plus récentes et exposant certaines controverses dans le domaine des biostatistiques appliquées en chirurgie cardiaque⁸⁻³⁴. Un de ces articles⁹ décortique point par point la méthode statistique utilisée pour un article de chirurgie thoracique oncologique³⁵ et fut pour moi un déclencheur dans mon engouement pour les biostatistiques appliquées.

Deux de ces articles m'ont introduit au concept du “*propensity score*” (PSc)^{10,14}. Dans une série d'articles de revue systématique, Austin s'est intéressé à la qualité de la méthodologie statistique des articles utilisant l'appariement selon le PSc dans la littérature médicale de 1996 à 2003 (47 articles revus)³⁶, puis de 2004 à 2006 dans la littérature en chirurgie cardiaque (60 articles tirés de *Annals of Thoracic Surgery*, *European Journal of Cardio-thoracic Surgery*, *Journal of Cardiovascular Surgery* et *Journal of Thoracic and Cardiovascular Surgery*)³⁷ et en cardiologie (44 articles tirés de *American Heart Journal*, *American Journal of Cardiology*,

Circulation, European Heart Journal, Heart, International Journal of Cardiology, Journal of the American College of Cardiology)³⁸. L'auteur rapporte peu d'amélioration dans la pauvre qualité méthodologique d'application du PSc, laquelle est légèrement inférieure en chirurgie cardiaque malgré une utilisation plus fréquente. En chirurgie cardiaque, 31 études (52%) ne donnent pas d'information suffisante sur la façon dont les groupes appariés ont été formés, 11 études (18%) ne rapportent pas si les groupes étaient bien balancés suivant l'utilisation de l'appariement par le PSc et aucune étude utilise une technique appropriée pour comparer les caractéristiques de base des groupes appariés. Seulement 8 études (13%) utilisent des analyses appropriées pour comparer les résultats des groupes appariés en tenant compte de l'appariement. Notons que les lignes directrices proposées par l'auteur ont suscité un peu de controverse sur quelques points techniques³⁹⁻⁴¹ ; la recherche sur les méthodes d'application du PSc est très active dans la littérature statistique et les opinions varient.

L'utilisation inappropriée des méthodes épidémiologiques et biostatistiques a été documentée également pour les essais cliniques^{42,43}, ainsi qu'en recherche clinique en médecine⁴⁴⁻⁴⁹, anesthésie^{50,51} et radiologie⁵². Il est possible que ces erreurs dans la littérature surviennent parce que les cliniciens ressentent une pression à mener des projets de recherche et à publier leurs résultats alors qu'ils ne disposent pas de connaissances suffisantes pour sélectionner les devis d'étude et les méthodes biostatistiques appropriées⁵³. Aussi, l'utilisation de techniques statistiques « à la mode », tel que le PSc, peut favoriser l'acceptation de manuscrits par certains évaluateurs même si la méthode d'application de la technique n'est pas appropriée.

Malgré la publication de lignes directrices telles que le CONSORT (*Consolidated Standards of Reporting Trials*)⁵⁴⁻⁵⁶ pour les essais cliniques et STROBE (*STrengthening the Reporting of OBservational studies in Epidemiology*)⁵⁷ pour les études observationnelles, d'importantes lacunes au point de vue épidémiologique et biostatistique sont toujours notées dans la littérature actuelle^{6,48,49}. Plusieurs journaux font maintenant appel à des experts en biostatistique et épidémiologie pour les manuscrits comportant des analyses trop complexes pour le clinicien qui revoit le manuscrit ou si un doute subsiste quant à l'utilisation appropriée des méthodes biostatistiques. En revanche, on constate que la plupart des manuscrits sont revus par des cliniciens experts dans un domaine médical particulier qui n'ont souvent pas d'expertise particulière en biostatistique.

Il est essentiel pour le clinicien-chercheur de bien comprendre les hypothèses qui sous-tendent chacun des tests pour éviter de tirer des conclusions trompeuses. L'avis de collaborateurs avec connaissances en épidémiologie et biostatistique s'avère une aide précieuse dans le choix des tests appropriés pour l'élaboration ou l'analyse d'un projet de recherche, au lieu de s'en tenir aveuglément à des logiciels de statistique performant et de plus en plus facile d'utilisation. Le défi est de parvenir à communiquer efficacement avec l'épidémiologiste ou le biostatisticien car ceux-ci ont souvent une connaissance limitée de l'aspect clinique, de l'importance de la question posée et des autres hypothèses que le clinicien a parfois peine à formuler : ce qui est statistiquement significatif n'est pas nécessairement cliniquement significatif. D'ailleurs, le comité de rédaction du *Journal of Thoracic and Cardiovascular Surgery* recommande depuis peu d'éviter l'emploi des termes « statistiquement significatif ».

MOTIVATION

Constatant l'état de l'application des statistiques en médecine et chirurgie, j'ai élaboré ce projet de doctorat en ayant pour objectif principal d'étudier en profondeur certaines techniques biostatistiques avancées en recherche évaluative en chirurgie cardiaque adulte. Je me suis intéressé plus particulièrement aux concepts d'analyse de survie, de "*propensity score*", et analyse des coûts. Les techniques chirurgicales étudiées sont brièvement décrites ci-dessous pour se concentrer davantage sur la description et l'évaluation des applications épidémiologiques et biostatistiques utilisées pour mes travaux. Durant mes études au *Massachusetts General Hospital* de la *Harvard University*, j'ai été exposé à toutes les facettes de la recherche clinique en chirurgie cardiaque adulte et j'ai complété une formation en épidémiologie et biostatistique appliquée à la *Harvard School of Public Health*⁵⁸⁻⁶⁷.

Dans le but de rendre mon travail plus accessible aux cliniciens avec une connaissance de base en biostatistiques, j'ai cherché à présenter et vulgariser les analyses biostatistiques et concepts épidémiologiques dans un langage le moins technique possible, que j'ai imagé à l'aide d'exemples concrets tirés de mes travaux antérieurs et autres (voir Chapitre I). Les analyses statistiques présentées dans chacun des manuscrits, ainsi que les analyses additionnelles présentées au Chapitre IV, sont plus avancées qu'elles ne le sont généralement pour des articles de cohorte rétrospective en chirurgie cardiaque adulte. L'originalité de cette thèse tient au fait qu'elle vise à pousser plus loin certaines méthodes biostatistiques dans le cadre de la recherche évaluative en chirurgie cardiaque adulte et contribuer ainsi à l'avancement des connaissances dans ce domaine d'expertise.

Divers concepts d'analyses de survie sont approfondis pour l'ensemble des travaux et le "*propensity score*" (PSc) pour les deux travaux comportant des comparaisons de groupes de patients. Pour le premier travail⁶⁸, j'ai opté pour des analyses de survie comprenant la décomposition des phases de risque, peu rapportées dans la littérature en dehors des articles générés par celui qui a proposé cette méthode statistique (Dr. E. H. Blackstone). L'intérêt de cette technique est décrit (i.e. plus grande puissance du modèle paramétrique pour analyser les facteurs de risque pour diverses portion du suivi, lorsque le modèle est bien défini) et les résultats obtenus comparés à des méthodes biostatistiques plus conventionnelles telles que le modèle d'analyse de risque proportionnel de Cox. Pour le deuxième travail⁶⁹, le PSc est étudié en détail incluant la comparaison de diverses applications du PSc et l'étude détaillée d'algorithmes complexes d'appariement selon le PSc. L'intérêt de l'utilisation du PSc et de la technique d'appariement pour ce travail est souligné dans les Chapitres I et IV incluant une comparaison des résultats obtenus avec des analyses statistiques plus conventionnelles. Pour le troisième travail⁷⁰, plusieurs méthodes d'analyse de coûts ont été explorées, incluant un modèle d'analyse de survie semi-paramétrique de Cox, avant d'en venir au modèle proposé. Étant donné que les méthodes statistiques utilisées dans les articles médicaux ne répondaient pas aux besoins de mon devis d'étude, j'ai eu à puiser dans la littérature biostatistique, mathématique et économétrique pour être en mesure de développer un modèle satisfaisant. Il s'agit du modèle Log-Normal, lequel fait aussi partie de la famille des modèles d'analyse de survie paramétrique. D'autres alternatives sont explorées et comparées aux résultats obtenus dans le manuscrit original au Chapitre IV.

DISSECTION AIGÛE DE L'AORTE ASCENDANTE

Le premier travail porte sur l'expérience du *Massachusetts General Hospital* (MGH), à Boston (Massachusetts, États-Unis), pour la chirurgie de la dissection aigüe de l'aorte ascendante⁶⁸. Avec 170 patients et un suivi moyen complet de $7,0 \pm 5,9$ ans (jusqu'à 27 ans), cette étude représente l'une des plus longue série chirurgicale contemporaine publiée à ce jour⁷¹⁻⁷⁶.

L'aorte se situe à la sortie du cœur gauche, ou ventricule gauche (Figure 1). Elle se divise en racine aortique (valve aortique, sinus de valsalva, jonction sino-tubulaire), aorte ascendante jusqu'au tronc brachio-céphalique, aorte transverse, et aorte descendante après l'artère sous-clavière gauche.

La dissection de l'aorte se caractérise par la déchirure soit complète (rupture) ou partielle de la paroi de l'aorte entraînant l'extravasation de sang dans la paroi de l'aorte avec ou non création d'une fausse lumière, habituellement au niveau de la média (Figure 2). La dissection de l'aorte est une urgence médicale et peut rapidement mener à un décès, même en présence d'un traitement optimal. La dissection est dite aigüe lorsque le patient se présente dans les deux semaines suivant ses premiers symptômes, le plus souvent une douleur thoracique déchirante avec ou sans instabilité hémodynamique.

Figure 1. Vue antérieure (frontale) du cœur ouvert^b Les flèches blanches indiquent le flux sanguin normal.

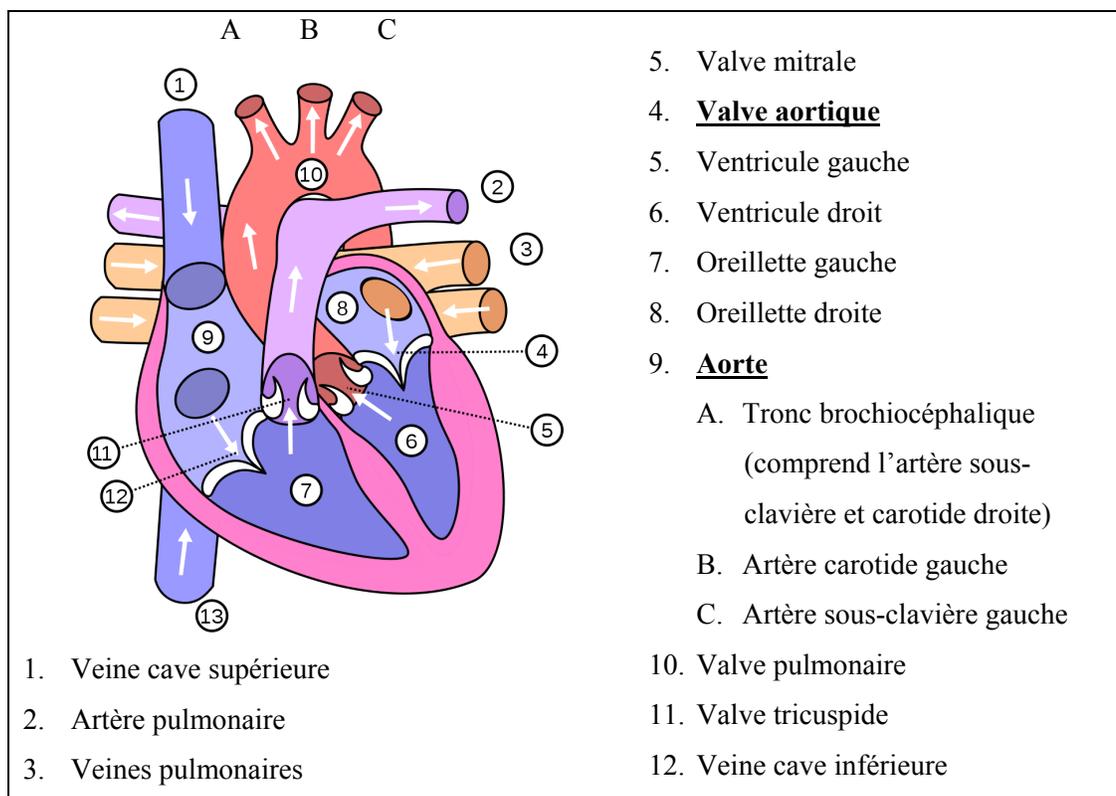
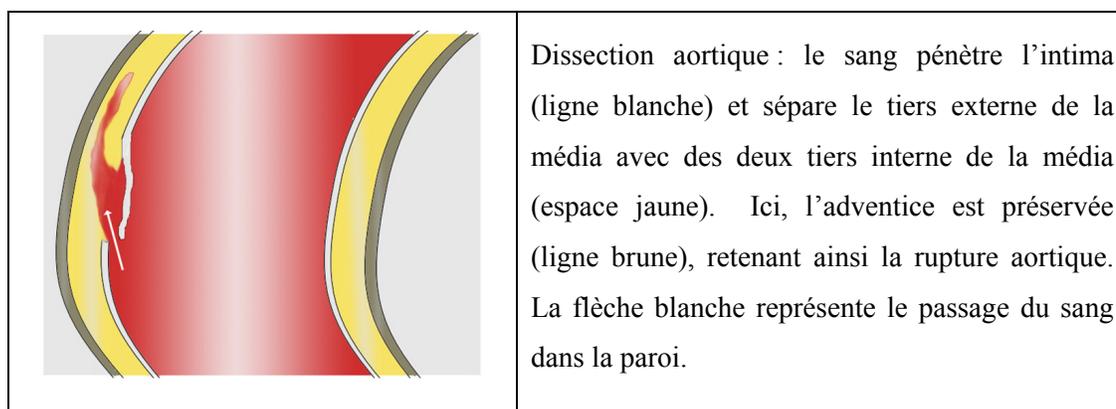


Figure 2. Schéma d'une dissection aortique^c.



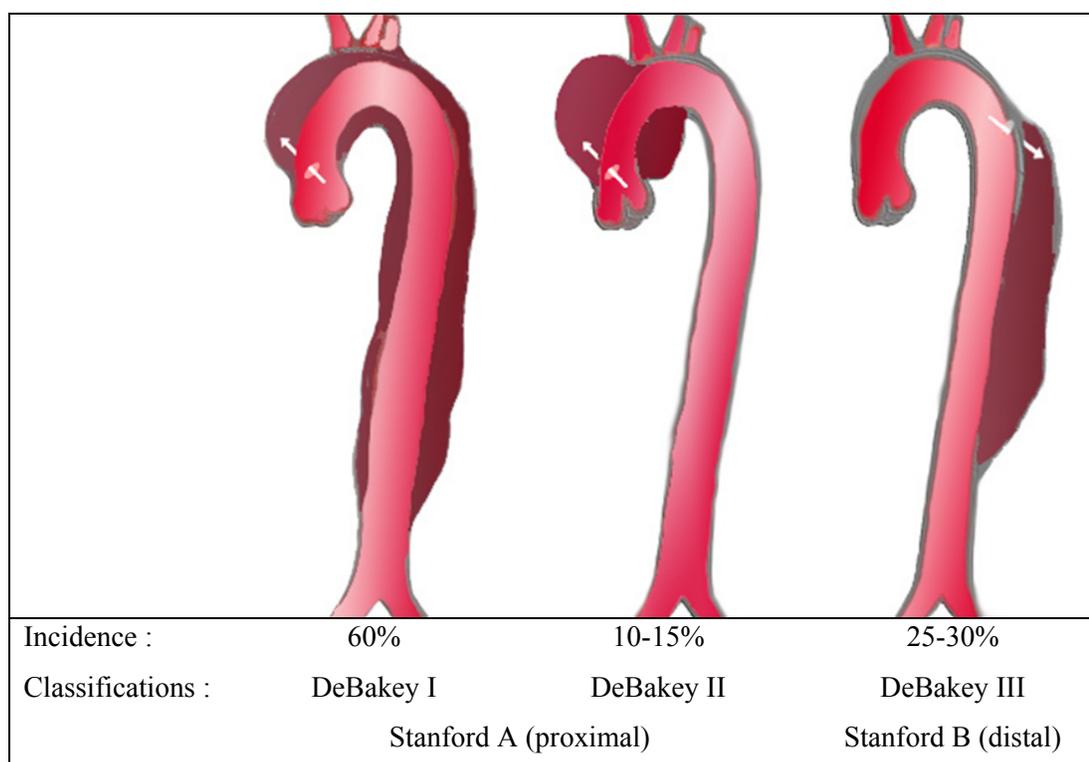
^b Copyright MesserWoland. Permission donnée par l'auteur de copier, distribuer et/ou modifier ces images selon les termes de la *GNU Free Documentation License* ([http://commons.wikimedia.org/wiki/File:Diagram_of_the_human_heart_\(multilingual\).svg](http://commons.wikimedia.org/wiki/File:Diagram_of_the_human_heart_(multilingual).svg))

^c Copyright J. Heuser. Permission donnée par l'auteur de copier, distribuer et/ou modifier ces images selon les termes de la *GNU Free Documentation License* (http://en.wikipedia.org/wiki/File:AoDissect_DeBakey1.png)

La classification de Stanford divise les dissections en deux types (Figure 3) :

- **Type A** implique l'aorte ascendante et nécessite habituellement une réparation chirurgicale. Elle comprend les DeBakey Type I (origine de l'aorte ascendante et se propage à l'aorte transverse et possiblement l'aorte descendante) et DeBakey Type II (confiné à l'aorte ascendante) ;
- **Type B** n'implique pas l'aorte ascendante. Elle est souvent traitée de façon conservatrice (médicalement). La réparation chirurgicale est indiquée lors de l'apparition d'une complication telle que l'atteinte d'un organe vital (syndrome de malperfusion), la rupture ou la dissection rétrograde dans l'aorte ascendante, ou pour les patients avec maladies du collagène (Marfan, Ehlers-Danlos) où le risque est plus grand.

Figure 3. Classifications de la dissection de l'aorte ascendante^b.



PONTAGE CORONARIENS APRÈS INTERVENTION CORONARIENNE PERCUTANÉE

La chirurgie de pontages coronariens est l'une des procédures chirurgicales majeures la plus fréquemment effectuée et étudiée en Amérique du Nord. La chirurgie cardiaque adulte est d'ailleurs un des champs d'expertise de la chirurgie où la recherche clinique est particulièrement active. Plusieurs institutions se sont dotées de larges bases de données cliniques prospectives dans un but d'amélioration de la qualité de l'acte ainsi que pour faire connaître leurs résultats. En Amérique, les États de New York, New Jersey, Pennsylvanie, Californie, Virginie et Massachusetts⁷⁷⁻⁷⁹ ont d'ailleurs institués des bases de données à participation obligatoire où les résultats à court terme des chirurgiens cardiaques sont disponibles publiquement. Leur expérience a démontré que les bases de données à vocation administrative (par opposition à clinique) sont insuffisantes, que la collecte et l'analyse de ces résultats doit être effectuée rigoureusement et que les résultats doivent être ajustés en fonction de la complexité, grandissante, des patients opérés en chirurgie cardiaque.

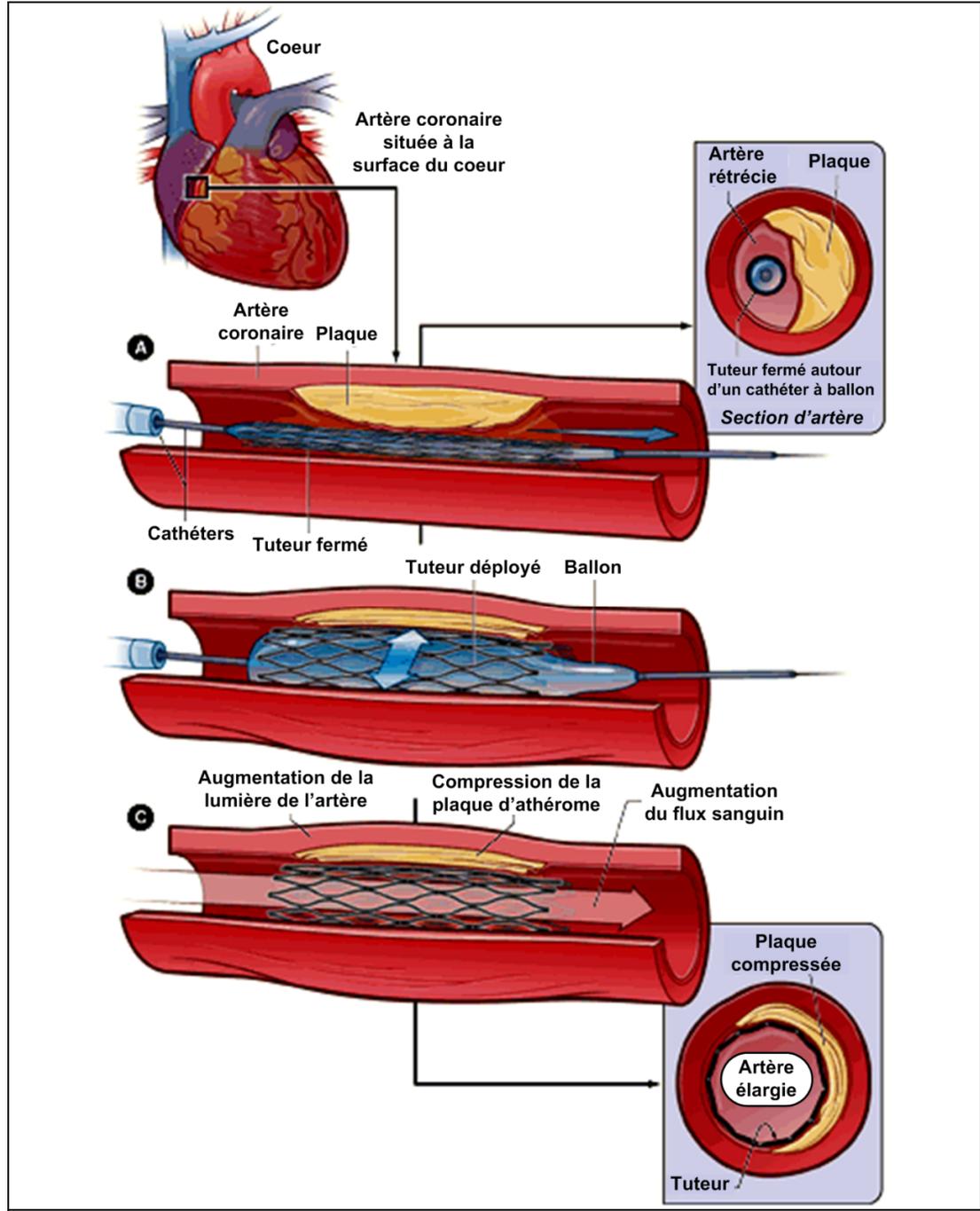
Le second travail s'intéresse aux résultats à court et long terme des pontages coronariens (PAC) après intervention coronarienne percutanée (ICP), utilisant les bases de données à participation obligatoire de l'État du Massachusetts⁶⁹. Après plusieurs démarches administratives, je suis parvenu à accéder à une version dénominalisée de ces données dans le cadre de mon doctorat et j'ai pu effectuer personnellement toutes les analyses statistiques planifiées.

Le patient présentant une ou des sténoses coronariennes sévères peut être traité médicalement, par dilatation (angioplastie), avec ou sans tutorisation (*stenting*) de la lésion par technique percutanée (Figure 4), ou par chirurgie de pontage coronarien (Figure 5).

Le pontage coronarien consiste en l'utilisation de greffon artériel et/ou veineux pour apporter du sang oxygéné au-delà de la sténose. Un pontage est dit *aorto*-coronarien lorsque le greffon artériel ou veineux est cousu (anastomosé) au niveau de l'aorte ascendante par opposition à pontage *mammo*-coronarien dont le flux sanguin provient d'une artère mammaire interne qui demeure connectée à son origine au niveau de l'artère sous-clavière. La coronarographie diagnostique est l'examen permettant d'injecter une solution radioopaque dans les artères coronariennes sous guidage fluoroscopique. Dépendamment de la présentation clinique, les caractéristiques du patient et des lésions coronariennes, le patient est référé pour l'un ou l'autre des traitements cités ci-haut.

Le nombre d'angioplasties coronariennes avec tutorisation (*stenting*) avec ou sans tuteur médicamenté a augmenté de façon importante ces vingt dernières années, et cette technologie s'est étendue à un sous-groupe de patients complexes, traditionnellement référés en chirurgie cardiaque, incluant les diabétiques^{80,81}, les patients avec dysfonction ventriculaire gauche sévère⁸², maladie du tronc commun⁸³⁻⁸⁵ et maladie tritronculaire sévère⁸⁶⁻⁹⁰.

Figure 4. Intervention coronarienne percutanée avec tutorisation (stenting)^c.



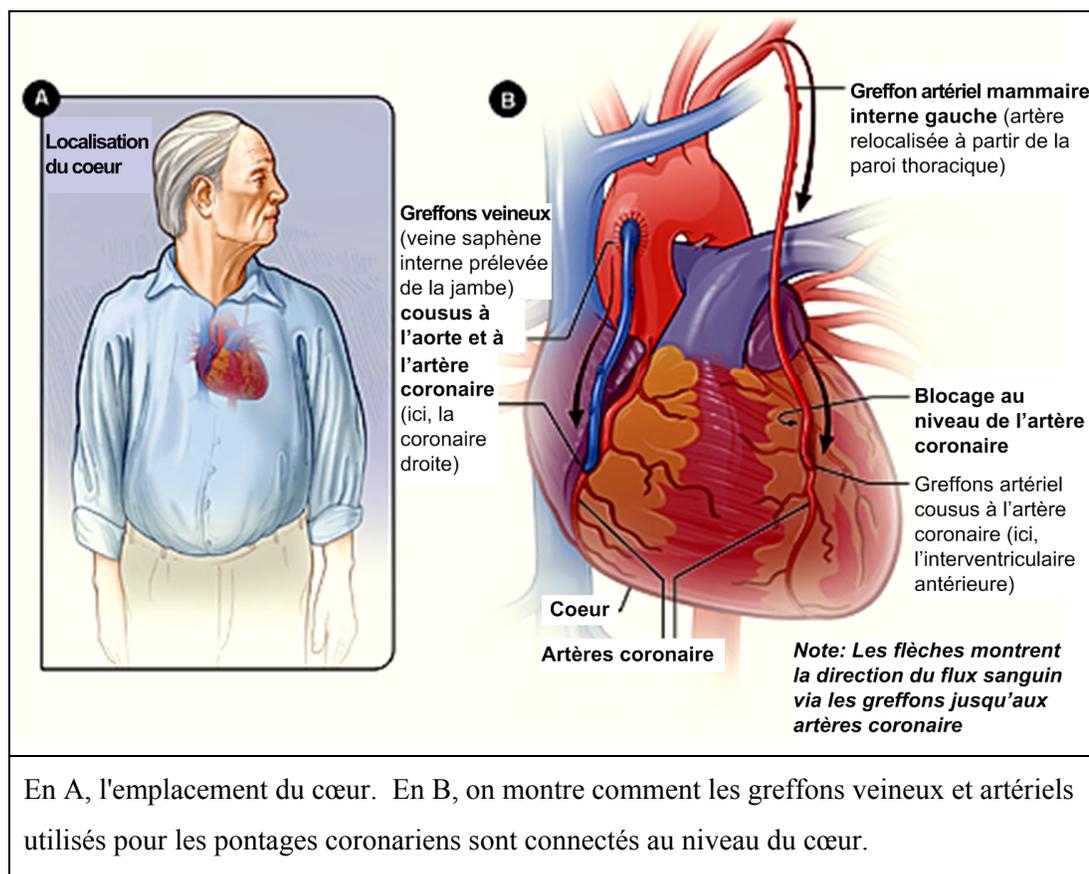
En A, le cathéter est inséré dans la lésion. En B, le ballonnet est gonflé avec expansion du tuteur et compression de la plaque. En C, le ballonnet est dégonflé. Le résultat de la pose du tuteur est apprécié schématiquement avant et après le déploiement du tuteur sur les tranches de section.

Les métaanalyses des essais cliniques randomisés récents comparant les ICP avec tutorisation vs la chirurgie de pontage coronarien, comportant un échantillon plutôt restreint de l'ensemble des patients traités, démontrent une incidence cumulative similaire de décès, accident vasculaire cérébral et d'infarctus du myocarde à cinq ans pour les deux techniques, mais une nette augmentation du taux de réintervention coronarienne chez les patients avec ICP⁸⁷. Inversement, d'importants registres comparant les résultats des ICP avec tutorisation à la chirurgie de pontage coronarien, et dont les résultats sont plus généralisables (« *real-world* »), démontrent une incidence plus faible de décès ou d'infarctus du myocarde avec la chirurgie de pontage coronarien^{86,88,90}. Dans les essais randomisés et les registres, les ICP sont associés à une augmentation de l'incidence de récurrence de symptômes et de réintervention pour revascularisation en comparaison à la chirurgie de pontage coronarien. Récemment, l'important essai clinique randomisé SYNTAX comparant les ICP avec tuteur médicamenté vs la chirurgie de pontage coronarien reconferme la place première de la chirurgie de pontage pour la maladie tritronculaire et du tronc commun, puisque celle-ci est associée à un plus faible taux du résultat combiné d'évènements cardiaques majeurs (décès, infarctus, réintervention) ou cérébrovasculaire à un an⁹¹ et trois ans⁹².

En chirurgie cardiaque, les greffons les plus souvent utilisés sont les artères mammaires (pontage mammo-coronarien ; PMC) et les veines saphènes (pontage aorto-coronarien veineux ; PAC) (Figure 5). Les PMC sont associés à une excellente perméabilité à long terme⁹³⁻⁹⁵ avec 90% des PMC greffés sur la coronaire interventriculaire antérieure perméables 10-20 ans après l'intervention chirurgicale.

La perméabilité des PMC demeure excellente lorsque utilisé en greffe libre (91% à 5 ans⁹⁶). D'un autre côté, les PAC veineux sont associés à une perméabilité à 10 ans d'environ 50-60%⁹⁷, mais jusqu'à 80% lorsque greffé sur le territoire de la coronaire interventriculaire antérieure⁹⁴.

Figure 5. Pontages coronariens^d.



^d Pas de copyright. Images éditées à partir d'images dans le domaine public provenant du *National Institutes of Health*.

ANALYSE DE COÛTS EN CHIRURGIE CARDIAQUE

Le troisième travail explore les conséquences financières de la création d'un nouveau programme de chirurgie cardiaque dans un centre hospitalier affilié situé dans une région historiquement desservie par un centre hospitalier universitaire (CHU)⁷⁰ offrant un service de chirurgie cardiaque. J'ai comparé les changements avant et après l'ouverture du programme pour les patients référés de cette région vs ceux référés des autres régions desservies par le CHU dans ce qui a trait à leurs caractéristiques et aux coûts et revenus ajustés⁹⁸ rattachés à leur hospitalisation.

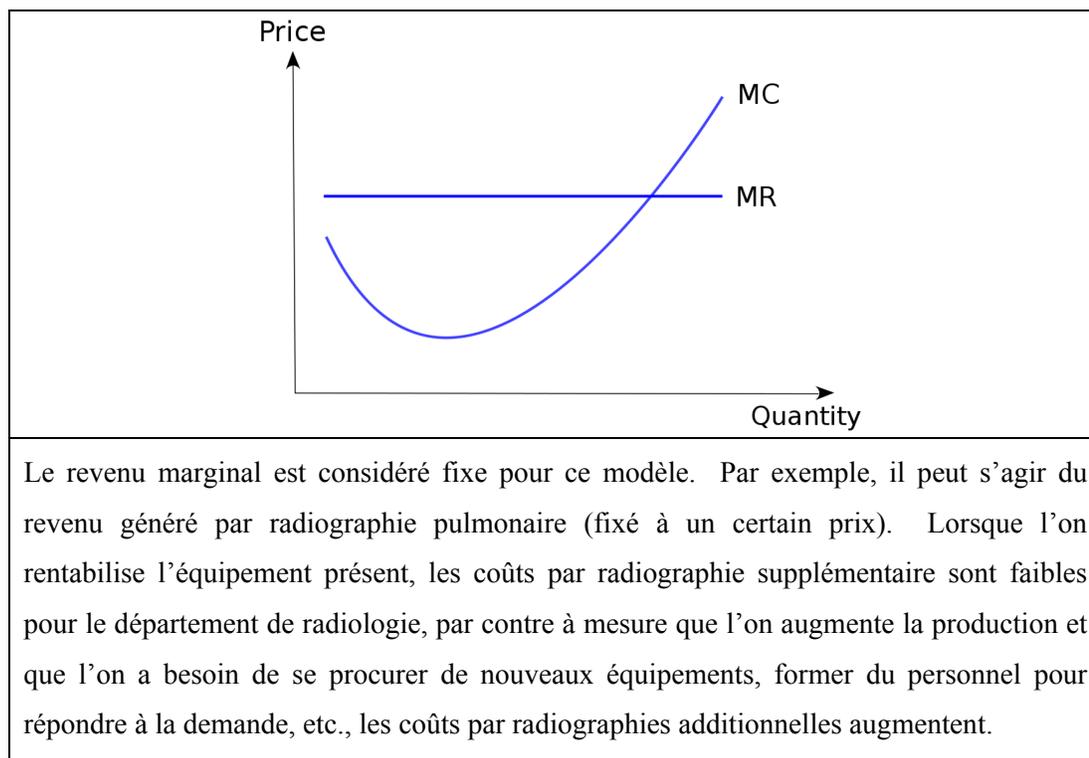
Durant la dernière décennie, le volume de chirurgie cardiaque adulte a diminué en Amérique du Nord alors que le nombre de centres de chirurgie cardiaque a paradoxalement augmenté en lien avec l'augmentation des programmes communautaires d'ICP⁹⁹. Les références en chirurgie de pontage coronarien isolés ont globalement diminuées en raison de la baisse des admissions pour symptômes coronariens aigus par les stratégies de cardiologie préventive (tel que cesser de fumer) et médicales (comme les statines)⁷⁷, mais aussi en raison de la croissance des ICP¹⁰⁰. Malgré une référence en chirurgie cardiaque de patients de plus en plus complexes, la mortalité de chirurgie cardiaque a continué de baisser. Selon la base de données de la STS, la population de patients subissant une chirurgie cardiaque est de plus en plus âgée et susceptible d'être de sexe féminin, avec tabagisme, diabète, insuffisance rénale, hypertension, accidents vasculaires cérébraux, maladie pulmonaire chronique et antécédents d'ICP, se présentant en classe fonctionnelle NYHA IV (dyspnée au repos) et avec une maladie tritonculaire avancée^{101,102}.

Les coûts reliés à une hospitalisation peuvent être divisés en 3 classes :

- *Variable directs* (liés au patient) : dépenses directement liées aux soins des patients à une date donnée, comme les tests de laboratoire, les radiographies et les fournitures jetables ;
- *Invariable (fixes) directs* (liés au département) : dépenses associées aux départements qui fournissent des soins directs aux patients, mais pas à un patient en particulier, tels que les équipements et les dispositifs médicaux ;
- *Indirects* (liés à l'institution) : dépenses engendrées en dehors des départements et individus, tels que l'administration, la sécurité et autres services de soutien des départements. Pris ensemble, les coûts liés aux départements et à l'institution sont appelés «frais généraux d'hôpital» (*hospital overhead*). Les honoraires médicaux ne sont habituellement pas inclus dans le calcul de ces coûts d'hospitalisation car le paiement de ces honoraires est fait séparément des frais exigés par l'hôpital.

Le profit généré pour une hospitalisation est défini comme la somme des revenus d'hospitalisation moins les coûts directs et indirects [profit = revenus – coûts]. Le « coût marginal » (*marginal cost*) correspond à la variation du coût total qui se produit lorsque la quantité produite est augmentée d'une unité (i.e. ensemble des coûts additionnels engendrés si l'on désire traiter un sujet de plus). La relation entre le nombre de sujets traités et le coût marginal n'est pas linéaire (Figure 6). On réalise des économies d'échelle si le coût de production d'une unité supplémentaire est moindre que le coût moyen de toutes les unités précédentes.

Figure 6. Courbe typique de coût marginal avec revenu marginal superposé^e.



La marge de contribution (*unit or margin contribution*) est le revenu généré par un sujet moins ses coûts variables, i.e. la « contribution » de ce sujet pour les frais généraux de l'institution [marge de contribution = revenu – coûts variables] (voir Figure 2 du troisième manuscrit⁷⁰ au Chapitre III). Cette contribution est plus ou moins grande selon le revenu généré et les coûts engendrés pour le sujet en question. Par exemple, elle est plus grande pour un patient de chirurgie cardiaque avec peu de comorbidité et dont le séjour hospitalier est court si le revenu versé par son assurance, en vertu de contrat de service, est le même qu'un sujet complexe qui requiert plus de ressources. L'économétrie consiste en l'utilisation de méthodes statistiques dans l'analyse de données économiques.

^e Copyright Jarry1250. Autorisation d'éditer l'image selon le « *Creative Commons Attribution-Share Alike 3.0 Unported License* ». http://en.wikipedia.org/wiki/File:Costcurve_-_Marginal_Cost_2.svg.

RÉGRESSION LOGISTIQUE ET LINÉAIRE

Il importe d'abord de définir quelques termes. Dans le cadre d'une analyse de régression, on cherche à expliquer un résultat (i.e. variable dépendante ou « *Y* ») à l'aide de caractéristiques (i.e. variable indépendante, facteurs de risque, covariables, exposition à certains facteurs). Chacune des caractéristiques étudiées explique ou non une partie de la variation dans les résultats pour les divers sujets de l'étude. Par exemple, on s'attend à ce que le risque de décès pour des pontages coronariens d'un patient âgé opéré en urgence soit plus grand qu'un jeune patient opéré électivement. L'analyse nous permet de déterminer quels sont les facteurs de risque de décès à 30 jours suivant une telle chirurgie expliquant cette « intuition » clinique.

Le modèle de régression utilisé dépend du type de résultat analysé (Table 1). Par exemple, pour un résultat mesuré sur une échelle continue, on utilise un modèle de régression linéaire, alors que pour un résultat binaire, on utilise un modèle de régression logistique. Un modèle de régression peut être utilisé pour élucider un mécanisme biologique (facteurs impliqués expliquant un événement précis) ou créer un score de prédiction (facteurs de risque de décès postopératoire).

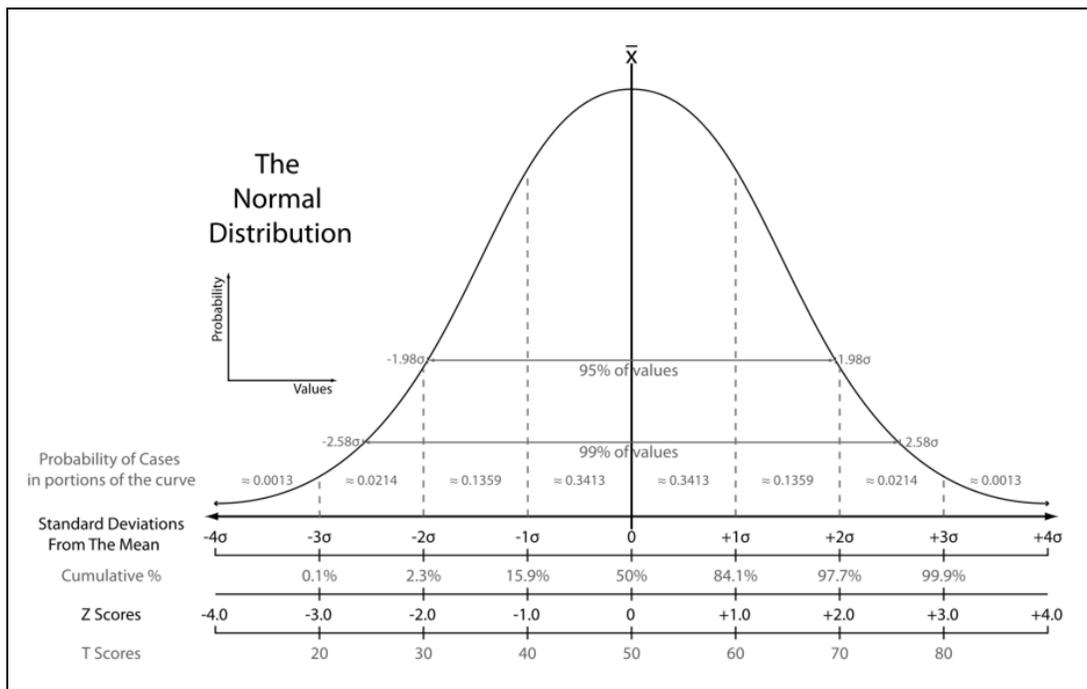
La régression est dite simple/univariée s'il n'y a qu'une seule variable indépendante ou facteurs de risque dans le modèle, ou multivariée lorsque d'autres variables indépendantes sont ajoutées au modèle pour évaluer l'impact relatif de chacune sur le résultat. Plusieurs stratégies existent pour faire le choix des variables que l'on décide de conserver dans le modèle multivarié final (*forward, backward, stepwise, score, ...*) et ne seront pas détaillées ici.

Table 1. Exemples de modèles de régression

Type de résultat	Type de régression	Autre types d'analyse possibles (si univariée)
Continu , ayant idéalement une distribution Normale (Figure 7)	Linéaire → Mesure d'effet : coefficient de régression, β	Nuage de points (<i>scatter plot</i>) (Figure 8) Coefficient corrélation
Binaire (ex. : décès à 30 jours suivant une chirurgie de pontages coronariens)	Logistique → Mesure d'effet : rapport de cote (<i>odds ratio</i> , e^β)	Graphiques en barre (<i>bar chart</i>) ou boîte avec mesure de dispersion (<i>box plot</i>) <i>t</i> test <i>Wilcoxon rank sum test</i> Analyse de variance
Ordinal (ex. : la classe fonctionnelle NYHA sur une échelle de I à IV) ou Nominal (catégorie sans gradation ; ex. : bleu, blanc, rouge)	Logistique ordinale ou multinomiale → Mesure d'effet : rapport de cote (<i>odds ratio</i> , e^β)	Graphiques à barre (<i>bar chart</i>) ou boîte avec mesure de dispersion (<i>box plot</i>) Analyse de variance Test de Kruskal-Wallis
Temps jusqu'à un évènement , avec censure (<i>censoring</i>) ^f (ex. : survie suivant une chirurgie cardiaque)	Cox ou paramétrique → Mesure d'effet : rapport de risque (<i>hazard ratio</i> , e^β)	Courbe de Kaplan-Meier
Décompte (ne peut pas être négatif)	Poisson	Régression linéaire avec variable dépendante transformée (log)

^f Les sujets peuvent avoir diverses période de suivi qu'ils aient ou non présenté l'évènement étudié à la fin du suivi, comme le décès. Voir section sur l'analyse de survie ci-dessous.

Figure 7. Distribution Normale^g.



Plus la valeur de la variable continue pour le sujet se rapproche de la valeur attendue (moyenne) dans une distribution Normale, plus on est susceptible de retrouver d'autres sujets ayant des valeurs semblables (plus la « densité de probabilité » est grande). Les valeurs statistiques utilisées dans l'estimation des tests standardisés selon la distribution Normale sont présentées dans la figure. Les échelles comprennent les écarts-types (*standard deviation*), pourcentages cumulatifs, les percentiles, et les scores Z et T.

Pour toute analyse multivariée, je m'assure que les variables choisies ne sont pas corrélées entre elles (i.e. qu'elles ne représentent pas un même concept comme « âge plus grand que 50 ans » et « ménopause »). Je me méfie des variables qui ont beaucoup de poids dans l'analyse (*leverage*), surtout si peu de sujets présentent cette caractéristique, car cela est associé à des estimés imprécis. Par exemple, si on analyse les facteurs de risque de mortalité à 30 jours pour la chirurgie de la dissection aigüe de l'aorte ascendante et que, par hasard, la majorité des quelques patients avec maladie du

^g Pas de copyright. Image libérée dans le domaine public fournie par l'auteur «Heds 1» du Wikipedia.

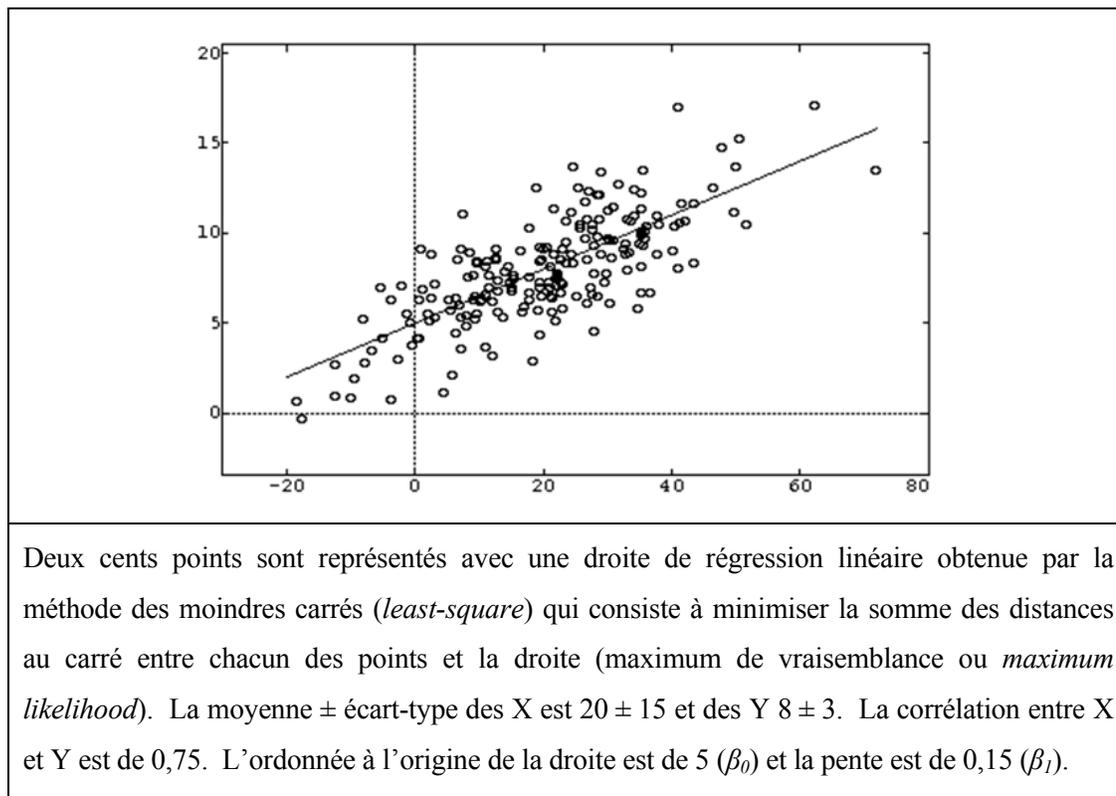
collagène (« Marfan ») sont décédés dans cette cohorte, cette variable « Marfan » risque d'avoir beaucoup de poids dans l'analyse multivariée. En présence d'une telle variable, je cherche à élucider le mécanisme expliquant cette augmentation du risque de décès pour ces patients. Ceci peut être difficile si peu de patients présentent la variable « Marfan » ou aucune autre variable ne représente le même concept de fragilité des tissus de l'aorte, comme par exemple un marqueur tissulaire qui aurait été mesuré pour tous les sujets de l'étude. Comme l'observation d'une augmentation du risque de décès pour les sujets « Marfan » est basée sur peu de patients et peut être le fruit du hasard, il est souhaitable d'éliminer cette variable du modèle pour que celui-ci soit plus généralisable. Je suspecte une telle variable lorsque peu de patients présentent cette caractéristique, que les intervalles de confiance pour les rapports de cote (*odds ratio*) sont grands ou que l'analyse avec "*bootstrapping*" ne donne pas les mêmes résultats (voir section sur le "*bootstrapping*").

Ce phénomène de « *leverage* » se voit aussi pour les variables continues, particulièrement si on ne tient pas compte si elles sont bel et bien distribuées selon la loi Normale. La variable « *creatinin* » comme marqueur de la fonction rénale est un bon exemple. Celle-ci est souvent distribuée de façon asymétrique avec la plupart des patients ayant une valeur de créatinine dans les limites de la normale et quelques patients avec des valeurs de créatinine très élevées (insuffisants rénaux avec ou sans dialyse). Si on utilise la variable créatinine sans la transformer (transformation logarithmique par exemple), les patients insuffisants rénaux décédés auront plus de poids dans l'analyse qu'ils ne le devraient et les estimés des rapports de cote deviennent imprécis (i.e. large intervalles de confiance).

Pour une régression linéaire, on a $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$, où :

- Y est le résultat ou variable dépendante ;
- β_0 est le coefficient de référence lorsque toutes les variables sont absentes (*baseline*) et $\beta_1, \beta_2, \dots, \beta_p$ sont les p coefficients mesurant la force relative de la relation entre X_1, X_2, \dots, X_p et Y . Dans le cas où il n'y a qu'une seule variable indépendante (X_1), β_0 représente aussi l'ordonnée à l'origine de la droite de régression linéaire et β_1 la pente (Figure 8) ;
- ε est la variabilité pour Y qui n'est pas expliquée par X_1, X_2, \dots, X_p .

Figure 8. Exemple de nuage de point (*scatter-plot*) et de régression linéaire^h.



^h Pas de copyright. Image libérée dans le domaine public sur le Wikipedia (http://en.wikipedia.org/wiki/File:Linear_regression.png).

$\beta_0, \beta_1, \beta_2, \dots, \beta_p$ sont estimés à partir des données disponibles chez les sujets de l'étude. Puisqu'ils approximent ce que l'on s'attend à retrouver pour l'ensemble de la population, on les dénote avec un chapeau par $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$. Pour un modèle univarié, l'analyse de régression cherche à déterminer s'il y a une association entre X_1 et Y . L'hypothèse nulle (H_0) est l'absence de relation entre X_1 et Y , c'est-à-dire $\beta_1 = 0$. Par exemple, pour un modèle évaluant l'association entre la hauteur d'un sujet (résultat) et son poids (facteur de risque), le poids d'un sujet peut permettre de prédire partiellement quelle est la hauteur de celui-ci, mais si le patient est obèse, musclé ou maigre, la hauteur peut varier considérablement pour un sujet d'un même poids. Ainsi, il est indiqué d'ajouter au modèle d'autres variables telles que le sexe du sujet et son tour de taille pour améliorer la prédiction du modèle.

Une interaction entre deux variables existe lorsque l'association entre ces deux variables est différente à divers niveaux d'une troisième variable. En termes épidémiologique, on parle de modification d'effet (*effect modification*). Par exemple, il est connu que l'effet bénéfique des médicaments hypolipémiants (statines) est plus grand pour les sujets avec antécédent de maladie coronarienne que la population générale. Dans l'équation suivante, l'interaction entre les variables X_1 et X_2 est représentée par X_1X_2 : $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2$. Si le coefficient β_3 est statistiquement différent de zéro, on est en présence d'une interaction entre ces variables. L'interaction est difficile à évaluer dans les modèles statistiques utilisés en recherche clinique lorsque le nombre de sujets dans la cohorte est petit ou que l'interaction identifie un sous-groupe de quelques sujets seulement.

PROPENSITY SCORE

Le “*propensity score*” (PSc) est la probabilité pour un sujet donné de recevoir un traitement vs un autre traitement (réel ou placebo) conditionnellement à (ou basé sur) un ensemble de variables observées chez le sujet en question (Table 1). Il permet de regrouper dans un score unique une quantité considérable d’information. Il s’avère fort utile pour tenir compte des différences, observables à la base, entre les groupes comparés et ainsi permet de réduire les biais générés par celles-ci dans l’estimation de l’effet du traitement. Il trouve son application surtout dans les études observationnelles non-randomisées puisque les études randomisées tiennent compte des différences observées, et non observées, par leur devis d’étude.

Le PSc a été proposé initialement par Rosenbaum et Rubin¹⁰³ dans un langage épidémiologique très technique incluant les théorèmes et preuves qui justifient son usage. La rationnelle du PSc se comprends mieux en utilisant le concept d’analyse de cause à effet¹⁰⁴. En théorie, pour être en mesure de formuler une relation de cause à effet, on doit pouvoir comparer le résultat que l’on aurait obtenu si le sujet avait reçu le traitement A vs le résultat que l’on aurait obtenu si ce même sujet avait reçu le traitement B. Dans la plupart des devis d’étude, on ne peut observer qu’une des réponses par sujet (réponse au traitement A ou B) et on doit spéculer sur la réponse qu’aurait produit l’autre traitement. Les analyses de cause à effet peuvent être considérées comme un problème de donnée manquante (*missing data*) car on ne sait pas avec certitude quelle aurait été la réponse au traitement B si le patient a eu le traitement A. La Table 1 présente un exemple simple de PSc.

Table 1. Exemple simple de “*propensity score*” pour deux groupes fictifs.

Sujet	Groupe	Âge	Diabète	FEVG	PSc_log	PSc_ad
1	A	45	Oui	>50%	0,650	0,643
2	A	31	Non	>50%	0,832	0,807
3	A	47	Non	<40%	0,536	0,535
4	A	71	Oui	<40%	0,231	0,267
5	A	55	Non	>50%	0,673	0,666
6	A	48	Non	>50%	0,727	0,712
7	A	64	Oui	>50%	0,481	0,501
8	A	51	Oui	<40%	0,384	0,403
9	A	66	Non	40%-50%	0,533	0,488
10	A	62	Oui	40%-50%	0,453	0,417
11	B	76	Oui	<40%	0,200	0,238
12	B	53	Non	<40%	0,481	0,488
13	B	52	Oui	40%-50%	0,544	0,494
14	B	71	Oui	>50%	0,418	0,448
15	B	67	Oui	<40%	0,258	0,292
16	B	56	Non	>50%	0,665	0,659
17	B	50	Oui	<40%	0,393	0,410
18	B	69	Non	>50%	0,553	0,565
19	B	60	Oui	40%-50%	0,471	0,433
20	B	36	Oui	<40%	0,519	0,517

FEVG : fraction d'éjection du ventricule gauche. Le “*propensity score*” a été calculé à l'aide de la régression logistique multivariée (PSc_log) et l'analyse discriminante (PSc_ad).

Le tableau rapporte l'âge des sujets (variable continue), la présence de diabète (variable binaire) et la classe de FEVG (variable ordinale). Les sujets du groupe A sont plus jeunes, présentent moins de diabète et ont de meilleures FEVG en moyenne que le groupe B. Le PSc est une prédiction basée sur un modèle statistique qui nous renseigne ici sur les sujets présentant davantage des attributs du groupe A (PSc près de 1) vs du groupe B (PSc près de 0). Les PSc obtenus par régression logistique multivariée et analyse discriminante sont similaires. Le sujet 2 du groupe A est le plus jeune, n'est pas diabétique et a une FEVG > 50%, son PSc est près de 1 ; par opposition, le sujet 11 du groupe B est le plus âgé, il est diabétique et présente une FEVG < 40%, son PSc est près de 0. Les sujets 7 et 12 sont du groupe A et B, respectivement, et ont des PSc semblables bien que leurs caractéristiques soient différentes. Le sujet 7 est plus âgé et diabétique mais conserve une bonne FEVG (>50%), alors que le sujet 12 est plus jeune, non-diabétique mais présente une pauvre FEVG (<40%). Puisque ces sujets ont des profils de caractéristiques qui les rapprochent autant des caractéristiques propres au groupe A que du groupe B, leurs PSc sont près de 0,5.

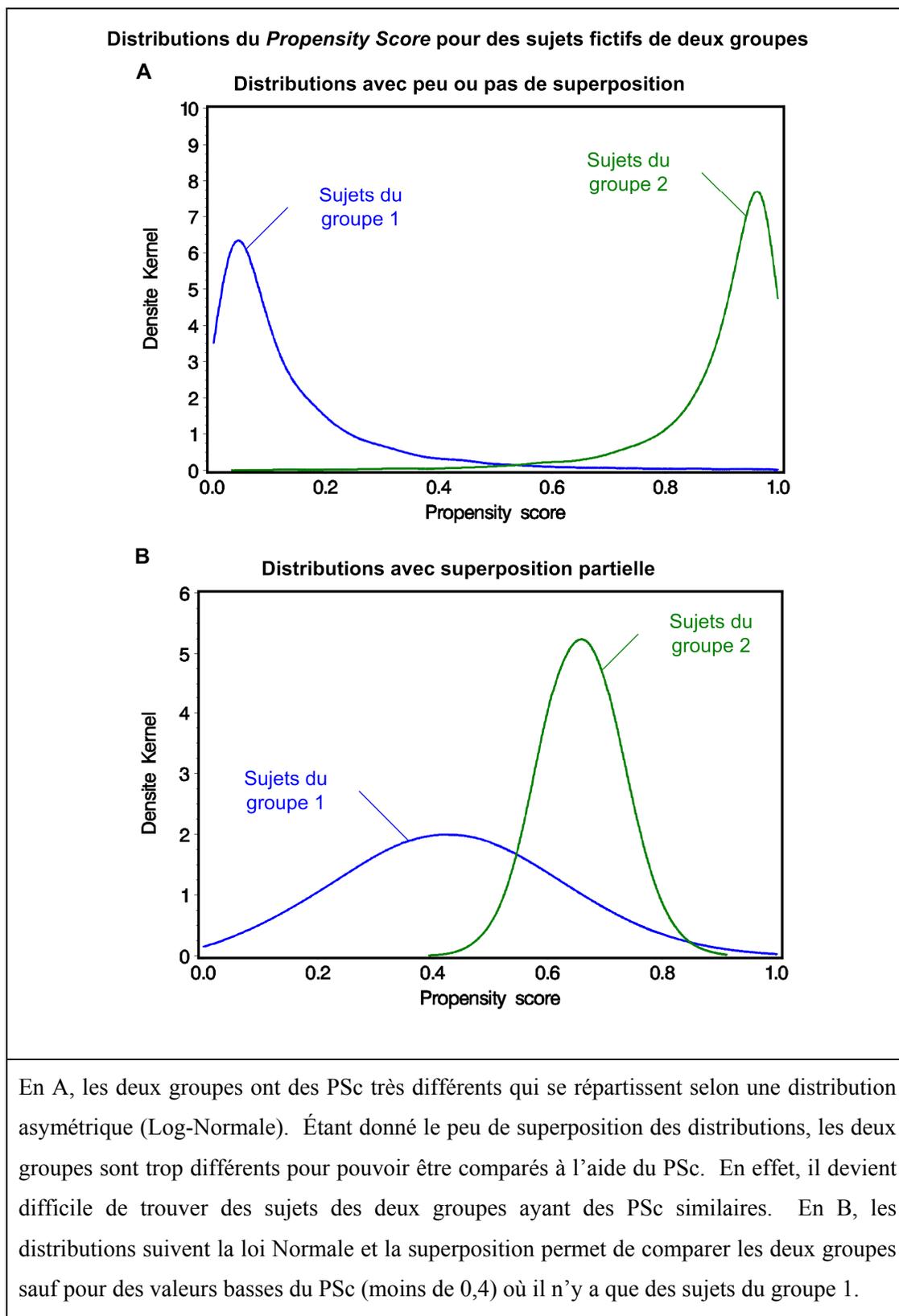
Dans une étude randomisée, les groupes de sujets ont de fortes chances d'être comparables pour les variables observées, de même que pour les variables non observées. Par exemple, le niveau socio-économique des participants dans chacun des groupes va tendre à être similaire même si l'étude ne collecte pas de données sur ces paramètres (non observée, non mesurée). La randomisation des sujets aux différents traitements garantit qu'en moyenne il ne devrait pas y avoir de différence systématique, observée ou non, entre les groupes de sujets assignés à chacun des traitements. Dans le contexte d'une étude randomisée, le PSc est utile lorsque l'on désire comparer l'effet de traitement selon le traitement reçu et non selon le traitement assigné (puisque certains sujets peuvent changer de traitement en cours d'étude – en anglais, *crossover*) ou pour réaliser certaines sous-études à partir des données de l'étude principale. Dans une étude observationnelle non-randomisée, on retrouve souvent des différences systématiques dans les caractéristiques des sujets selon le traitement attribué. Puisque les chercheurs n'ont pas de contrôle sur l'attribution des traitements, la comparaison directe des effets du traitement entre chaque groupe risque de mener à des conclusions trompeuses. L'utilisation d'un PSc permet de réduire ces biais et d'améliorer la précision des effets de traitement rapportés.

Le *balancing score* décrit par Rosenbaum et Rubin¹⁰³ est fonction des variables observées si bien que la distribution de ces variables, conditionnelle au *balancing score*, est la même pour les sujets recevant le traitement A vs ceux recevant le traitement B. En d'autres mots, en tenant compte du *balancing score*, les groupes deviennent « similaires » pour les variables observées et on obtient un estimé potentiellement non biaisé de l'effet de traitement A comparé au traitement B.

Pour obtenir un estimé non biaisé, il est important que les variables observées prennent en considération les facteurs principaux qui influencent l'attribution du traitement pour que les sujets soient effectivement similaires pour un niveau donné du *balancing score*. Si une variable importante est omise (« non-observée »), le *balancing score* ne peut en tenir compte et l'estimé obtenu est biaisé. Le PSc est la forme la plus grossière du *balancing score* puisqu'il résume toutes les données en une probabilité allant de 0 à 1 (Table 1).

La Figure 9 présente des distributions fictives du PSc avec ou sans superposition. Par exemple, si le PSc est défini par l'âge des sujets et que tous les sujets de plus de 65 ans reçoivent le traitement 1 (groupe 1) et qu'aucun sujet de 65 ans ou moins ne reçoit le traitement 1 (groupe 2), il n'est pas possible de comparer le résultat que le sujet de 75 ans a connu avec le traitement 1 avec celui qu'il aurait connu sans le traitement 1 (Figure 9A). Aussi, dans la situation où tous les sujets « instables » sont opérés (groupe 1) et que les sujets « stables » peuvent être traités de façon chirurgicale (groupe 1) ou médicale (groupe 2) (Figure 9B), il convient de redéfinir la population à l'étude si l'on désire comparer les patients du groupe 1 vs 2 en excluant les sujets « instables » pour lesquels il n'est pas possible de comparer leurs résultats avec des sujets « instables » qui n'ont pas été opérés. Les valeurs extrêmes du PSc (près de 0 ou de 1), importent peu en autant que les distributions du PSc soient suffisamment superposables pour que les groupes puissent être comparées (Figure 10). Si le PSc est distribué selon la loi Log-Normale et que les distributions sont suffisamment superposables, j'utilise une transformation logarithmique (\log) du PSc avant d'utiliser celui-ci dans mes modèles.

Figure 9. Exemples de distributions du PSc.



Rosenbaum et Rubin exposent cinq théorèmes¹⁰³ avec des preuves détaillées à partir de règles statistiques et épidémiologiques. Les conclusions sont les suivantes :

- i. Le PSc est un *balancing score*.
- ii. Tout score plus fin qu'un PSc est un *balancing score* ; de plus, l'expression la plus détaillée des variables utilisées pour estimer le PSc, ou distribution, définissent le *balancing score* le plus fin (comme si l'âge dans l'exemple en Table 1 était représenté en variable ordinale au lieu de variable continue) et le PSc est le *balancing score* le plus grossier.
- iii. Si l'attribution du traitement est aléatoire une fois que l'on a tenu compte de toutes les variables observées, elle est également aléatoire si on tient compte du *balancing score*.
- iv. À n'importe quel niveau du *balancing score*, la différence entre les résultats pour le traitement A et le traitement B est un estimé non biaisé de l'effet de traitement moyen à cette valeur du *balancing score* si l'attribution du traitement est aléatoire à ce niveau. Ceci justifie l'utilisation des techniques d'appariement (*matching*), stratification et ajustement par régression (voir ci-dessous).
- v. En tenant compte du *balancing score*, on obtient une distribution similaire des variables observées pour les deux groupes. Donc, en moyenne, les groupes deviennent similaires pour les variables observées qui font partie du *balancing score*.

Le PSc est souvent estimé par régression logistique. Cette technique consiste simplement à créer un modèle de régression logistique multivariée prédisant le traitement reçu (groupe A vs B) en incluant les variables observées pertinentes dans le modèle et, au besoin, des interactions entre ces variables, des transformations quadratiques, etc. Plusieurs variables peuvent être incluses dans ces modèles car le nombre d' « événements » (traitement A vs traitement B) est grand, contrairement au modèle prédisant le résultat d'intérêt (où l' « événement » que le modèle cherche à prédire peut être défini comme une mortalité, un infarctus, ... et où le nombre d' « événements » est habituellement plus petit). Les modèles de régression logistique multivariée peuvent inclure classiquement une variable indépendante par dix « événements ». Ainsi, à chaque 10 sujets faisant partie du groupe A, on peut inclure une variable explicative au modèle de régression logistique. Enfin, l'équation finale tirée du modèle de régression logistique multivariée est utilisée pour prédire la probabilité de recevoir le traitement A vs traitement B pour chacun des sujets, i.e. leur PSc, peu importe le groupe dont ils font partie réellement.

Le modèle de régression logistique simple utilisé pour l'exemple de la Table 1 se résume par l'équation suivante :

$$\Pr(\text{GroupeA}|X) = \frac{1}{(1 + e^{-\beta'x})} = \frac{1}{(1 + e^{-(\beta_0 + \beta_1 \text{Age} + \beta_2 \text{Diabete} + \beta_3 \text{FEVGbas} + \beta_4 \text{FEVGmid})})}, \text{ où :}$$

- $\beta'x$ est un vecteur de coefficients dont chacun est associé à une des variables du modèle. Il se traduit par l'équation suivante pour l'exemple de la Table 1 : $\beta_0 + \beta_1 \text{Age} + \beta_2 \text{Diabete} + \beta_3 \text{FEVGbas} + \beta_4 \text{FEVGmid}$.
- Sans vouloir entrer dans les détails de la théorie des matrices, $\beta'x$ est

- simplement la notation matricielle utilisée pour désigner $\sum_{i=1}^p \beta_i X_i$, où β' représente le vecteur des coefficients $\beta_1, \beta_2, \dots, \beta_p$ et x le vecteur des variables X_1, X_2, \dots, X_p . Elle est utilisée ici pour simplifier les équations ;
- $\Pr(\text{Groupe A} | X)$ est la probabilité (entre 0 et 1) pour un sujet de faire partie du groupe A étant donné ses caractéristiques propres dénotées par x ;
 - $\frac{1}{(1 + e^{-\beta'x})}$, aussi appelé « *logit* », est l'expression mathématique du modèle de régression logistique dont le résultat est compris à l'intérieur de 0 et 1 (plus l'expression $e^{-\beta'x}$ est grande, plus le résultat se rapproche de 0 et plus l'expression $e^{-\beta'x}$ est petite, plus le résultat se rapproche de 1) ;
 - On remarque que deux variables ont été créées pour la variable catégorielle FEVG. Lorsque le patient présente une FEVG normale (groupe de référence), les deux variables FEVG du modèle sont égales à zéro, si la FEVG est entre 40 et 50%, $FEVG_{mid}$ est égal à 1 (et $FEVG_{bas}$ est égal à 0) et enfin lorsque la FEVG est basse (<40%), $FEVG_{bas}$ est égal à 1 (et $FEVG_{mid}$ est égal à 0).

Le PSc est également estimable par analyse discriminante (*discriminant analysis*), une technique plus ancienne qui était particulièrement utile avant la survenue des logiciels de statistiques et la présence d'ordinateurs de plus en plus performants. La régression logistique est maintenant plus souvent utilisée que l'analyse discriminante puisque qu'elle présente moins de risque de faillir aux

suppositions du modèle (les variables indépendantes n'ont pas à être distribuées selon la loi Normale, reliées linéairement, ou de présenter des variances “*within-group*” égales), elle est robuste, tient compte autant des variables catégorielles que continues, et les coefficients produits sont plus facile à interpréter. L'analyse discriminante peut être utile lorsque les suppositions de régression linéaire sont acceptables puisqu'elle a alors plus de puissance statistique que la régression logistique (moins de chance d'une erreur de type II, i.e. d'accepter une fausse hypothèse nulle). Par contre, en présence de variables indépendantes qui ne sont pas distribuées selon la loi Normale, l'analyse discriminante a tendance à surestimer la magnitude des estimés des coefficients du modèle. Puisque l'analyse discriminante est rarement utilisée en pratique de nos jours, elle ne sera pas détaillée ici⁶².

Techniques d'application du “*propensity score*”

Il y a trois principales techniques d'application du PSc :

- 1) l'appariement ;
- 2) la stratification ;
- 3) l'ajustement par régression.

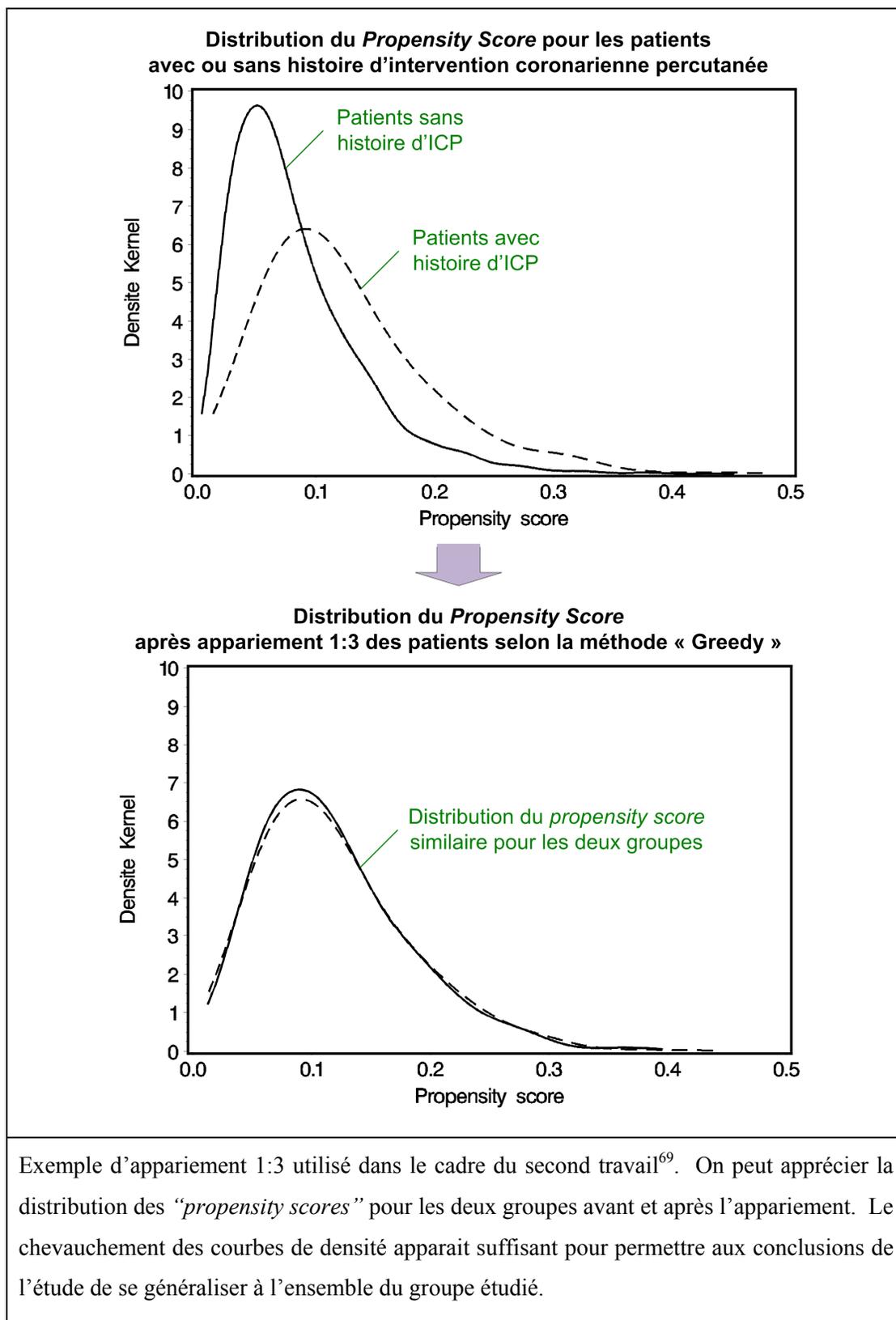
Ces techniques d'application sont décrites en plus amples détails au Chapitre IV, avec des exemples concrets à l'appui. Je présente ci-dessous les grandes lignes de chacune de ces techniques d'application du PSc avec leurs avantages et inconvénients respectifs.

Appariement (matching)

La technique d'appariement (*matching*) est la technique la plus simple. On choisit au hasard un sujet ayant reçu le traitement A et on cherche le sujet ayant reçu le traitement B qui a un PSc similaire à l'intérieur d'une étendue de PSc définie à l'avance (par exemple, $\pm 0,001$). Ces deux sujets sont alors mis de côté et on recommence jusqu'à ce que l'on ne soit plus en mesure de trouver de sujets ayant un appariement satisfaisant. On peut ainsi effectuer l'appariement un pour un, un pour deux, etc. s'il y a moins de sujets pour un groupe par rapport à l'autre (Figure 10). Le plus souvent, les sujets qui ne sont pas appariés sont ceux qui n'ont pas de sujet comparable dans l'autre groupe et pour lesquels il est difficile conceptuellement d'estimer ce qu'aurait été le résultat s'ils avaient fait partie de l'autre groupe. La technique d'appariement selon le PSc est plus efficace que l'appariement selon plusieurs variables prises séparément car il peut être difficile de trouver deux sujets ayant exactement le même profil de variables.

L'appariement peut se faire de façon automatisée par des programmes d'appariement où on a défini un intervalle d'appariement. Le programme sélectionne la paire la plus proche (*nearest available match*) pour chaque sujet. Les techniques d'appariement "*greedy*" utilisent un processus itératif en cherchant d'abord les meilleures paires en appariant à 5 décimales, puis à 4 décimales, etc., jusqu'à une décimale. Un processus itératif est une séquence d'instructions destinée à être exécutée/répétée autant de fois que l'on en a besoin.

Figure 10. Distribution du “*propensity score*” et effet de l’appariement 1:3⁶⁹.



La technique d'appariement présente plusieurs avantages :

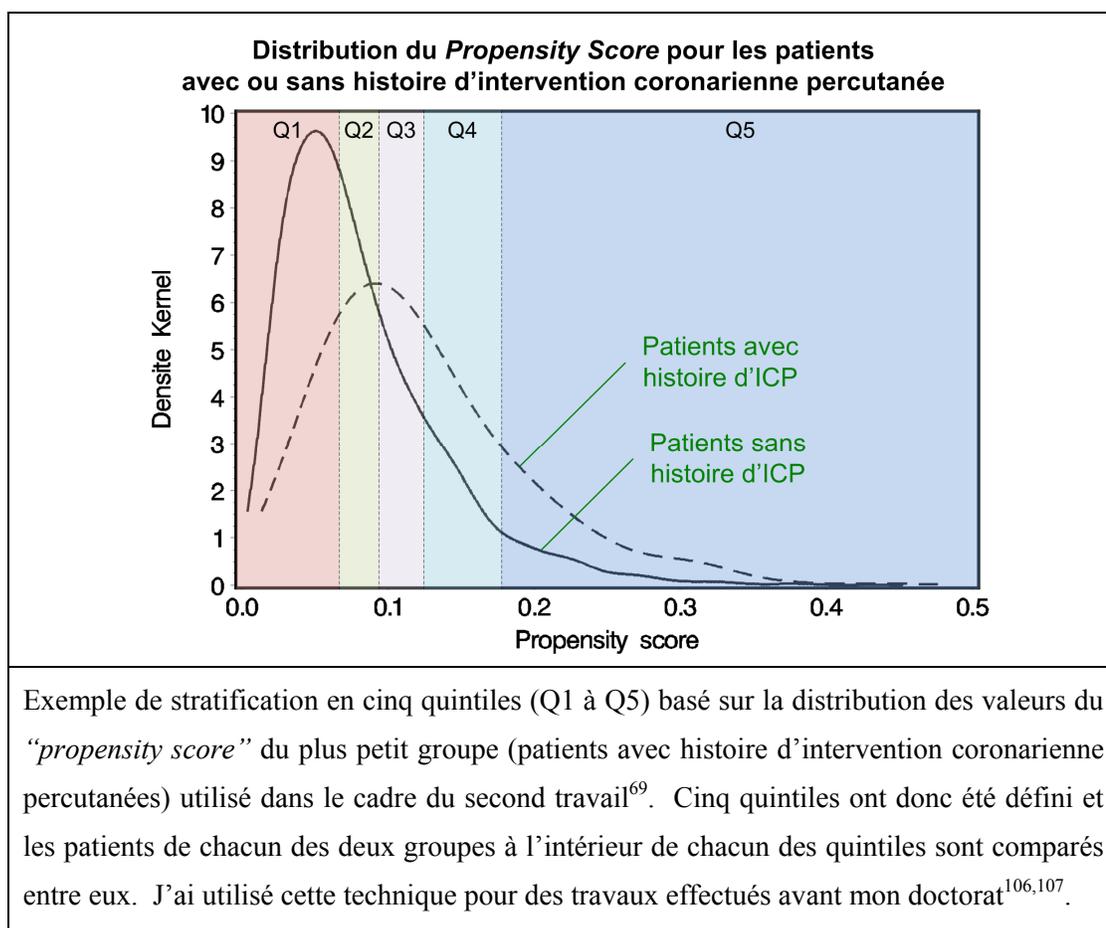
- Elle est simple à saisir conceptuellement pour un chercheur avec des connaissances statistiques de base ;
- Elle permet d'obtenir deux groupes chez lesquels il est facile de démontrer les similitudes par rapport aux variables observées « à l'œil » ou à l'aide de méthodes statistiques simples à calculer (différence standardisées) ;
- La variance des estimés d'effet de traitement est moins grande que si les groupes avaient été échantillonnés au hasard du fait que la distribution des variables est similaire dans chacun des groupes. Par contre, on doit en tenir compte dans l'analyse par l'utilisation de méthodes statistiques appropriés aux groupes appariés (tests de McNemar et de t apparié [*paired t test*], régression logistique inconditionnelle, *generalized estimating equation*) ;
- Elle est plus robuste que les techniques basées sur la modélisation mathématique (ajustement par régression) car elle se trouve à exclure les sujets qui n'ont pas de sujet comparable dans l'autre groupe.

L'inconvénient de l'appariement est que l'on se trouve parfois à exclure un nombre important de sujets (selon la distribution des PSc ; voir Figure 10). Les conclusions de nos analyses portent alors seulement sur le sous-groupe de patients appariés et ne reflètent pas nécessairement l'ensemble du groupe que l'on souhaite étudier. Les sujets ayant un PSc dans les extrêmes de la distribution de celui-ci sont souvent difficiles à appairer. Si on souhaite que les résultats des analyses reflètent le groupe en entier, on doit supposer que l'effet du traitement n'est pas différent, ou modifié, chez les sujets qui ne peuvent être appariés, ce qui peut parfois être irréaliste – par exemple, si on compare un traitement chirurgical vs médical alors que l'alternative médicale n'est pas envisageable pour certains sujets.

Stratification

La technique de stratification consiste habituellement à diviser le groupe en quintiles selon le PSc et de comparer les effets du traitement à l'intérieur de chaque quintile (Figure 11). Il a été démontré que le biais est réduit de 90% par l'utilisation de quintiles¹⁰⁵. On peut également obtenir un estimé général de tout le groupe en combinant les estimés obtenus au sein de chaque quintile par des méthodes de moyenne pondérées (*Mantel-Haenszel weighted average methods*). La technique est parfois limitée par le nombre de patients de chaque groupe retrouvé dans chaque quintile et elle est moins intuitive que l'appariement où deux groupes sont comparés.

Figure 11. Exemple de distribution du “propensity score” et effet de la stratification⁶⁹.



Ajustement par régression

La technique d'ajustement par régression consiste à inclure le PSc (valeur de 0 à 1), ou une transformation de celui-ci, comme une des variables du modèle multivarié évaluant l'effet de traitement, et ce au même titre que les autres variables indépendantes (facteurs de risque)⁷⁰. Il est important que la forme du PSc, ou une transformation de celui-ci, soit bien spécifiée dans le modèle d'ajustement par régression (i.e. on doit tenir compte de sa distribution), sinon on court le risque d'obtenir un estimé biaisé. Pour ce faire, je vérifie si les résidus du modèle d'ajustement par régression ont une distribution Normale. Un résidu se définit par la différence entre la valeur prédite et obtenue calculée pour chaque sujet (autrement dit le terme qui n'est pas expliqué par les autres variables). L'avantage de la technique d'ajustement par régression est qu'elle permet d'inclure tous les sujets de l'étude (contrairement à l'appariement). En théorie, cette technique est plus puissante pour tenir compte du PSc, en autant que le modèle soit bien spécifié. Certains auteurs suggèrent d'inclure le PSc avec d'autres variables indépendantes importantes (ex. : âge, diabète, etc.), même si celles-ci font partie du PSc^{9,10}. Ces auteurs croient que l'on contrôle de façon plus robuste pour ces variables importantes, surtout lorsque le PSc n'est résumé que par une variable continue. En pratique, l'ajout de ces variables, en plus du PSc, a habituellement peu d'effet si le modèle incluant le PSc est bien spécifié. D'autres suggèrent d'inclure les variables qui différencient surtout les deux groupes pour l'estimation du PSc, alors que l'on réserve les variables qui sont davantage associées avec le résultat étudié (ex. : décès) dans le modèle final¹⁰⁸.

Ajustement par régression pondérée (weighted regression)

La technique d’ajustement par régression pondérée attribue un poids relatif pour chaque sujet en fonction de leur PSc. Les avantages de cette technique sont les suivants : 1) elle permet d’utiliser tous les sujets ; 2) elle permet d’étudier des modèles complexes de relations de cause-à-effet et d’analyses avec variables variant dans le temps ; et 3) elle permet de standardiser en fonction de la population sur laquelle on désire que nos inférences s’appliquent. Les inconvénients sont : 1) qu’elle est difficile à conceptualiser ; 2) que les erreurs-type peuvent être difficile à estimer ; 3) et que cette méthode est très sensible aux patients qui se situent aux extrêmes de la distribution du PSc. Une première méthode de pondération utilise l’estimateur “*inverse-probability-of-treatment*” (IPTW) : les patients avec le traitement A sont pondérés (*weight*) en utilisant $\left[\frac{1}{PSc} \right]$ alors que les sujets avec le traitement B sont pondérés avec $\left[\frac{1}{(1 - PSc)} \right]$. Conceptuellement, on compare l’effet attendu en traitant tous les sujets avec le traitement A vs l’effet attendu en traitant tous les patients avec le traitement B. Une seconde méthode utilise l’estimateur “*standardized mortality/morbidity ratio*” (SMR) où les patients ayant le traitement A sont pondérés avec [1] alors que l’on utilise un “*propensity odds*” pour les non-traités, soit $\left[\frac{PSc}{(1 - PSc)} \right]$ ¹⁰⁹. Conceptuellement, on se demande ce qui serait survenu si on avait utilisé le traitement B pour tous ceux qui ont reçu le traitement A.

Dans la plupart des cas, peu importe la méthode d'application du PSc utilisée, les résultats obtenus sont similaires entre eux et aux méthodes traditionnelle de régression sans ajustement avec le PSc¹¹⁰⁻¹¹². Lorsque ceux-ci diffèrent, on doit se demander ce que nos analyses étudient vraiment (par exemple, un sous-groupe de la population totale pour la technique d'appariement)¹¹³. La révision des méthodes d'application du PSc dans les articles de recherche clinique publiés a démontré que la méthode d'ajustement par régression est la plus utilisée, bien que les auteurs de ces articles indiquent rarement s'ils ont vérifié si le modèle était bien spécifié¹¹⁰⁻¹¹². Lorsque la relation entre le PSc et le résultat est linéaire, un ajustement direct peut être réalisé en ajoutant simplement le PSc comme une variable continue dans le modèle de régression du résultat d'intérêt. Dans ce cas, l'ajustement par régression est préférable à la stratification ou l'appariement car on est en mesure d'estimer l'effet du traitement avec une variance plus faible tout en tenant compte des biais statistiques^{103,104}. Lorsque la relation du PSc avec le résultat n'est pas linéaire, l'ajustement par régression demande plus de soins (par exemple, on peut avoir à utiliser une transformation du PSc, etc.) ou une autre méthode d'application doit être utilisée. Tel que présenté ci-haut, chaque méthode d'application a ses avantages et ses inconvénients. Des exemples concrets d'applications du PSc sont donnés dans le Chapitre IV. J'ai utilisé la technique du PSc pour quatre travaux de recherche à ce jour^{69,70,106,107}.

ANALYSES DE SURVIE

L'analyse de survie s'intéresse à l'analyse de résultats de temps jusqu'à un évènement, c'est-à-dire l'analyse de variable dépendante de valeur positive représentant un intervalle de temps à partir d'un moment défini (chirurgie, randomisation, etc.) jusqu'à un évènement d'intérêt (décès, survenue d'un premier infarctus, rémission pour un cancer, etc.). On parle de « temps de survie », puisque le sujet a « survécu » un certain temps jusqu'à l'évènement d'intérêt. La distribution des temps de survie est habituellement asymétrique avec la queue sur le côté droit de la distribution plus longue que le côté gauche, qui est limité à zéro, et la plupart des valeurs (y compris la médiane) se trouvant à gauche de la moyenne (*positive or right-skewed*). Puisque l'évènement a souvent une connotation négative (décès, complication), on caractérise souvent l'évènement d'« échec » au suivi (*failure*). On assume que seulement un évènement est noté par sujet, bien que des modèles plus avancés permettent l'inclusion de plus d'un évènement par sujet (modèle à risque concurrent [*competing risk*] et modèles longitudinaux). Des modèles similaires d'analyse de survie sont utilisés en ingénierie (*failure-time analysis, reliability analysis*), sociologie (*event-history analysis*) et économique (*transition analysis*).

L'objet d'intérêt principal de l'analyse est la fonction de survie, conventionnellement noté $\hat{S}(t)$, qui est défini comme : $\hat{S}(t) = P(T > t)$, où :

- P est une probabilité ;
- T est une variable aléatoire désignant le temps de survie ;
- t est un point ou moment spécifié dans le temps.

Dit autrement, la fonction de survie est la probabilité que le sujet survive plus longtemps qu'un certain temps (t) spécifié. La fonction de survie décroît nécessairement dans le temps : $\hat{S}(u) \leq \hat{S}(t)$ si $u > t$. Cela reflète l'idée que la survie à un âge plus avancé n'est possible que si le sujet a survécu aux âges plus jeunes. Au temps zéro, avant même que le premier évènement ne survienne, la probabilité de survie est 100% [$\hat{S}(t) = 1$]. La fonction de survie est généralement supposée tendre vers zéro alors que t augmente à l'infini [$\hat{S}(t) \rightarrow 0$ si $t \rightarrow \infty$] à moins que l'on admette que la vie éternelle soit possible pour un modèle de survie donné (i.e. certains matériaux inertes) ou que les patients décèdent avant de présenter l'évènement d'intérêt lorsqu'il ne s'agit pas de décès (*competing risk*).

La fonction de risque instantané (*hazard function*) est une mesure du risque que l'évènement étudié survienne par unité de temps, habituellement un intervalle de temps très court, parmi les sujets qui ont survécu jusqu'au temps t . Il correspond donc à un taux (compris entre 0 et l'infini) et non une probabilité (comprise entre 0 et 1). Il s'exprime mathématiquement comme ceci :

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \left[\frac{P(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} \right], \text{ où :}$$

- $\lambda(t)$ est la fonction de risque ;
- Δt est un intervalle de temps très court, qui tend vers zéro.

L'intérêt de la fonction de risque par rapport à la fonction de survie est que :

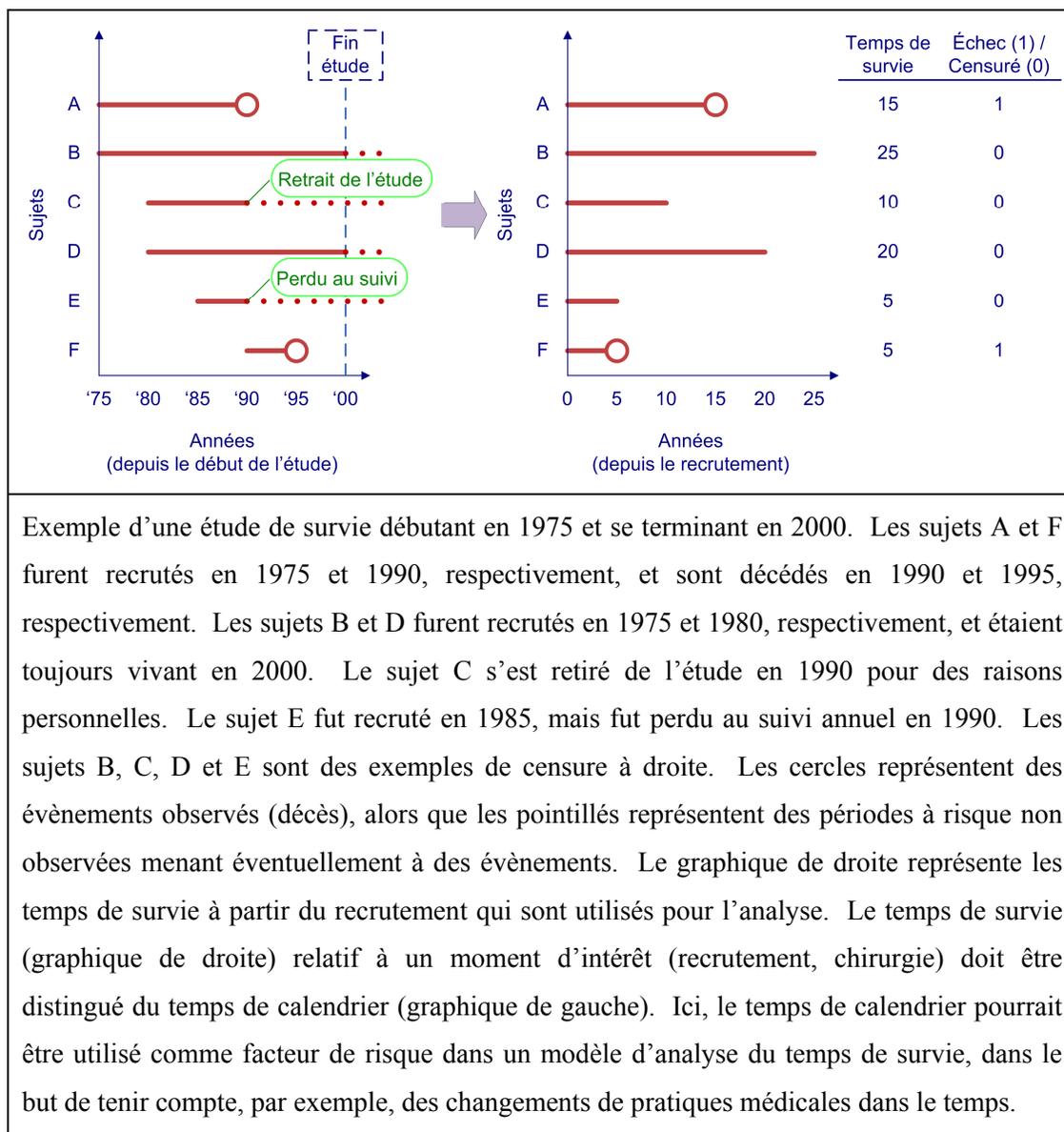
- Elle donne une mesure du risque instantané au lieu d'une mesure cumulative ;

- Elle peut être utilisée pour identifier un modèle spécifique d'analyse paramétrique qui convient mieux aux données (exponentiel, Weibull, Log-Normale ...). Le modèle de régression paramétrique exponentiel est un modèle particulier où le risque instantané est constant dans le temps, ce qui n'est habituellement pas le cas avec les autres modèles paramétriques ou le modèle semi-paramétrique de Cox ;
- Elle permet la modélisation mathématique des données de survie ;
- Si on connaît la fonction de risque, on peut dériver mathématiquement la fonction de survie, et vice versa.

Censure

Une des caractéristiques importantes de l'analyse de survie est que le temps de survie peut être censuré pour certains sujets (*censoring*). Pour les sujets censurés, on possède donc un peu d'information sur leur temps de survie - jusqu'au moment où le suivi s'arrête - mais on ne connaît pas exactement le réel temps de survie car l'évènement n'a pas encore eu lieu. Il y a censure « fixe » si le sujet ne présente pas l'évènement avant la fin du suivi (*fixed censoring*), ou censure aléatoire (*random censoring*) si le sujet est perdu au suivi durant la période d'étude, ou si un sujet se retire de l'étude en raison d'un décès (si l'évènement d'intérêt n'est pas le décès), ou pour une autre raison (effet secondaire, raison personnelle). On parle de censure à droite (*right-censoring*), puisque l'évènement d'intérêt se situe, en théorie, « à droite » dans le continuum temps après la censure (Figure 12).

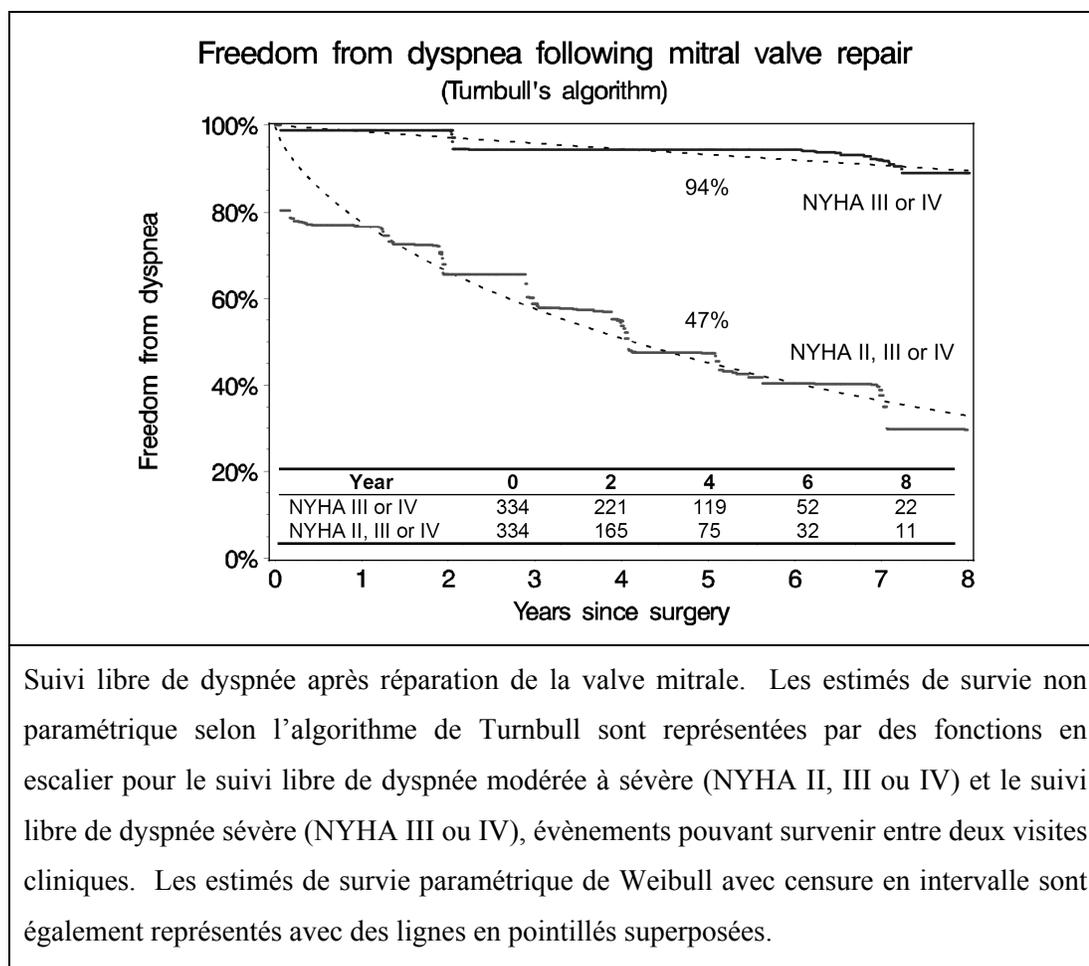
Figure 12. Censure à droite (*right-censoring*).



Dans d'autres devis d'étude, il est possible d'être en présence de censure à gauche (*left-censoring*) lorsque l'évènement d'intérêt est survenu avant que le monitoring ne débute : par exemple, si l'on s'intéresse au temps jusqu'à la perte de la première dent pour des enfants à la maternelle, il est possible que certains d'entre eux aient déjà perdu une dent avant même d'être en âge pour entrer en maternelle. Enfin,

on parle de censure en intervalle (*interval-censoring*) si l'évènement survient quelque part entre deux rendez-vous de suivi : par exemple, la survenue de fuite sévère d'une valve cardiaque entre deux échocardiographies de suivi. J'ai étudié ce dernier type de censure en intervalle pour un travail portant sur les résultats des patients suivant une réparation de la valve mitrale¹¹⁴ (Figure 13). J'avais alors utilisé un modèle non-paramétrique utilisant l'algorithme de Turnbull¹¹⁵ ainsi qu'un modèle paramétrique (Weibull¹¹⁶) me permettant de tenir compte de l'intervalle entre deux visites annuelles dans l'estimation des courbes de survie.

Figure 13. Exemples d'estimés de survie avec censure en intervalle¹¹⁴.



Pour que le modèle d'analyse de survie soit approprié, il importe que la censure soit indépendante du résultat et des facteurs de risque étudiés. On dit que la censure est « non-informative » par rapport au résultat, i.e. que le sujet qui est censuré au temps t est représentatif des autres sujets qui ont survécu jusqu'au temps t et qui n'ont pas été censuré. Il n'est donc pas adéquat de censurer un sujet qui va mal, ou de censurer un sujet juste avant son décès, parce que dans ces situations, la censure est directement reliée avec le résultat (décès). Il est convenable, par contre, de déterminer une date de fin d'étude commune à tous les sujets (*fixed censoring*).

Dans le cas de censure aléatoire (*random censoring*), il est primordial de la limiter le plus possible par un suivi de près et de s'assurer que les temps de survie ne soient pas dépendant du mécanisme de censure. Il est connu que les patients les plus malades ont tendance à se retirer des études peu avant leur décès. Ceci peut mener à des estimés de survie biaisés vers le haut avec impression d'une meilleure survie que la survie réelle de la population étudiée. Il est possible d'en explorer l'impact en incluant dans les modèles d'analyse des variables reliées à la fois au mécanisme suspecté de censure et au temps de survie, tel que la sévérité de la maladie.

Mesures descriptives

Les trois objectifs de l'analyse de survie sont 1) estimer et interpréter les fonctions de survie et/ou de risque instantané à partir de données de survie ; 2) comparer les fonctions de survie et/ou de risque instantané et 3) évaluer la relation entre des variables et les temps de survie.

Plusieurs méthodes existent pour la construction de courbes de survie :

- La méthode de table de survie (*life table*) où les périodes de calcul du taux de survie sont prédéterminées et de même durée (1 mois, 3 mois, 6 mois, ...). Ainsi, les "marches d'escalier" déterminées par cette méthode sont toutes de longueurs égales ; certains suggèrent même de relier les points par les lignes droites sans marche d'escalier. Cette méthode a pour avantage de limiter le nombre de points et de donner des estimés moyens de survie par intervalle, mais ce faisant elle perd une quantité importante d'information.
- La méthode de Kaplan-Meier (voir ci-dessous), la plus utilisée actuellement, où l'on calcule un taux de survie à chaque survenue d'un épisode de l'événement étudié. On obtient alors des "marches" de longueurs inégales.
- La méthode de Cox (voir ci-dessous) a l'avantage de produire des courbes de survie ajustées selon les variables présentes dans le modèle en autant que celui-ci soit bien spécifié.
- Les méthodes paramétriques (exponentielle, Weibull, etc.). Je les décris en plus amples détails dans la section sur l'analyse de la décomposition des phases de risque (voir ci-dessous).

La médiane de survie (ou survie médiane) correspond au délai après lequel 50% des sujets sont encore vivants alors que le temps de survie moyen (ou survie moyenne, \bar{T}) se définit comme le temps moyen de survie des sujets depuis le début de l'étude. Le nombre de sujet-années se calcule en additionnant les temps de survie pour tous les sujets de l'étude. Le taux de risque moyen (*average hazard rate*, $\bar{\lambda}$) se

calcule simplement en divisant le nombre d'évènements par la somme des temps de survie observés et suppose que le taux de risque est constant dans le temps, ce qui peut être irréaliste pour certains modèles. Ces mesures donnent donc un aperçu général mais n'informent pas précisément sur l'évolution des sujets dans le temps.

Les probabilités de survie avec leurs intervalles de confiance peuvent être calculées pour des moments précis au suivi. Tel que décrit plus tôt, la probabilité de survie à un temps t se définit comme la chance de survie passé le temps t , i.e. $\hat{S}(t) = P(T > t)$. Dans la situation où aucun des sujets n'est censuré, la probabilité de survie est simplement la proportion de sujets survivant plus longtemps que le temps t : par exemple, si cinq sujets sont décédés sur vingt sujets à un an, la probabilité de survie est 75% à un an [15/20].

S'il y a censure, la probabilité de survie est obtenue à l'aide de l'estimateur (inconditionnel) de Kaplan-Meier¹¹⁷ aussi appelé « équation de produit limité » (*product limit formula*) car elle consiste en une série de produits de probabilité de survie en partant du début de la courbe au temps zéro jusqu'au point d'intérêt :

$$\hat{S}(t) = \prod_{t_{(i)} \leq t} \frac{n_i - d_i}{n_i}, \text{ où :}$$

- $\hat{S}(t)$ est l'estimé de la fonction de survie ;
- $t_{(i)}$ correspond à l'un des « m » évènements ordonnés dans le temps $(t_{(1)}, t_{(2)}, \dots, t_{(m)})$;
- n_i représente le nombre de sujets à risque d'un évènement au temps $t_{(i)}$;
- d_i indique le nombre d'évènements observés au temps $t_{(i)}$.

Par exemple, si on suit une cohorte de 20 sujets et que trois décèdent à six mois, un sujet décède à neuf mois, un sujet est censuré à dix mois, un autre décède à un an, deux sont censurés à treize mois et trois décèdent à quatorze mois, la probabilité de survivre :

- ... plus de zéro mois est de 100% ;
- ... plus de six mois est de $\left[\left(\frac{17}{20}\right) = 85\%\right]$;
- ... plus de neuf mois s'exprime comme la probabilité conditionnelle de survivre plus de neuf mois pour ceux qui ont survécu plus de six mois $\left[\left(\frac{16}{17}\right) = 94,1\%\right]$ multiplié par la probabilité de survivre plus de six mois pour tous les sujets $\left[\left(\frac{17}{20}\right) = 85\%\right]$, soit $\left[\left(\frac{17}{20}\right) \times \left(\frac{16}{17}\right) = \left(\frac{16}{20}\right) = 80\%\right]$;
- ... plus de dix mois est également de 80% puisque rien de bien ou de mal n'est survenu à dix mois et que l'on estime que le sujet censuré est représentatif des autres sujets ;
- ... plus d'un an est $\left[\left(\frac{17}{20}\right) \times \left(\frac{16}{17}\right) \times \left(\frac{14}{15}\right) = 74,7\%\right]$. On note que vu la censure, le dénominateur du dernier produit est 15 et non 16 puisque le nombre de sujet à risque de l'évènement à un an est de 15 et non de 16. Ceci nous empêche aussi de simplifier l'expression comme nous l'avons fait pour la survie à neuf mois.
- ... plus de treize mois de 74,7%.
- ... plus de quatorze mois de $\left(\frac{17}{20}\right) \times \left(\frac{16}{17}\right) \times \left(\frac{14}{15}\right) \times \left(\frac{9}{12}\right) = 56\%$.

On obtient un estimé de la variance de la courbe de survie estimée, i.e. de la précision des estimés de survie, en utilisant la formule classique de Greenwood¹¹⁸ :

$$\hat{V}[\hat{S}(t)] = [\hat{S}(t)]^2 \sum_{t_{(i)} \leq t} \frac{d_i}{n_i(n_i - d_i)}, \text{ où :}$$

- $\hat{V}[\hat{S}(t)]$ est l'estimé de la variance de $\hat{S}(t)$;
- $\sqrt{\hat{V}[\hat{S}(t)]}$ est l'estimé de l'erreur-type (*standard error*) de $\hat{S}(t)$;
- $\hat{S}(t) \pm 1.96\sqrt{\hat{V}[\hat{S}(t)]}$ donne un intervalle de confiance à 95% autour de $\hat{S}(t)$.

Comparaison de courbes de survie

La méthode de Kaplan-Meier permet de représenter graphiquement les courbes de survie de plus d'un groupe sur un même graphique avec leurs intervalles de confiance respectifs. Ceci permet de comparer « à l'œil » s'il y a des différences entre les groupes et où se situent ces différences dans le temps. On utilise classiquement des intervalles de confiances à 95%, mais certains auteurs ont proposé des intervalles de confiance plus larges (68,3%) qui correspondent en fait à l'équivalent d'un écart-type¹¹⁹⁻¹²².

Dans la situation où tous les sujets présentent un évènement et que tous les temps de survie sont connus, il est possible de comparer simplement les temps de survie de deux groupes de sujets avec un test non-paramétrique, tel que le test de Wilcoxon, puisque la distribution de ces temps de survie n'a souvent pas une distribution Normale.

Dans la plupart des analyses par contre, il y a un certain degré de censure avec des sujets qui se retirent de l'étude, qui sont perdus au suivi, ou qui ne présenteront pas l'évènement avant la fin de l'étude. On peut ainsi comparer la survie de deux groupes à un moment prédéfini dans le temps à l'aide d'une simple table 2x2 et d'un test de Chi-carré (χ^2) ou test exact de Fisher. Par exemple, on utilise ce type d'analyse pour la survie à 30 jours suivant une procédure chirurgicale (mortalité opératoire) ou bien pour la survie à un an dans un essai randomisé où chaque sujet est suivi de près pour au moins un an et où le degré de censure aléatoire est négligeable. Les inconvénients de cette approche sont que :

- Il ne doit théoriquement pas y avoir de censure avant le moment prédéfini pour l'analyse ;
- Le moment prédéfini est arbitraire ;
- Si le suivi est suffisamment long, l'analyse n'est pas informative (puisque 100% des sujets décèderont éventuellement) ;
- Il y a une perte d'information (puissance statistique) puisque l'on ne tient pas compte des temps jusqu'à chacun des évènements ;
- L'analyse s'intéresse seulement au fait qu'il y a eu ou non un évènement avant un moment prédéfini mais ne s'intéresse pas à la distribution de ces évènements dans le temps. Par exemple, une technique chirurgicale peut être associée à une plus grande mortalité précocement et une meilleure survie au suivi alors qu'un traitement médical est associée à une meilleure survie précocement et une moindre survie au suivi – dépendamment du moment prédéfini pour l'analyse, les conclusions de l'analyse pourraient être fort différentes.

Alternativement, les groupes peuvent être comparés à l'aide du (Mantel-Haenszel) log-rank test. Sommairement, il s'agit d'un test de Chi-carré (χ^2) stratifié où l'on forme en quelque sorte des tables 2x2 pour tous les moments (stratum) où un évènement est noté. On compare les décomptes d'évènements observés vs attendus pour chacun des groupes à chacun des temps où il y a eu des évènements. Ce test peut aussi comparer plus de deux groupes ensemble. L'hypothèse nulle est qu'il n'y a pas de différence entre les courbes de survie. Les valeurs de P (*p-values*) sont obtenues à partir des tables de la distribution de χ^2 avec un degré de liberté de moins qu'il n'y a de groupes comparés.

Le test de log-rank s'intéresse à l'ordre dans lequel les évènements surviennent - et pas au moment précis où les évènements surviennent - et au taux d'évènement dans chaque groupe. Le test devient positif si l'ordre des évènements est généralement plus précoce pour un groupe par rapport à l'autre, ou si le taux d'évènement est plus élevé pour un groupe par rapport à l'autre. Ce test peut être modifié pour donner plus de poids aux évènements survenant tôt dans le temps par un test de log-rank pondéré (*weighted*). Ces tests log-rank pondérés (Wilcoxon-Breslow, Tarone-Ware, Peto, Flemington-Harrington) sont peu utilisés en pratique. Si on décide de les utiliser par contre, le choix doit se faire *a priori* selon une rationnelle clinique précise, i.e. si on s'attend à ce que les courbes de survie divergent de telle ou telle façon selon que l'on s'attend à ce que l'effet de traitement soit noté surtout précocement ou tardivement. Le choix du test ne doit pas être basé sur la plus petite valeur de P obtenue avec tous ces tests log-rank pondérés qui sont souvent calculés automatiquement par les logiciels statistiques.

Modèles d'analyse de survie multivariés semi-paramétrique (Cox) et paramétriques

Pour l'analyse des facteurs de risque dans des modèles multivariés, on peut utiliser des modèles de régression à risque proportionnel de Cox (*Cox proportional hazards regression*) ou bien des modèles de régression paramétrique avec censure (exponentiel, Gamma, Weibull, Log-Normal). Pour l'analyse de survie, ces modèles sont généralement préférables au modèle de régression logistique multivarié puisque ce dernier ne tient pas compte des temps de survie et de la censure, le résultat étant binaire.

Les modèles paramétriques sont relativement simples et ont beaucoup de puissance, même en présence de données limitées. Certains de ces modèles plus simples étaient populaires avant que la technologie informatique et les logiciels de statistique deviennent davantage disponibles aux chercheurs. Par contre, il peut être difficile de vérifier si les suppositions de distribution de survie sont adéquates surtout si on ne dispose pas de beaucoup de données de survie. Si les données de survie sont limitées, on a le choix de se fier aux données recueillies ou bien à d'autres données historiques. J'ai exploré ces modèles pour mon premier travail⁶⁸ (décomposition dans le temps du niveau de risque et survie appariée à la population générale selon les table de survie gouvernementales qui suivent un modèle exponentiel – voir ci-dessous), mon troisième travail⁷⁰ (modèle Log-Normale) ainsi que d'autres travaux^{114,123,124} (modèle Weibull pour la censure par intervalle et survie appariée).

Le modèle de régression à risque proportionnel de Cox (*Cox proportional hazards regression*), introduit par David Cox en 1972^{58,59,63,65,125,126}, permet d'établir une relation paramétrique entre les facteurs de risque et la distribution des temps de survie, sans donner une forme particulière à celle-ci. Il s'agit donc d'un test semi-paramétrique. Plus spécifiquement, il est basé sur la fonction de risque instantané (*hazard function*). Contrairement à la régression paramétrique, la régression de Cox peut accommoder n'importe quelle fonction de risque instantané puisque aucune supposition sur la forme de la fonction de risque instantané n'est faite. L'ajout de variables est permis dans le modèle. L'effet de ces variables est considéré indépendant dans le temps, c'est-à-dire que l'effet multiplicatif de chaque variable est le même peu importe où on se situe dans le temps (notion de risque proportionnel). Par exemple, si le risque instantané de décès pour un sujet diabétique est deux fois plus élevé que le risque instantané de décès d'un sujet non diabétique à 30 jours, il en est de même à 1 ou 5 ans ; cet effet multiplicatif est constant dans le temps.

L'équation pour le modèle de Cox est $\lambda_1(t) = \lambda_0(t)e^{\beta'x} = \lambda_0(t)e^{\sum_{i=1}^p \beta_i X_i}$ ou de façon équivalente, $\ln(\lambda_1(t)) = \ln(\lambda_0(t)) + \sum_{i=1}^p \beta_i X_i = \ln(\lambda_0(t)) + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$, où :

- $\lambda_0(t)$, ou fonction de risque de base, est le risque instantané de base au temps t pour un patient référence avec toutes les variables explicatives égales à 0. Par exemple, si diabète est codé [1 : présent] et [0 : absent], $\lambda_0(t)$ représente le risque pour les non-diabétiques. Si l'âge est codé en variable continue avec « 0 » correspondant à la naissance, $\lambda_0(t)$ représente en théorie le risque

pour quelqu'un qui vient de naître (soit souvent une extrapolation des données du modèle si on s'intéresse aux sujets adultes) ; alors que si on convertit l'âge en variable centrée-réduiteⁱ, $\lambda_0(t)$ représente maintenant le risque pour quelqu'un ayant l'âge moyen du groupe ;

- $\lambda_1(t)$ est le risque instantané au temps t pour un patient avec un profil de variables X_1, X_2, \dots, X_p ;
- $\beta_1, \beta_2, \dots, \beta_p$ sont les coefficients des variables explicatives X_1, X_2, \dots, X_p .
- $\beta'x$ est la notation matricielle de $\sum_{i=1}^p \beta_i X_i$ (voir ci-haut).

L'équation de Cox est donc le produit de deux quantités : la fonction de risque de base qui est fonction du temps (dont la forme n'est pas spécifiée à l'avance, d'où le « *semi* »-paramétrique) et une expression exponentielle incluant coefficients et variables explicatives qui ne varient pas dans le temps. L'aspect « paramétrique » du modèle tient du fait que l'on doit spécifier les coefficients $\beta_1, \beta_2, \dots, \beta_p$. Une propriété importante du modèle de Cox est que les coefficients sont situés dans une expression exponentielle, ce qui permet de s'assurer que le modèle résolu donne toujours des risques instantanés non négatifs, entre zéro et l'infini ($[0 \leq \lambda_1(t) < \infty]$). Par définition, les risques instantanés doivent toujours être non-négatifs.

ⁱ En anglais “*standard score*” ou “*z-score*”. Consiste à soustraire la moyenne du groupe à chacune des valeurs initiales (centré) et à diviser toutes les valeurs par l'écart-type du groupe (réduite). Les variables centrées-réduites ont une moyenne nulle et un écart-type de 1.

Des estimés sont obtenus pour chacun des coefficients β du modèle par la méthode de vraisemblance partielle (*partial likelihood*), une solution mathématique proposée par David Cox pour ce modèle particulier qui ne peut pas être résolu par la technique d'estimation de vraisemblance maximale (*maximum likelihood estimation*, *MLE*) à l'instar des modèles complètement spécifiés (paramétriques) où la forme de la fonction de risque instantané est défini à l'avance.

Le concept de vraisemblance maximale a été introduit plus tôt avec la régression linéaire (voir Figure 8) et consiste en une méthode pour estimer les paramètres d'un modèle statistique. Cette méthode cherche à obtenir la valeur des paramètres du modèle qui représentent le mieux les données recueillies (concept de distribution des paramètres du modèle ayant la plus grande probabilité d'être vraie). Dans l'exemple en Figure 8, une droite était tracée dans un nuage de points de façon à minimiser la somme des distances au carré entre chacun des points et la droite, soit la méthode des moindres carrés, ou *least square*, une méthode d'estimation de vraisemblance maximale. De façon similaire, il est possible d'obtenir une solution pour les coefficients d'une courbe de survie paramétrique en utilisant d'autres formules de vraisemblance maximale. Pour le modèle de Cox, la forme de vraisemblance utilisé est dite partielle car elle considère les probabilités chez les sujets présentant l'évènement ainsi que les temps de survie avant censure, mais ne tient pas compte des sujets après leur censure. Supposons que l'on dispose d'une base de données avec n observations et k temps de survie distincts. L'estimation de Cox procède d'abord en triant les temps de survie de façon à ce que : $t_1 < t_2 < \dots < t_k$, où t_j représente le temps de survie pour le $j^{\text{ème}}$ sujet. Pour les sujets censurés,

définissons δ_j à 0 si le sujet censuré à droite (n'a pas présenté l'évènement à la fin du suivi), et 1 s'il a présenté l'évènement étudié (décès, complication, etc.). Enfin, définissons $R(t_j)$ comme le nombre de sujets à risque de présenter un évènement au temps t_j . La fonction de vraisemblance partielle est ainsi donnée par cette équation :

$$L_p = \prod_{j=1}^k \left[\frac{e^{\beta'x_j}}{\sum_{l \in R(t_j)} e^{\beta'x_l}} \right]^{\delta_j}, \text{ où } \prod \text{ est un « opérateur de produit » et désigne une}$$

séquence d'expressions multipliées entre elles (à l'instar de \sum pour l'addition).

Pour illustrer la fonction de vraisemblance partielle, reprenons l'exemple de la Figure 12. Les sujets sont triés par temps de survie : $t_E = t_F < t_C < t_A < t_D < t_B$, ou $5 = 5 < 10 < 15 < 20 < 25$. Dans cet exemple, seuls les sujets A et F sont décédés. Imaginons une série de facteurs de risque de décès et leur coefficients désignés par $\beta'x_A$ pour le sujet A , $\beta'x_B$ pour le sujet B , etc. (les valeurs des coefficients sont les mêmes pour tous les sujets, seuls les valeurs des variables changent selon les sujets, i.e diabétique ou non, etc.). La fonction de risque partielle est ainsi :

$$L_p = \left[\frac{e^{\beta'x_E}}{e^{\beta'x_A} + e^{\beta'x_B} + e^{\beta'x_C} + e^{\beta'x_D} + e^{\beta'x_E} + e^{\beta'x_F}} \right]^0 \times \left[\frac{e^{\beta'x_F}}{e^{\beta'x_A} + e^{\beta'x_B} + e^{\beta'x_C} + e^{\beta'x_D} + e^{\beta'x_E} + e^{\beta'x_F}} \right]^1 \times \left[\frac{e^{\beta'x_C}}{e^{\beta'x_A} + e^{\beta'x_B} + e^{\beta'x_C} + e^{\beta'x_D}} \right]^0 \times \left[\frac{e^{\beta'x_A}}{e^{\beta'x_A} + e^{\beta'x_B} + e^{\beta'x_D}} \right]^1 \times \left[\frac{e^{\beta'x_D}}{e^{\beta'x_B} + e^{\beta'x_D}} \right]^0 \times \left[\frac{e^{\beta'x_B}}{e^{\beta'x_B}} \right]^0$$

$$L_p = \left[\frac{e^{\beta'x_F}}{e^{\beta'x_A} + e^{\beta'x_B} + e^{\beta'x_C} + e^{\beta'x_D} + e^{\beta'x_E} + e^{\beta'x_F}} \right]^1 \times \left[\frac{e^{\beta'x_A}}{e^{\beta'x_A} + e^{\beta'x_B} + e^{\beta'x_D}} \right]^1$$

Des estimés sont obtenus pour les coefficients β' en maximisant cette dernière expression (c'est-à-dire en tentant d'obtenir la plus grande valeur possible de L_p). Le processus est effectué de façon itérative, i.e. répétée, par des algorithmes mathématiques en tentant de « deviner » la solution (dérivés partiels de la fonction logarithmique de L_p) et en modifiant la valeur des estimés de β' jusqu'à ce que la valeur maximale de L_p soit obtenue.

Plusieurs observations peuvent être notées dans cet exemple : 1) au moment où le sujet A est décédé, les sujets à risque étaient les sujets A , B et D (i.e. faisaient partie du groupe de sujets encore suivis au moment du décès du sujet A) ; 2) seuls les sujets ayant présenté l'évènement (sujets A et F) sont considérés au numérateur ; 3) les sujets qui n'ont pas présenté l'évènement (sujets « censurés »), mais qui ont fait partie de l'un des groupe « à risque », sont considérés dans le dénominateur. Il n'est pas nécessaire de spécifier la forme de la fonction de risque de base ($\lambda_0(t)$). L'intervalle entre les évènements (qu'il soit de 1 an ou de 10 ans) n'est pas informatif pour la fonction de vraisemblance partielle.

Le modèle de Cox est populaire car il est robuste : il permet d'obtenir des estimés fiables des coefficients régression et des rapports de risque et des courbes de survie ajustées pour un large éventail de données cliniques avec un minimum de suppositions. Même en présence d'un modèle paramétrique, le modèle de Cox permet une approximation relativement juste de ces modèles. Lorsque les données suivent de près un modèle paramétrique, il peut être avantageux d'utiliser ce modèle paramétrique ; dans le doute, il est souvent plus sûr de s'en tenir au modèle de Cox.

Rapport de risque (hazard ratio)

Le rapport de risque (*hazard ratio*) est défini comme le risque d'un sujet divisé par le risque d'un autre sujet ayant des facteurs de risque différents (X).

$$\hat{HR} = \frac{\lambda_j(t)}{\lambda_l(t)} = \frac{\lambda_0(t)e^{\beta x_j}}{\lambda_0(t)e^{\beta x_l}} = \frac{e^{\beta x_j}}{e^{\beta x_l}} = e^{\beta(x_j - x_l)}, \text{ où :}$$

- \hat{HR} est l'estimé du rapport de risque (*hazard ratio*) ;
- $\lambda_j(t)$ et $\lambda_l(t)$ sont les fonctions de risque instantané de sujets avec le profil de variables x_j et x_l , respectivement. On note que la fonction de risque instantané de base, $\lambda_0(t)$, s'élimine dans l'équation.

On note que si les sujets ne diffèrent que pour une variable binaire [$X_1 = (0,1)$], le rapport de risque se résume par $[e^{\beta(x_j - x_l)} = e^{\beta_1(X_{1j} - X_{1l})} = e^{\beta_1(1-0)} = e^{\beta_1}]$

Le rapport de risque s'interprète de la même façon qu'un rapport de cotes (*odds ratio*) pour la régression logistique. Un rapport de risque de 1 signifie qu'il n'y a pas de relation (i.e. le risque est le même pour les sujets exposés ou non au facteur de risque), un rapport de risque de 10 signifie que le risque des sujets exposés au facteur de risque est 10 fois plus élevé, alors qu'un rapport de risque de 1/10 signifie que le risque des sujets exposé au facteur de risque est dix fois moindre.

Évaluation du risque proportionnel

Trois méthodes générales sont disponibles pour évaluer si la supposition de risque proportionnel est réaliste.

Techniques graphiques. La transformation log-log des courbes de survie Kaplan-Meier de deux groupes nous permet d'évaluer graphiquement si ces dernières sont parallèles. Il s'agit d'effectuer la transformation logarithmique (logarithme naturel en base e) deux fois des estimés de survie des courbes de survie. Mathématiquement, on décrit la courbe log-log comme la « valeur négative de la transformation logarithmique de la valeur négative de la transformation logarithmique de l'estimé de survie » ou $-\log(-\log(S_1(t)))$, puisque le logarithme d'une proportion donne une valeur négative et que l'on ne peut effectuer une seconde transformation logarithmique sur une valeur négative. Pourquoi une transformation log-log est elle utile pour évaluer l'hypothèse de risque proportionnel? Prenons l'équation de survie de Cox et effectuons la transformation log-log :

$$\begin{aligned} -\log(-\log(S_1(t))) &= -\log(-\log(S_0(t)e^{\beta'x})) = -\log(-e^{\beta'x} \times \log(S_0(t))) \\ &= -\log(-e^{\beta'x}) - \log(-\log(S_0(t))) = -\beta'x - \log(-\log(S_0(t))) \end{aligned}$$

On remarque donc que l'expression finale obtenue est composée de la somme de deux termes, le premier étant la somme linéaire des coefficients et leur variables pour tous les sujets ($-\beta'x$ ou $-\sum_{i=1}^p \beta_i X_i$), qui est indépendante du temps, et le deuxième la transformation $-\log(-\log)$ de la fonction de survie de base $S_0(t)$. Supposons que deux groupes de sujets ne diffèrent que pour le diabète, la différence entre les deux courbes $-\log(-\log(S_1(t)))$ définies par cette variable « diabète » devrait

être la même peu importe le temps t et correspondre au coefficient de la variable diabète puisque tous les autres termes de l'équation s'annulent. Si les courbes $-\log(-\log(S_1(t)))$ sont grossièrement non-parallèles, l'hypothèse de risque proportionnel pour la variable diabète n'est pas satisfaite graphiquement.

Pour des variables continues, on divise le groupe en quelques catégories (2 ou 3) selon des limites cliniquement pertinentes (ex. : catégorisation de la variable « index de masse corporelle » selon les limites définies pour la maigreur et l'obésité) et on compare les courbes $-\log(-\log(S_1(t)))$ de ces groupes comme pour la variable binaire diabète. On peut aussi comparer les transformations log-log de courbes de survie stratifiées pour une variable dont on souhaite évaluer l'hypothèse de risque proportionnel (ex. : diabète), lesquelles ont été préalablement ajustées selon le modèle de Cox pour des variables répondant déjà au risque proportionnel (ex. : hypertension, dyslipidémie, ...). Ce modèle permet d'estimer des courbes de survie avec des fonctions de survie de base différentes mais pour lesquelles les valeurs des autres coefficients pour les autres variables sont identiques (voir le modèle de Cox stratifié).

Alternativement, on compare les courbes de survies « observées » selon la méthode de Kaplan-Meier pour les sujets avec ou sans le facteur de risque (ex. : courbes de survie de Kaplan Meier pour les diabétique, puis pour les non-diabétique) aux courbes de survie « attendues » selon le modèle de Cox, où le facteur de risque est inclus dans le modèle. Il s'agit en fait d'un équivalent graphique du test de validité de l'ajustement (voir ci-dessous).

Toutes ces méthodes graphique demeurent subjectives. Puisque le modèle de Cox est très robuste (i.e., estimés non-biaisés dans la plupart des situations), on assume que le risque est proportionnel à moins de forte évidence de non parallélisme pour les courbes log-log ou de courbes de survie très différentes pour le test des courbes de survies « observées » versus « attendues ».

Test de validité de l'ajustement (*goodness of fit test*)¹²⁷. Ce test utilise les résidus de Schoenfeld (*Schoenfeld residuals*)¹²⁸ définis pour chacune des variables à chaque moment que l'un des sujets présente un évènement, ainsi que le rang des temps de survie des sujets présentant un évènement. Les résidus de Schoenfeld correspondent sommairement à la différence entre la valeur observée pour une variable pour un sujet présentant un évènement au temps t et la valeur moyenne de cette même variable pour les autres sujets, pondérée selon le risque de chacun. On effectue alors pour chacune des variables du modèle un test de corrélation entre ces résidus de Schoenfeld et les rangs des temps de survie. Si l'effet de la variable ne varie pas dans le temps, il ne devrait pas y avoir de corrélation entre les résidus de Schoenfeld et les rangs des temps de survie. Avec ce test, on ne peut jamais prouver l'hypothèse nulle (i.e. que la supposition de risque proportionnel est respectée à coup sûr) ; tout ce que l'on peut dire est que l'on n'a pas suffisamment de données pour rejeter l'hypothèse nulle. Les valeurs de P peuvent être influencées par la taille d'échantillon : une importante déviation de la proportionnalité peut ne pas être apparente pour un petit groupe alors qu'une légère déviation de la proportionnalité peut apparaître comme hautement significative pour un large groupe. Le test est donc plus objectif, mais il est recommandé de le considérer conjointement avec des méthodes graphiques avant de décider du sort d'une variable.

Modèle de Cox incluant des variables variant dans le temps.

La forme de la fonction de risque de ce modèle est la suivante :

$$\lambda_1(t) = \lambda_0(t) e^{\beta'x + \delta X_t \times g(t)}, \text{ où}$$

- δ est le coefficient de la variable variant dans le temps ;
- X_t est une variable variant dans le temps selon la fonction de temps $g(t)$;
- $g(t)$ est une fonction du temps qui peut correspondre au temps t comme tel, à une transformation de celui-ci ($\log(t)$) ou bien à un point ou des points précis dans le temps (≤ 30 jours vs > 30 jours après la chirurgie).

On teste alors si le coefficient δ de cette variable variant dans le temps X_t selon $g(t)$ est statistiquement différent de zéro. Si tel est le cas, la supposition de risque proportionnel n'est pas respectée. Si l'hypothèse nulle ne peut être rejetée (i.e. l'ajout de la variable variant dans le temps n'apporte rien au modèle), le modèle de risque instantané se réduit au modèle de Cox pour risque proportionnel $\lambda_1(t) = \lambda_0(t) e^{\beta'x}$. L'inconvénient de cette approche est qu'il dépend de la façon dont on définit $g(t)$: un autre choix de $g(t)$ peut mener à une conclusion différente. Par exemple, si on a choisi une méthode d'expression du temps avec un point précis dans le temps (30 jours après la chirurgie) et que les sujets ne décèdent généralement pas dans les 30 jours après la chirurgie, on n'est pas en mesure d'évaluer adéquatement une variable dont l'effet commence à se voir cinq ans après la chirurgie, mais pas moins de cinq ans après la chirurgie.

Modèle de Cox stratifié

Lorsque des variables ne respectent pas la supposition de risque proportionnel, on peut utiliser un modèle de Cox stratifié pour cette variable. Dans ce modèle, on permet différentes fonctions de risque de base pour les variables stratifiées mais on restreint les coefficients des autres facteurs de risque à être identiques pour chacune des équations définies par la variable de stratification ; ces coefficients sont « ajustés » pour les variables de stratification, tout comme les variables de stratification sont « ajustées » selon les autres variables dans le modèle.

Dans ce modèle stratifié, les marches d'escalier des courbes de survie ajustées sont différentes pour les courbes de survie définies par les variables de stratification (vu que leur fonction de risque de base est différente). Toutefois, on n'est pas en mesure d'obtenir un coefficient, ou un rapport de risque, pour les variables de stratification. J'ai utilisé ce modèle de Cox stratifié pour obtenir les courbes de survies ajustées selon les hôpitaux ou bien selon le facteur d'appariement (définissant chacun des sous-groupes 1:3 incluant 4 sujets avec un score de propension similaire) pour mon second travail⁶⁹. Dans cette situation, je n'étais pas vraiment intéressé à obtenir des rapports de risque pour les variables « hôpitaux » et « facteur d'appariement » mais seulement à ce que les rapports de risque pour la variable « histoire d'ICP » soient ajustés pour ces variables de stratification « hôpitaux » ou « facteur d'appariement ».

Courbes de survie ajustées en utilisant le modèle de Cox

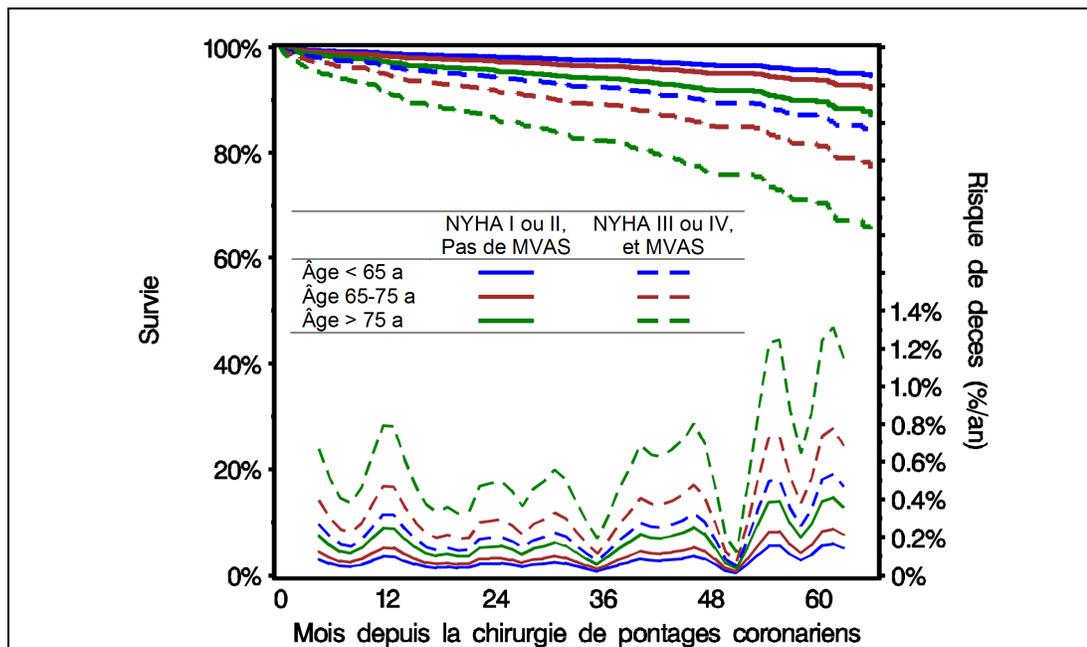
La fonction de risque instantané de Cox peut être convertie en une fonction de survie ajustée de Cox.

$$\lambda_1(t) = \lambda_0(t)e^{\beta'x} \Leftrightarrow S_1(t) = S_0(t)e^{-\beta'x}, \text{ où :}$$

- $S_0(t)$ ou fonction de survie de base, est la survie de base au temps t pour un patient référence avec toutes les variables explicatives égales à 0 ;
- $S_1(t)$ est le pourcentage de survie au temps t pour un patient avec un profil de variables X_1, X_2, \dots, X_p .

On obtient une courbe de survie ajustée en spécifiant la valeur des facteurs de risque du modèle. Pour les variables binaires, j'utilise habituellement le mode (soit la valeur la plus fréquente, i.e. si 35% des sujets sont diabétique, je représente la courbe ajustée pour les non-diabétique) et pour les variables continues, j'utilise la moyenne ou la médiane. À l'instar de la courbe de survie de Kaplan-Meier, ces courbes de survie ajustées sont représentées par des courbes en escalier. On peut aussi démontrer graphiquement l'impact d'un facteur de risque (Figure 14). Ces courbes tirées du modèle donnent l'impression que les courbes sont approximativement parallèles et les marches d'escalier apparaissent aux mêmes moments pour les deux groupes. Rappelons que dans le modèle de Cox classique (sans variable variant dans le temps), on considère que l'effet des facteurs de risque ne varie pas dans le temps (risque proportionnel) et que leurs effets sont indépendants des fonctions de survie ($S_0(t)$) et de risque de base ($\lambda_0(t)$).

Figure 14. Exemple courbes de survie et de risque de Cox ajustées.



Modèle ajusté pour le sexe masculin, la présence d'hypertension, dyslipidémie, FEVG > 39%, assurance santé privée, utilisation de l'artère mammaire, et l'absence d'ICP, diabète, arythmie, maladie cérébrovasculaire, insuffisance rénale, maladie pulmonaire chronique, choc cardiogénique, lésion >50% du tronc commun et d'utilisation d'artère radiale. MVAS : maladie vasculaire périphérique sévère. Les courbes épaisses du haut sont les courbes de survie et les courbes minces du bas représentent le risque instantané de décès.

Exemple de courbes de survies ajustées selon le modèle de risque proportionnel de Cox à partir d'un échantillon aléatoire de 800 patients tiré du deuxième travail⁶⁹. La taille de l'échantillon a été réduite pour permettre de bien visualiser la présence des « marches d'escalier » de la courbe de survie ajustée. On remarque que chaque courbe de survie présente des marches d'escalier au même moment et que celles-ci sont associées à des pics d'augmentation du risque instantané. La courbe de survie épaisse hachurée en vert est ajustée pour un patient de plus de 75 ans (rapport de risque (HR) de 1,69 par rapport au groupe référence des 65-75 ans) qui présente une classe NYHA III ou IV (HR = 1,52) avec MVAS (HR = 2,11). Les rapports de risque entre chacun des groupes sont proportionnels dans le temps (les courbes de risque instantané sont parallèles entre-elles dans le temps sur l'échelle logarithmique). Elle est comparée graphiquement aux autres sous-groupes de patients définis par ces facteurs de risque, où on a tenu compte des autres variables.

Pour vérifier que les deux groupes comparés ont bel et bien des fonctions de risque de base similaires (i.e. que le risque est proportionnel dans le temps), je compare les fonctions de risque de base des deux groupes et je m'assure que ces courbes ne se croisent pas, i.e. que le risque de base d'un groupe est un multiple du risque de base de l'autre groupe pour n'importe quel temps t . Cette situation où les fonctions de risque se croisent peut se rencontrer en clinique lorsque l'on compare la survie d'un groupe médical à un groupe chirurgical. Le risque de décès est plus grand immédiatement après une procédure chirurgicale, mais celle-ci peut avoir un plus grand bénéfice (meilleure survie) à plus long terme comparé au traitement médical. Notons que pour comparer les fonctions de risque de deux groupes (ex. : diabétique vs non-diabétique), les estimés de la fonction de risque doivent être obtenus en analysant ces groupes séparément (i.e. en stratifiant en fonction de la variable « groupe »). Les estimés tirés d'un modèle de Cox non stratifié où la variable « groupe » fait partie des variables du modèle ne se croiseront pas car ils sont tirés du modèle de Cox où nous avons déjà fait la supposition qu'elles ne se croiseront pas (Figure 14).

Pour la Figure 14, la fonction de risque de base a été obtenue à partir des estimés de survie du modèle de Cox pour les six sous-groupes. Étant donné que la fonction de risque instantané est exprimée pour un intervalle de temps infiniment petit, les estimés instantanés de risque sont très grands à chaque marche d'escalier pour revenir à zéro entre les pics. Dans le but d'obtenir une valeur moyenne, j'utilise un algorithme calculant la valeur moyenne de la fonction de risque par intervalle de temps pour une fenêtre définie de temps (*moving average*). Dans cet exemple, j'ai choisi arbitrairement une fenêtre de trois mois, raison pour laquelle la courbe de la fonction de risque de base « sur trois mois » estimée débute à trois mois de suivi.

DECOMPOSITION DANS LE TEMPS DU NIVEAU DE RISQUE (*HAZARD*) EN PHASES

Suivant une intervention chirurgicale, le risque de décès est relativement élevé immédiatement, diminue rapidement à un bas niveau puis augmente progressivement dans le temps à mesure que le patient vieillit. Plusieurs autres évènements postopératoires suivent la même structure, comme par exemple la survenue d'infarctus du myocarde après une intervention de pontages coronariens. À l'inverse, d'autres évènements présentent une structure de risque au suivi différente. Par exemple, la survenue de dégénérescence structurelle d'une prothèse valvulaire biologique suivant son implantation où le risque est minimal immédiatement après la chirurgie et pour les 5-10 années suivantes, puis le risque augmente progressivement par la suite.

Dans leurs articles publiés^{122,129-131}, ouvrages de référence¹³² et documents d'information distribués sur leur site web¹¹⁹⁻¹²¹, Dr Blackstone et ses collègues proposent une méthode d'analyse de survie complètement paramétrique. La méthode décrite permet la décomposition dans le temps du niveau de risque (*hazard*) suivant une intervention en une à trois phases complémentaires et superposables. On décrit un maximum de trois phases de risque : précoce, constante et tardive. Chacune des phases est construite à partir de différentes fonctions génériques du temps exprimées à l'aide d'équations provenant d'une même famille d'équations (*nested*). Chacune des phases est définie pour l'ensemble du suivi ($0 \leq t < \text{infini}$), bien que l'effet de chacune de ces phases est plus marqué pour une partie du suivi en particulier. La

phase précoce tend à présenter le maximum de son effet dans les jours ou mois suivant l'intervention (dépendamment du type d'intervention) alors que la phase continue a un impact tout au long du suivi et que la phase tardive a un impact maximal à la toute fin du suivi, si celui-ci est suffisamment long.

Initialement, l'estimation de la fonction de risque cumulative (*cumulative hazard function*) par la méthode non paramétrique de Kaplan-Meier est utile pour nous indiquer la structure de la distribution des événements, telle que la présence d'une phase précoce (partie abrupte au tout début), une phase constante (augmentation linéaire) et/ou une phase tardive (phase accélérée courbant vers le haut). Par contre, on détermine si une phase est requise ou non en effectuant un test de rapport de vraisemblance (*likelihood ratio test*), c'est à dire que l'on compare à quel point l'équation suit bien les données avec ou sans la phase en question. Il peut être démontré que la différence entre les deux vraisemblances suit approximativement une distribution de Chi-carré (χ^2) avec un nombre de degré de liberté équivalent à la différence entre le nombre total de paramètres entre les deux équations. De là, on formule un test d'hypothèse et on détermine *a priori* un niveau de significativité (habituellement $\alpha = 0,05$, i.e. on rejette l'hypothèse nulle si $P < 0,05$) :

- H_0 : « Est-ce que les deux vraisemblances sont similaires ? » (auquel cas la phase en question n'est pas requise – hypothèse nulle)
- H_A : « ...ou sont-elles statistiquement différentes ? » (hypothèse alternative)

Le nombre de phases qu'il est possible d'estimer dépend entre autres :

- de la structure sous-jacente de distribution des évènements ;
- de la durée du suivi ;
- de la courbe produite selon les paramètres choisis pour chacune des phases.

L'estimation des divers paramètres de ces équations est effectuée suivant la technique d'estimation du maximum de vraisemblance (*maximum likelihood estimation ; MLE*), i.e. qui permet de choisir la valeur des paramètres qui représentent le mieux les données. Dépendamment des équations, certains paramètres sont estimés directement alors que d'autres doivent être estimés itérativement (par essai-erreur) à l'aide d'algorithmes mathématiques. On retrouve un paramètre de mise à l'échelle (*scaling parameter*) par phase et les autres paramètres ont un impact sur la forme de la courbe (*shaping parameters*).

Une fois ces paramètres estimés, on obtient par algèbre une mesure de dispersion (variance/covariance, écart-type, intervalle de confiance). Suivant une approximation de la loi de Normalité, on formule des tests d'hypothèse pour chacun de ces paramètres :

- H_0 : « Est-ce que ce paramètre est équivalent à zéro ? » (auquel cas ce paramètre n'est pas requis – hypothèse nulle)
- H_A : « ...ou est-il statistiquement différent de zéro ? » (hypothèse alternative)

Habituellement, on choisit une p -value $< 0,05$ (et parfois $< 0,1$) pour déterminer ce qui est statistiquement différent de zéro. On suit le principe de parcimonie par lequel on tente de représenter les données avec le moins de phases (précoce, constante et/ou tardive) et de paramètres par phase (*scaling* et *shaping parameters*).

Une fois que les paramètres des équations sont estimés, on peut analyser l'impact de facteurs de risque sur chacune des phases prises séparément. Ceci permet de déterminer à quelle phase du suivi le facteur de risque en question a le plus d'impact. Des variables sont alors ajoutées à chacune des phases pour évaluer leur impact. Leurs coefficients et mesures de dispersion sont estimés de la même façon que pour les paramètres, i.e. par la méthode d'estimation du maximum de vraisemblance. Lors de l'estimation des coefficients de ces variables, les paramètres de forme de courbe (*shaping parameters*) sont fixés à la valeur estimée dans le modèle de base, mais on permet aux paramètres de mise à l'échelle (*scaling parameters*) de varier selon les variables présentes dans le modèle.

Des programmes permettant l'application de cette méthode sont disponibles à partir du site Web du *Cleveland Clinic Heart Center* à l'adresse suivante : <http://www.clevelandclinic.org/heartcenter/hazard/default.htm>. Sur ce site, on retrouve également trois autres documents qui décrivent la méthode en plus amples détails¹¹⁹⁻¹²¹. J'ai utilisé cette méthode statistique pour le premier manuscrit⁶⁸.

BOOTSTRAPPING

Le “*bootstrapping*” est une technique statistique pour obtenir un estimé robuste de la distribution d'échantillonnage (variance, écart-type [*standard deviation*]) d'un paramètre de la population tel qu'une moyenne, une médiane, une proportion, un rapport de cotes (*odds ratio*), un coefficient de corrélation ou de régression^{60,67,133-136}. Il permet aussi de construire des tests d'hypothèse. Le nom “*bootstrap*” vient de l'expression anglaise « *pull up by your own bootstraps* »¹³⁷ qui signifie « compter sur ses propres ressources », puisque le ré-échantillonnage est effectué à même la cohorte originale sans nécessiter de nouvelles ressources. L'échantillonnage est effectué avec remplacement, ce qui signifie qu'après avoir sélectionné au hasard une observation, on la remet dans le bassin avant de sélectionner l'observation suivante. Ainsi, une même observation ou sujet peut être sélectionné plus d'une fois. Dans le but d'obtenir une distribution de coefficient de régression, le modèle de régression logistique est recalculé sur cet échantillon de “*bootstrapping*” et le processus est répété à de multiples reprises (> 1000).

La technique de “*bootstrapping*” permet de s'assurer de la validité interne des résultats d'une étude pour que le modèle de régression logistique final ne suive pas de trop près la distribution des variables du modèle (*overfit*). La technique suppose que la cohorte à l'étude est représentative de la population à l'étude et que chacun des ré-échantillonnages sont indépendant les uns des autres et proviennent d'une même distribution de la population. Ainsi, si la cohorte originale compte peu de sujets, on ne peut pas faire confiance à la distribution de “*bootstrapping*” comme

reflet de la distribution de la population sur laquelle on désire ultimement effectuer des inférences statistiques.

L'avantage de la technique de "*bootstrapping*" est sa simplicité. Il est relativement facile d'appliquer la technique de "*bootstrapping*" pour obtenir des estimés de variance, écart-type et intervalles de confiance pour des estimés avec des distributions complexes, tels que les percentiles, les proportions, les rapports de cotes et les coefficients de corrélation.

Parmi ses limites, on note que le processus peut parfois être long et être exigeant pour l'ordinateur (ce qui est de moins en moins un facteur limitant de nos jours) et il ne donne pas de garantie d'un résultat final et unique. Lorsque l'on utilise un modèle de régression pré-spécifié, les résultats tendent à être trop optimistes (i.e. intervalles de confiance trop étroits) puisque les échantillons de "*bootstrapping*" sont puisés à même la cohorte qui a défini les coefficients du modèle de régression pré-spécifié. Cette limite peut être contournée en utilisant un processus de sélection de variable automatisé dans la boucle de "*bootstrapping*", tel que le mode de sélection pas à pas (*stepwise*).

Enfin, la technique de "*bootstrapping*" est perçue par certains chercheurs comme un trucage statistique, bien que ses bases théoriques et pratiques soient bien établies. Les principales objections à la technique du "*bootstrapping*" sont reliées à la taille des échantillons de "*bootstrapping*" (si la cohorte est petite, les échantillons de "*bootstrapping*" le sont aussi peu importe le nombre d'échantillons de "*bootstrapping*" qui sont créés, et les estimés peuvent être potentiellement biaisés),

aux estimés trop optimiste tirés de la distribution du “*bootstrapping*” dans certaines situations, et à l’hypothèse d’indépendance des échantillon de “*bootstrapping*” qui est difficile à vérifier.

Voici les étapes à suivre pour valider les résultats d’une régression logistique par la technique de “*bootstrapping*” :

1. On estime d’abord le modèle de régression logistique multivariée avec la population d’étude originale et on obtient un estimé du coefficient de la variable d’intérêt ($\hat{\beta}_1$).
2. S’il y a 200 patients dans la population à l’étude, on échantillonne au hasard 200 patients avec remplacement à partir de la population d’étude originale. Ainsi un même sujet peut être inclus plus d’une fois ou pas du tout dans un échantillon ;
3. On répète le modèle sur le « nouvel » échantillon et on obtient un nouvel estimé : $\hat{\beta}_{1,1}$;
4. On répète les étapes 2 et 3, disons 1000 fois, et on obtient un estimé de $\hat{\beta}_1$ à chaque fois : $\hat{\beta}_{1,1}, \hat{\beta}_{1,2}, \hat{\beta}_{1,3}, \dots, \hat{\beta}_{1,1000}$;
5. On calcule la variance « empirique » de “*bootstrapping*” de $\hat{\beta}_1$:

$$Var^*(\hat{\beta}_1) = \sum_{i=1}^N \frac{(\hat{\beta}_{1,i} - \bar{\hat{\beta}}_1)^2}{(N-1)} \text{ où } N = 1000 ;$$

6. On obtient finalement la nouvelle p-value de “*bootstrapping*” en utilisant le

$\hat{\beta}_1$ et la variance $Var^*(\hat{\beta}_1)$ nouvellement calculée selon la distribution de t :

$$t = \frac{(\hat{\beta}_1 - 0)}{\sqrt{Var^*(\hat{\beta}_1)}}$$

J’ai utilisé la technique de “*bootstrapping*” dans mon premier travail pour effectuer une « validation interne » des résultats d’un modèle de régression logistique multivariée⁶⁸ en utilisant d’une autre façon les données disponibles. En présence de données externe sur une cohorte similaire, j’aurais également pu effectuer une validation en comparant les résultats prédits par le modèle de régression logistique avec les résultats obtenus pour la cohorte externe.

Enfin, j’ai utilisé la technique de “*bootstrapping*” pour obtenir une mesure de dispersion pour des proportions dont le dénominateur et le numérateur étaient obtenus à partir de différents sondages et bases de données publiques^{138,139}. Alternativement, j’aurais pu soit utiliser une simplification de l’équation de mesure de dispersion pour une proportion en considérant certaines de ces valeurs comme connues avec certitude (fixées) ou j’aurais pu utiliser des équations complexes pour tenter d’estimer la variance d’une telle proportion (par exemple, la technique delta utilisant les dérivés de second degré des séries de Taylor – celle-ci ne sera pas abordée ici). La technique de “*bootstrapping*” m’apparaissait indiquée dans cette situation vu sa simplicité et sa capacité d’adaptation aux distributions qui ne suivent pas parfaitement la loi Normale.

CHAPITRE II

HYPOTHÈSES ET BUTS

OBJECTIFS GÉNÉRAUX

J'ai élaboré les travaux présentés dans le but d'explorer en profondeur certaines méthodes épidémiologiques et statistiques appliquées à la chirurgie cardiaque adulte, plus particulièrement les analyses de survie (paramétrique ou non), les applications du "*propensity score*", et les analyses de coûts. Ces travaux m'ont aussi permis de parfaire mes connaissances dans l'élaboration de devis d'étude, la gestion de larges bases de données, la planification et la réalisation des analyses biostatistiques appropriées incluant les analyses de régression logistique, linéaire, Log-Normale et autres incluant des applications du "*bootstrapping*". En tant que clinicien, mon intérêt était dans le choix judicieux d'analyses biostatistiques et leur application dans le cadre de la recherche évaluative en chirurgie cardiaque adulte. Je me suis également intéressé aux analyses longitudinales à mesures répétées dans le temps pour d'autres travaux^{114,123,124,140-143}, une extension des analyses de survie en un sens, mais celles-ci ne seront pas détaillées ici.

DISSECTION AIGÛE DE L'AORTE ASCENDANTE

Pour le premier travail, l'objectif était d'étendre nos connaissances sur la survie à long terme suivant la chirurgie pour une dissection aigüe de l'aorte ascendante en révisant l'expérience d'un centre hospitalier universitaire sur plus de 25 ans. Je me suis également intéressé à l'impact des caractéristiques des patients et de l'évolution de l'évaluation préopératoire, des techniques chirurgicales et du suivi médical sur les résultats obtenus. L'hypothèse de travail était que les résultats des patients suivant une chirurgie pour une dissection aigüe de l'aorte ascendante se sont perfectionnés graduellement suivant l'amélioration de la médecine et de la chirurgie.

Le projet a nécessité une révision détaillée de plusieurs dossiers médicaux avec suivi téléphonique. J'ai obtenu les données de survie ainsi que les causes de mortalité en faisant appel au registre national des décès aux États-Unis (*National Death Index*)¹⁴⁴. Ce projet de recherche a été l'occasion d'explorer des méthodes biostatistiques avancées telles que le "bootstrapping" pour la régression logistique, analyses de survie utilisant les méthodes de Kaplan-Meier, de Cox et de décomposition dans le temps du niveau de risque (*hazard*) en phases (*multiphase hazard regression*)¹²²) avec estimation de courbes de survie paramétriques et estimation de courbe de survie comparée (*matched*) à la population du Massachusetts pour l'âge, le sexe et la race d'après les tables de statistiques de survie des États-Unis (*US Life Tables*)¹⁴⁵). La décomposition dans le temps du niveau de risque a été peu rapportée dans la littérature en dehors des articles provenant de celui qui l'a proposé (Dr. E. H. Blackstone).

PONTAGE CORONARIENS APRÈS INTERVENTION CORONARIENNE PERCUTANÉE

Pour le second travail, j'ai émis l'hypothèse que les patients ayant une intervention coronarienne percutanée (ICP) avant leur chirurgie de pontage coronarien (PAC) ont une diminution de la survie postopératoire comparativement aux patients PAC sans histoire d'ICP. J'ai supposé que cette différence était plus marquée pour les patients ayant un intervalle court entre l'ICP et le PAC. L'objectif était de décrire et comparer les caractéristiques et les résultats de ces patients. Le résultat d'intérêt de l'étude était la mortalité à moyen terme et les résultats secondaires étaient la mortalité à 30 jours, la morbidité hospitalière et la durée de séjour hospitalier.

J'ai relié la base de données cliniques obligatoire en chirurgie cardiaque de l'État du Massachusetts à la base de données administratives hospitalières pour connaître quels patients, et à quel moment ceux-ci, ont bénéficié d'une ICP dans les cinq ans précédant les PAC. J'ai exclu les patients ayant bénéficié de leur ICP dans une situation d'infarctus aigu du myocarde ainsi que ceux venant d'un autre État puisque le suivi de survie est effectué à l'aide du registre de décès de l'État du Massachusetts. J'ai analysé séparément les patients se présentant pour leur PAC deux semaines ou moins après leur ICP de ceux s'étant présentés plus tard. J'ai apparié ces patients (1:3) à des patients n'ayant pas bénéficié d'une ICP en utilisant un PSc^{104,146}. J'ai analysé les résultats à court et moyen terme de ces patients utilisant des méthodes biostatistiques adaptées aux analyses avec appariement.

ANALYSE DE COÛTS EN CHIRURGIE CARDIAQUE

L'objectif de ce troisième travail était d'évaluer les changements dans les 1) volumes de référence de patients en chirurgie cardiaque ; 2) caractéristiques des patients ; et 3) profit réalisé par un centre hospitalier universitaire (CHU) avant et après l'implantation d'un nouveau centre de chirurgie cardiaque satellite (SAT). L'hypothèse de travail était que l'implantation du SAT a entraîné une diminution des volumes de référence en chirurgie cardiaque, que les patients référés étaient plus malades que ceux retenus pour être opérés au SAT et que les profits réalisés par le CHU s'en trouvaient diminués puisque le mode de compensation par forfait de Medicare, Medicaid, des organismes de prise en charge de la santé (*Health Maintenance Organization*) et des compagnies privées d'assurance santé aux États-Unis dédommagent défavorablement pour les patients plus lourds.

L'analyse des données financières est compliquée par leur distribution très asymétrique (*right skewed*) qui limite l'utilisation de modèles statistique plus communs (qui nécessitent une distribution suivant la loi Normale)¹⁴⁷. J'ai exploré plusieurs techniques statistiques, dont les analyses non paramétriques, les analyses de survie appliquées aux coûts (en utilisant les « coûts » au lieu du « temps de survie » comme variable résultat), les analyses paramétriques avec distribution Log-Normale et Gamma, la transformation logarithmique des coûts (*log-coûts*) et la technique de "*bootstrapping*"¹⁴⁸⁻¹⁵¹. J'ai retenu la distribution Log-Normale (paramétrique) pour les modèles statistiques puisqu'elle représente la meilleure approximation (*fit*) pour ces données, permet de réaliser les inférences dans l'unité de coût non transformé (\$) au lieu de *log-\$*) et permet la modélisation multivariée.

CHAPITRE III

CORPS DE LA THÈSE

ARTICLES ORIGINAUX

CHAPITRE III

EVOLUTION IN SURGICAL MANAGEMENT AND OUTCOME FOR ACUTE ASCENDING AORTIC DISSECTION

Evolution in Surgical Management and Outcome for Acute Ascending Aortic Dissection

Louis-Mathieu Stevens, MD^a, Joren C. Madsen, MD, DPhil^a,
Eric M. Isselbacher, MD^b, Paul Khairy, MD, PhD^c, Thomas E. MacGillivray, MD^a,
Alan D. Hilgenberg, MD^a, Arvind K. Agnihotri, MD^a

Word count : 3698

Keywords : Acute Type A Aortic Dissection ; Long-Term Follow-up ; Statistics,
Survival Analysis.

Division of Cardiac Surgery^a and Division of Cardiology^b, Massachusetts General
Hospital, Boston, Massachusetts, USA and Department of Medicine^c, Montreal Heart
Institute, Montreal, Quebec, Canada.

Address for reprints : A.K. Agnihotri, MD, Department of Cardiothoracic Surgery,
Massachusetts General Hospital, 55 Fruit Street, COX 630, Boston, Massachusetts,
USA, 02114. Tel. : 617-726-5608 ; Fax : 617-726-8469 (Email :

)

ABSTRACT

Objective : To assess early and late survival and cardiovascular specific mortality following surgical repair of acute ascending aortic dissection (AAAD), and the impact of differences in surgical technique, patient characteristics, and preoperative diagnostic testing.

Methods : Between 1979 and 2003, 195 consecutive patients underwent repair for AAAD within 2 weeks of the onset of symptoms. Mean follow-up was 7.0 ± 5.9 years (range : 0-26 years) and was 100% complete.

Results : Patients were aged 62 ± 15 years on average, were mostly male (66%) and hypertensive (69%). Risk of death early and late after operation decreased over the study period, with hospital mortality declining from 21% to 4% comparing the first and most recent quartiles (χ^2 for trend $P = 0.007$). At 1, 5, 10, and 20 years postoperatively, survival was 84%, 69%, 55%, and 30%, respectively, and freedom from cardiovascular death was 86%, 80%, 71%, and 51%, respectively. Additional independent risk factors for death were : older age ($P < .001$), renal dysfunction ($P < .003$), syncope ($P = 0.007$), and peripheral vascular disease ($P = 0.006$). During the study period, echocardiographic and CT diagnostic imaging replaced routine aortic angiography and operative techniques involved more frequent use of open distal anastomosis, retrograde cerebral perfusion, earlier restoration of antegrade perfusion, and conservative approach to aortic arch repair. Freedom from reoperation on the aorta or aortic valve was 93% and 84% at 5 and 10 years.

Conclusions : Early and late survival following repair of AAAD has improved progressively over 25 years in association with noticeable changes in preoperative and intraoperative management. Aortic reoperations were infrequent during follow-up.

ULTRAMINI-ABSTRACT

In a 25-year follow-up study, early and late survival following surgical repair of acute ascending aortic dissection has improved in association with changes in management. Current hospital mortality is 4%, and survival at 5, 10 and 20 years was 69%, 55% and 30% respectively. Aortic reoperations were infrequent during follow-up.

Word count : 50 words

Patients presenting with ascending aortic dissection within two weeks of the onset of symptoms are considered to have *acute* ascending aortic dissection (AAAD), while *chronic* presentation is defined when there is at least 2 weeks from the onset of symptoms. The importance of this diagnostic time frame is underscored by the reported 65-75% death rate of patients associated with untreated AAAD¹⁵². Hospital mortality for repair of AAAD has decreased since the 40% mortality rate reported by DeBakey in 1965¹⁵³ and now ranges from 5% to 26%^{72-76,152,154-164}. Among studies reporting long-term results following repair of AAAD, few reports are based on mean follow-up greater than 5 years, with 10-year survival estimates ranging from 37 to 71%^{72,73,75,165,166}.

This study sought to extend information regarding long-term cardiovascular and all cause survival after repair of AAAD, by review of the experience in a single institution over a 26 year period. Additionally, we sought to assess the impact of patient characteristics, and measure the changes in outcome over time which may have occurred with changes in preoperative evaluation and surgical techniques.

Material and Methods

Patients

A retrospective study was undertaken of patients having surgical repair of AAAD from October 1979 to December 2003. A database of all cardiac surgical operations performed at Massachusetts General Hospital was queried for mention of aortic dissection. Two hundred and seventy three patients were reviewed. We excluded 76 patients (28%) for the following reasons : chronic dissection (n = 35 ; 13%), descending aortic dissection (DeBakey type III / Stanford type B) (n = 34 ; 12%) and intraoperative aortic dissections or traumatic aortic injuries (n = 7 ; 3%). Two patients, excluded from this analysis, were considered inoperable after surgical exploration due to extensive dissection in the descending aorta with multiple rupture sites and/or cardiac arrest prior to cardiopulmonary bypass. The final cohort consisted of 195 patients who underwent surgical intervention for ascending aortic dissection within 2 weeks of symptom onset.

Patient demographics, comorbidities, operative details and postoperative outcomes were extracted retrospectively by chart review (Supplemental Methods). Involvement of aorta and its branches in the dissection process was determined by preoperative imaging and intraoperative assessment of the dissection process. Ascending aorta repair was completed with conservative techniques adapted to the dissection extension and involvement of the aortic valve including : ascending aorta replacement (AAR) without aortic valve replacement if aortic valve reconstruction was feasible or the disease process limited to the ascending aorta above the

sinotubular junction ; AAR with aortic valve replacement if aortic valve was dysfunctional and aortic sinuses were not involved ; and AAR with composite root replacement otherwise¹⁶⁷. In general, aortic arch repair was performed only when the intimal tear extended in the aortic arch. *Hemi-arch repair* was defined as a replacement that included the lesser curvature of the aortic arch beyond the level of the innominate artery but did not involve the aortic arch vessels. *Total arch repair* involved replacement of anterior and posterior aortic arch, with reimplantation of the branch vessels, either as an island, or as individual branch grafts. *Antegrade perfusion* refers to restoration of antegrade cerebral and corporal perfusion through the aortic graft, aortic arch or subclavian artery following distal aortic graft anastomosis.

Survival follow-up data were obtained by linking to the National Death Index (NDI) for all patients up to December 31, 2005. Date and cause of death were confirmed with NDI data. Aortic reoperation follow-up was obtained by telephone interview, mailed questionnaires to patients, family and primary care physician / cardiologist, querying the patient records and correspondence. *Aortic reoperation* was defined as any surgical procedure performed on thoracic or abdominal aorta or aortic valve after patient discharge from index admission. Ethics approval for this study was obtained from the Institutional Review Board.

Hospital mortality included death during the hospitalization and death within 30 days of surgery. *Cardiovascular death* included hospital death for index admission and follow-up death caused by cardiac ischemic disease, endocarditis,

other cardiac valve diseases, or vascular disease, including stroke and vascular surgery related deaths. *Hypertension* was recorded if the patient was clinically diagnosed with hypertension or was treated with antihypertensive medication prior to the index admission. *Peripheral vascular disease* comprised any known atheromatous involvement of systemic arteries with the exception of coronary arteries, and included cerebrovascular disease, abdominal aorta atheromatosis and/or aneurismal disease, intermittent claudication or previous vascular surgery. *Marfan syndrome* was defined according to the 1996 revised criteria¹⁶⁸ and histo-pathologic confirmation was sought whenever possible. Most patients were hypertensive following dissection. *Cardiogenic shock* was reported if the preoperative systolic blood pressure was less than 80 mmHg, or their cardiac index was less than 1.8 mL/m², or they required the use of IV inotropes. Patients were considered *unstable* if they experienced accelerating pain or hemodynamic instability (uncontrollable hypertension or hypotension). *Salvage* was applied if patients were undergoing cardiopulmonary resuscitation en route to the operating room or prior to anesthesia induction.

Statistical analysis

All statistical analyses were performed using SAS release 9.1 (SAS Institute Inc., Cary, NC, USA). Data are expressed as mean \pm standard deviation or median and range for continuous variables and as numbers (percentages) for categorical variables. Univariate analyses included two-tailed Wilcoxon rank sum or t test for continuous variables and Fisher's Exact Test for discrete variables (see Supplemental Methods for the list of variables). For continuous variables, the linearity assumptions were assessed for all models and transformations considered where appropriate. For multivariate analyses, missing data were handled by dummy coding for dichotomous covariates. Covariates with more than 5% missing data were discarded from the multivariate analyses. There was no missing data for continuous covariates. Stability of effect estimates in the final model was assessed by completing a bootstrap procedure with 5,000 random samples of 195 patients chosen with replacement from the original dataset. The standard deviations of the means of the bootstrap estimates were used to test the original parameter estimates from the final model by means of a χ^2 test. For all statistical analyses, p-values < 0.05 were considered to be statistically significant.

Several variables were strongly associated with year of surgery (and each other) and were, by necessity, excluded from the multivariate analysis (see Results). These were related to preoperative or treatment variables that became available during the study period, and thus may not have been possible, or were very difficult to apply, in the early part of the series. In order to assess competing diagnostic or

surgical strategies across time, we defined mutually exclusive categories, such as angiography alone, angiography with CT-scan, angiography with echocardiography, CT-scan alone, etc. A multinomial logit model was fitted using the categories as the outcome variable and date of surgery as the independent variable, thereby producing predicted probabilities across time, each representing the likelihood of being assigned to one of the category (and summing to 100% at each time point). Categories with similar trend and/or few cases were combined if clinically meaningful and the model was re-fitted.

Survival. Survival and aortic reoperation were estimated by the Kaplan-Meier technique, with log-rank tests to compare stratified groups. Time-specific effects were estimated with parametric, multiphase (mixture distribution components) hazard regression, using analytical tools available from the Cleveland Clinic Heart Center Hazard website (<http://www.clevelandclinic.org/heartcenter/hazard/default.htm>). A non-automated directed technique of stepwise variable entry was used. Multiphase parametric hazard regression was used allowing assessment of the impact of baseline predictors on different phases of follow-up separately. Kaplan-Meier and parametric estimate of the survival of patients discharged alive (Weibull distribution) were compared to the survival of a matched cohort for age, gender and race from Massachusetts using 1989-1991 State lifetables¹⁴⁵. Risk factors for aortic reoperation after discharge were assessed using a Cox model.

Results

Presentation

Patient characteristics, operative data, aortic repair techniques and concomitant procedures are reported in Table 1. Twenty-seven patients (14%) had prior aortic disease, 25 patients (13%) had prior aortic valve dysfunction and/or bicuspid aortic valve and 31 patients had prior cardiovascular surgery (7 CABG, 10 valve surgery, 10 major vascular surgery and 4 other). Thirteen patients (7%) were diagnosed with Marfan syndrome. EKG findings were most often non specific (44%), but 42 patients (23%) presented either ST elevation or depression. Nineteen patients (10%) underwent coronary artery catheterization prior to surgery.

Surgical techniques

Ascending aorta and aortic valve surgical procedures included 137 (60%) ascending aortic grafts including aortic valve resuspension, 47 (24%) composite valve and ascending aortic grafts (39 mechanical, 8 biological valves), 9 (5%) ascending aortic grafts and aortic valve replacements (6 mechanical and 3 biological valves) and 2 primary repairs (aneurysmoplasty without aortic graft). The average ascending aorta graft was 28 ± 3 mm and the median aortic valve replacement size was 27 mm (range 19 mm – 31 mm).

Initial arterial cannulation was most commonly performed at the femoral artery (96%), and rarely at the axillary artery (2%) or aortic arch (2%). Venous cannulation was performed using a two-staged right atrial canula (64%) or bicaval cannulas (26%) and/or through the femoral vein (31%). Most patients were given cold intermittent blood cardioplegia (98%) and 3 patients (2%) received warm continuous blood cardioplegia. Hypothermic circulatory arrest was used in 148 patients (76%) and the circulatory arrest lasted 32 ± 15 min with a minimal core temperature of 17 ± 4 °C.

Early results

Hospital mortality over the entire study period was 13% (26 patients), but improved dramatically during the study interval. When the study period was divided in 4 equal subgroups of consecutive patients, we observed a gradual decrease in hospital mortality : 21% (10/48 ; 1979-‘90), 18% (9/49 ; ‘90-‘96), 10% (5/49 ; ‘96-‘00) and 4% (2/49 ; ‘00-‘03) (χ^2 for trend $P = 0.007$). Intraoperative deaths occurred in 6 (23% of deaths) and were due to uncontrollable hemorrhage (3), post-CPB heart failure (2) or pulmonary artery rupture (1). Postoperative deaths were caused by heart failure (3), massive hemorrhage (2), stroke (9), renal failure (1) and multiple system

failure (5). Univariate predictors for hospital mortality are presented in Table 1. Independent predictors of hospital mortality are reported in Table 2.^j

In hospital complications included cardiac reoperation in 35 patients (18%), delirium in 50 (26%), persistent neurologic dysfunction in 37 (19%), pneumonia documented by positive sputum culture and chest imaging in 42 (22%), pleural effusion in 37 (19%), and temporary or permanent dialysis in 18 (9%). Most patients received transfusion of packed red blood cells (use 92% ; median 6 units), platelets (use 82% ; median 15 units) and fresh frozen plasma or cryoprecipitate (use 86% ; median 8 units). Patients with stroke had longer postoperative length of stay ($P = .02$) and their hospital mortality was 2.3 times higher compared to patients without stroke (24% vs. 11% ; $P = 0.03$). Independent predictors of postoperative stroke are listed in table 3. The mean maximum postoperative creatinine level was 2.3 ± 2.0 mg/dl and decreased to 1.3 ± 1.0 mg/dl upon discharge or death. Hospital mortality was 2.9 times higher among patients requiring dialysis postoperatively compared to those who were not dialysed (33% vs. 11% ; $P = .02$). Of the 169 patients discharged alive, 93 (55%) were discharged home including 24 who were visited by a nurse, with the remaining 76 patients (45%) discharged to a rehabilitation center.

^j Marfan syndrome was initially selected as a significant independent predictor of hospital mortality (Odd Ratio 24.61 [2.08-291.1]; $P = 0.01$), but the variable was ultimately dropped from the list of independent risk factors after we found that stability of this estimate was in question. With the bootstrap procedure, the variable was rejected with a p-value of 0.25; all other estimated p-values remained significant with $P < .001$. The final model had a c-statistics of 0.85 and a Hosmer and Lemeshow Goodness-of-Fit Test p-value of 0.6.

Long-term survival

The mean follow-up was 7.0 ± 5.9 years (range : 1 day to 26 years), or 8.1 ± 5.6 years in patients discharged alive, and was 100% complete. Overall survival was 84%, 69%, 55%, 48% and 30% at 1, 5, 10, 15, and 20 years, respectively and freedom from cardiovascular death was 86%, 80%, 71%, 67% and 51% at 1, 5, 10, 15 and 20 years, respectively (Figure 1). The figure also shows the high early risk of death followed by a low but slowly rising late hazard phase. Late deaths were due to ischemic cardiac diseases (14 ; 22%), cerebrovascular disease (4 ; 6%), vascular diseases (14 ; 22%), cardiac valve diseases including endocarditis (4 ; 6%), pulmonary diseases (6 ; 10%), septicaemia (3 ; 5%), cancer (9 ; 15%), gastro-intestinal diseases (4 ; 6%), external causes / accidents (3 ; 5%) and other unrelated causes of death (2 ; 3%). Survival for patients discharged alive is graphically compared to an age, gender and race-matched contemporary general population in Figure E1.

Early and late phase multivariate predictors for all-cause death and cardiovascular death are shown in Table 4. Note that the independent predictors of mortality in the early hazard phase included the same variables identified for hospital mortality by logistic methods, plus “salvage surgery” and “cross-clamp time” for all-cause death or “involvement of the carotid arteries” for cardiovascular death. The impact of age on early and long term survival is represented in Figure 2A. Figure 2B shows the multivariable-adjusted 1, 5, 10 and 15-year survival by year of surgery.

Freedom from aortic reoperation

The mean aortic reoperation follow-up for patients discharged alive was 6.7 ± 5.8 years and was 91% complete (16 patients lost to follow-up). Freedom from aortic reoperation was $97 \pm 1\%$, $93 \pm 2\%$, $84 \pm 4\%$, $66 \pm 8\%$ and $58 \pm 10\%$ at 1, 5, 10, 15, and 20 years, respectively (Figure E2). Aortic reoperations included aortic valve replacements (2 ; 8.3%), composite root replacements (6 ; 25%), arch repair (2 ; 8.3%), thoracoabdominal aneurysm repairs (8 ; 33%), abdominal aneurysm repairs (4 ; 17%) and surgical fenestrations of the abdominal aorta (2 ; 8.3%). Independent risk factors for aortic reoperation included year of surgery (Hazard Ratio [95% Confidence Interval] for 5 year increment : 2.30 [1.36-3.90] ; P = 0.002) and Marfan syndrome (HR [95%CI] : 5.97 [2.22-16.0 ; P <0.001).

Trends in management

Several variables were correlated with later date of surgery including more sophisticated diagnostic imaging modalities such as computed tomography (Pearson $r = 0.56$) and preprocedural transesophageal echocardiography ($r = 0.57$), open distal anastomosis ($r = 0.73$), antegrade cerebral perfusion ($r = 0.41$), use of antifibrinolytics ($r = 0.64$) and biological glue ($r = 0.51$) (all $P < .001$). The time trend in competing diagnostic imaging modality, circulatory arrest use and cerebral protection, and surgical techniques used is shown in Figure 3. Overall, aortic angiography was used in 97 patients (50%), computed tomography in 112 patients

(57%), preprocedural transesophageal echocardiography in 130 patients (67%) and magnetic resonance imaging in 4 patients (2%). There was a trend towards more preservation of the aortic valve and less aortic valve replacement ($P=0.05$), whereas the percentage of composite root replacement remained stable.

Discussion

Dissection of the ascending aorta remains one of the most common emergent procedures in cardiac surgery. Despite advances in stent graft technology, the relation of the ascending aorta to the aortic valve, coronary ostia, and arch vessels have left open repair as the only treatment in all but a few cases in select centers. Repair of AAAD typically involves a “palliation” of a more extensive disease process. It is uncommon that the operation removes the entire area of involved vasculature, with the focused goal being only to halt extension, improve branch flow, prevent rupture, and improve aortic valve function. Since the early series of surgical repair for acute ascending aortic dissection¹⁵³, these goals have remained unchanged, but changes have occurred with the incorporation of new preoperative testing methods, mainly computed tomography and echocardiography, and use of improved techniques, primarily related to myocardial and cerebral protection strategies.

In order to better characterize both the impact of the evolution in management of this disease, and the long-term survival following surgical treatment, this study was undertaken. We report long-term results of aortic repair for acute ascending aortic dissection over a 26 years period in a single, large volume referral center.

Long-term Survival

Few studies present 15+ year survival data following repair for AAA. In the early to the late '90, Crawford et al.¹⁶⁵, the Stanford group¹⁶⁶ and Bachet et al.⁷⁵ reported 10- and 15-year survival ranging from 37%-56% and 24%-46%, respectively. Likewise, we report 10-, 15- and 20-year survival of 55%, 48% and 30%, respectively. The risk of death is markedly reduced after the perioperative period, although the survival of patients discharged alive remains lower compared to an age-, gender- and race-matched general population.

Patient Factors Effecting Survival

Age is an established risk factor for early and late mortality^{74,158,166}. Some studies have focused on the outcomes of surgery for acute ascending aortic dissection in patients over 70 years^{169,170}. Elderly patients differ markedly from their younger counterparts for comorbidities such as long-standing hypertension, atherosclerosis, and are more likely to present in cardiogenic shock.

Extent of renal impairment measured by serum creatinine is a strong predictor of early mortality as also reported by others^{154,164,166,171}. Acute renal failure is associated with mesenteric ischemia, drug-resistant hypertension and aortic branch vessel compromise¹⁷¹. Some have suggested that delayed surgery or medical treatment with percutaneous fenestration or stenting has acceptable early outcomes in the setting of severe renal or intestinal malperfusion^{172,173}. However, this strategy

has to be weighed against the increased risk of rupture, which is difficult to quantify, and generally poor results reported for medical therapy^{161,164}.

Late death was predicted by peripheral vascular disease, which relates to the systemic involvement of arteries and prior vascular surgery. Patients with extensive peripheral vascular disease may be unable to compensate for reduced flow in the acute setting leading to end-organ malperfusion, with a significant toll on mortality^k.

Management Trends Associated with Improved Outcomes

During this era, we observed incorporation of a variety of new methods, and an associated reduction in hospital mortality from 21% in the early series to 4% recently and improvement in adjusted 5-year survival from 65% to 89%. We identified factors associated with this improvement, most important of which were advances in diagnostic imaging and improvement in ascending aortic repair techniques including liberal use of hypothermic circulatory arrest with open distal anastomosis and cerebral protection techniques.

Diagnostic imaging modalities have undeniably improved over the last two decades with a shift from invasive procedures with high contrast loads (aortography) to faster modalities with low contrast loads (computed tomography) or applicable at

^k *La maladie vasculaire périphérique sévère (MVAS) est un processus qui affecte toute la vascularisation artérielle et les organes cibles tels que le cerveau, le cœur et les reins. Elle est classiquement associée à une augmentation de la mortalité au suivi. L'ajout d'un phénomène de bas débit en aigu à la perfusion marginale due à la MVAS peut entraîner des atteintes plus ou moins sévères des organes cibles qui influenceront à leur tour la mortalité au suivi. L'effet sur la mortalité à court terme dû au bas débit est probablement expliquée par les variables « salvage » et « syncope ».*

bedside (echocardiography).¹ As technology improved, aortography was rarely requested as a unique modality and was gradually combined with computed tomography, preoperative echocardiography or both. As we move forward in the years of treatment, angiography was seldom requested whereas tomography and echocardiography were generally combined. The International Registry of Acute Aortic Dissection (IRAD)¹⁷⁴ and others¹⁷⁵ have shown that all diagnostic modalities have a high sensitivity for ascending aortic dissection (computed tomography 93-100%, transesophageal echocardiography 90-98% and aortography 87-88%), but the false negative rate is still considerable and mandate more than one modality to conclusively exclude the diagnosis. Magnetic resonance imaging was seldom used in our series.

Surgical techniques changed significantly during the study period. The practice of hypothermic circulatory arrest, visual inspection of the arch vessels and open distal anastomosis was introduced in the 1980's and became routine soon afterwards. In this series, the initial arterial cannulation site of choice was the femoral artery for rapid access, with early reestablishment of flow following distal aortic graft anastomosis and, more recently, conversion to antegrade perfusion through the aortic graft. Cerebral protection techniques were introduced in the 1990's and antegrade cerebral protection is now more popular. David et al.⁷⁶ suggested that avoidance of aortic cross-clamping during cooling and resection of the primary tear

¹ *L'amélioration des résultats de la chirurgie de la dissection aigüe de l'aorte ascendante dans le temps peut s'expliquer partiellement par un diagnostic plus précoce et plus précis de la pathologie aortique en raison de l'amélioration des techniques d'imagerie. Il n'est pas possible dans cette étude de commenter sur les patients avec dissection aigüe de l'aorte ascendante qui ont été imagés mais dont la maladie était jugée inopérable sur la base d'une imagerie de pointe plus précise ou d'autres facteurs clinique.*

improved early and late outcomes for acute ascending aortic dissection. They further suggested a decrease in the incidence of stroke with this technique, which was not shown in this series. Instead, prior cerebral deficit, aortic rupture into pericardium and use of circulatory arrest without cerebral protection were shown to be independent predictors of stroke. In patients undergoing circulatory arrest and presenting without prior cerebral deficit or aortic rupture, the use of simple cerebral protection techniques was associated with a trend towards a decrease in the rate of stroke (21% to 11% ; $P = 0.2$). As antegrade perfusion techniques have evolved, we tend to use more frequently the axillary artery or the aortic arch and its branches as the initial cannulation site when the patient is hemodynamically stable.

Year of surgery was also associated with improved late survival in addition to the impact seen on early survival. This may be related to better postoperative follow-up of these complicated patients with extensive vascular disease at specialized clinic or attentive follow-up from others. Diagnostic imaging techniques are becoming less invasive and more precise, both of which facilitates repeated follow-up examinations. The earlier detection of patients with vascular complication during follow-up and aggressive treatment of patient with Marfan syndrome may explain why patients are more likely to undergo an aortic reoperation in the later part of the series. The overall risk of aortic reoperation was low and comparable with published literature^{72-75,166}.

Limitations

This study represents the experience of a large academic referral center over a 26-year period with a mean follow-up of 7.0 ± 5.9 years, or 8.1 ± 5.6 years in patients discharged alive, and may not be generalizable to others centers. As a referral center, many patients are transferred some hours after the initial presentation and there may be some selection bias in the subpopulation of patients that actually underwent surgery, knowing that there must have been a few who died or were deemed inoperable at some point in their evaluation.

Aortic reoperation follow-up was limited. Also, information regarding long-term outcome would ideally have included additional information on patient functional status collected prospectively.

Attempts to understand the factors which lead to observed changes over time were made difficult due to the fact that multiple “new” variables were appearing during the interval. These were highly correlated, and thus identifying (or excluding) any individual factor became statistically impossible in some cases. Other factors may account for the improvement in early and late survival, including earlier referral, improved intensive care unit performance, and secondary or tertiary prevention measures.

Conclusion and clinical inferences

Over the past 25 years, early and late outcomes for acute ascending aortic dissection have improved in association with the incorporation of new methods and technologies, including echocardiography, CT imaging, use of open distal anastomosis, and improved cerebral protection techniques. The reported, current, early mortality of less than 5% was achieved using a strategy that emphasizes : avoidance of coronary angiography, combined use of CT and TEE data, meticulous surgical technique involving routine use of hypothermic circulatory arrest with open distal anastomosis and conservative approach to aortic arch repair.

Acknowledgements

We are grateful for the research assistance provided by Sandra Debronkart, Karen E. Lynch, Jerene Bitondo and Suyog Mokashi. Salary support (to LMS) was provided by a fellowship award from the Canadian Institutes of Health Research's Clinical Research Initiative and the Rosetti Fund (Massachusetts General Hospital, Boston, Massachusetts).

Table 1. Patient characteristics, operative data and hospital mortality

	Overall	Hospital mortality*			
		<i>Exposed</i>	<i>Unexposed</i>	<i>Odds Ratio</i>	<i>P</i>
Overall	195 (100%)	26/195 (13%)	–	–	–
<i>Patients characteristics</i>					
Age, years	62 ± 15	–	–	1.84† [1.26-2.70]	0.002
Male gender	128 (66%)	16/128 (13%)	10/67 (15%)	0.81 [0.35-1.91]	0.6
Hypertension	134 (69%)	21/134 (16%)	5/61 (8%)	2.08 [0.75-5.81]	0.2
Coronary artery disease	32 (16%)	6/32 (19%)	20/163 (12%)	1.65 [0.60-4.50]	0.3
Chronic lung disease	23 (12%)	6/23 (26%)	20/172 (12%)	2.68 [0.95-7.60]	0.06
Peripheral vascular disease	38 (19%)	5/38 (13%)	21/157 (13%)	0.98 [0.34-2.80]	>.9
Serum creatinine (mg/dl) at admission	1.24 ± 0.51	–	–	2.47 [1.26-4.84]	0.009
Prior cardiac valve dysfunction	35 (18%)	3/35 (9%)	23/160 (14%)	0.56 [0.16-1.97]	0.4
<i>Presenting symptoms</i>					
Sudden onset pain	144/190 (76%)	18/144 (13%)	8/46 (17%)	0.68 [0.27-1.68]	0.4
Anterior chest pain	143/190 (75%)	17/143 (12%)	9/47 (19%)	0.57 [0.23-1.38]	0.2

	Hospital mortality*				
	Overall	Exposed	Unexposed	Odds Ratio	P
Head/neck pain	45/190 (24%)	3/45 (7%)	23/145 (16%)	0.38 [0.11-1.33]	0.1
Back pain	93/190 (49%)	10/93 (11%)	16/97 (16%)	0.61 [0.26-1.42]	0.3
Abdominal pain	39/190 (21%)	3/39 (8%)	23/151 (15%)	0.46 [0.13-1.63]	0.2
<i>Clinical status</i>					
Hemodynamic status					0.02
- Stable (ref.)	112 (57%)	9/112 (8%)	–	1	
- Unstable	71 (36%)	13/71 (18%)	–	2.57 [1.03-6.37]	
- Salvage	12 (6%)	4/12 (33%)	–	5.72 [1.44-22.8]	
Cardiogenic shock	45 (23%)	11/45 (24%)	15/150 (10%)	2.91 [1.23-6.91]	0.02
Tamponade	26 (13%)	6/26 (23%)	20/169 (12%)	2.24 [0.80-6.23]	0.1
Syncope	33 (17%)	12/33 (36%)	14/162 (9%)	6.04 [2.47-14.80]	<.001
Altered mental status	23 (12%)	5/23 (22%)	21/172 (12%)	2.00 [0.67-5.95]	0.2
Focal neurological deficit pre-op	12 (6%)	3/12 (25%)	23/183 (13%)	2.32 [0.58-9.20]	0.2
EKG ST-elevation or ST-depression	42/186 (23%)	8/42 (19%)	17/144 (12%)	1.76 [0.70-4.42]	0.2

	Overall	Hospital mortality*			P
		Exposed	Unexposed	Odds Ratio	
<i>Arteries involved in the dissection process</i>					
DeBakey type I	162 (83%)	20/162 (12%)	6/33 (18%)	0.63 [0.23-1.72]	0.4
DeBakey type II	33 (17%)	6/33 (18%)	20/162 (12%)	1.58 [0.58-4.29]	
Coronary arteries	40 (21%)	6/40 (15%)	20/155 (13%)	1.19 [0.44-3.20]	0.7
Carotid arteries	43 (22%)	8/43 (19%)	18/152 (12%)	1.70 [0.68-4.24]	0.3
<i>Surgical data</i>					
Interval symptom-surgery (hours)	24 [0.9-480]	–	–	1.01 [0.90-1.14]	0.9
CPB time (min)	166 ± 70	–	–	1.02‡ [0.96-1.07]	0.6
Cross-clamp time (min)	108 ± 44	–	–	0.93‡ [0.83-1.05]	0.2
Circulatory arrest time (min)	32 ± 15	–	–	0.81‡ [0.54-1.23]	0.3
Open distal anastomosis	146 (75%)	15/146 (10%)	11/49 (22%)	0.40 [0.17-0.93]	0.03
Retrograde cerebral perfusion	24 (12%)	0/24 (0%)	26/171 (15%)	–	–
Antegrade cerebral perfusion through arch or graft	56 (29%)	5/56 (9%)	21/139 (15%)	0.55 [0.20-1.54]	0.3

	Hospital mortality*				
	Overall	Exposed	Unexposed	Odds Ratio	P
Concomitant hemi-arch repair	18 (9%)	2/18 (11%)	24/177 (14%)	0.80 [0.17-3.69]	0.8
Concomitant total arch repair	5 (3%)	1/5 (20%)	25/190 (13%)	1.65 [0.18-15.36]	0.7
Concomitant CABG	14 (7%)	3/14 (21%)	23/181 (13%)	1.87 [0.49-7.22]	0.4
Antifibrinolytic use (Aprotinin or Amicar)	108 (55%)	7/108 (6%)	19/87 (22%)	0.25 [0.10-0.62]	0.003
Surgical glue use	44 (23%)	1/44 (2%)	25/151 (17%)	0.12 [0.02-0.89]	0.04
Year of surgery (since Oct 1 st , 1979)	15 ± 7	–	–	0.65§ [0.48-0.88]	0.005

* Hospital mortality in patients exposed (presence of) or unexposed (absence of) to each categorical risk factor (ex.: Male vs. Female, Hypertensive vs. Non-Hypertensive, etc.) with corresponding odds ratio and univariable p-values from logistic regression models. † Effect for 10-years difference in age. ‡ Effect for 10-minutes difference in CPB, cross-clamp or circulatory arrest time. § Effect for 5-year difference in the year of surgery.

CPB, cardiopulmonary bypass.

Table 2. Independent predictors of hospital death

	Odds Ratio [95% C.I.]	<i>P</i>
Age	2.40 [†] [1.44-3.99]	<.001
Serum creatinine (in mg/dL) at admission*	4.67 [1.18-18.52]	0.03
Syncope	4.66 [1.68-12.88]	0.003
Year of surgery	0.54 [‡] [0.36-0.80]	0.002

* Natural logarithmic transformation of serum creatinine (in mg/dL). [†] Effect for 10-years difference in age. [‡] Effect for 5-year difference in the year of surgery.

Table 3. Independent predictors of postoperative stroke

	Odds Ratio [95% C.I.]	<i>P</i>
Altered mental status	2.96 [1.22-7.17]	0.02
Aortic rupture into pericardium	3.64 [1.29-10.25]	0.01
Circulatory arrest		0.04
- None (reference)	1	-
- Without cerebral protection	3.22 [1.12-9.26]	0.01
- With cerebral protection	1.33 [0.40-4.41]	0.5

Table 4. Incremental risk factors for all-cause and cardiovascular death

	Risk factors	Coefficient ± SD	P
<u>ALL-CAUSE DEATH</u>	<i>Early hazard phase</i>		
	- Age	0.06 ± 0.018	0.002
	- Serum creatinine (in mg/dL) at admission*	1.7 ± 0.57	0.003
	- Syncope	1.1 ± 0.41	0.007
	- Salvage	1.3 ± 0.57	0.02
	- Cross-clamp time†	1.7 ± 0.74	0.02
	- Year of surgery‡	-0.5 ± 0.15	<.001
	<i>Late hazard phase</i>		
	- Age	0.06 ± 0.013	<.001
- Peripheral vascular disease	1.0 ± 0.34	0.006	
- Year of surgery	-0.05 ± 0.020	0.02	
<u>CARDIOVASCULAR DEATH</u>	<i>Early hazard phase</i>		
	- Age	0.07 ± 0.018	<.001
	- Serum creatinine (in mg/dL) at admission*	1.3 ± 0.54	0.02
	- Syncope	1.3 ± 0.38	<.001
	- Involvement of the carotid arteries	1.0 ± 0.42	0.02
	- Year of surgery‡	-0.4 ± 0.13	0.002
	<i>Late hazard phase</i>		
- Age	0.07 ± 0.023	0.001	

* Natural logarithmic transformation of serum creatinine (in mg/dL). † Natural logarithmic transformation of cross-clamp time (in min). ‡ Squared transformation of years since October 1st, 1979.

Figure Legends

- Figure 1 Survival (A) and freedom from cardiovascular death (B) after surgery for all patients. Actuarial survival estimates with 1 SD (68% confidence limits) and size of risk sets at 1, 5, 10, 15 and 20 years after surgery are provided in the embedded table. The predicted survival from each parametric multiphase model is overlaid as a bold solid line in the same illustration with a bold dashed line indicating 1 SD. The thin solid line represents the hazard function (as percent death per year) for each model, with the dotted line indicating 1 SD. The left vertical axis represents percent survival and the right vertical axis percent death per year.
- Figure 2 Multivariable-adjusted survival for patient aged 35, 50, 65 and 80 years (A) and multivariable-adjusted 1, 5, 10 and 15-year survival by year of surgery (B). Predictions are presented with 1 SD (68% confidence limits) and are truncated according to the maximum follow-up observed in each group. In these models, continuous variables are fixed at their mean values (except for age in model A and year of surgery for model B) and binary variables are set to zero.
- Figure 3 Changing patterns in management of acute ascending aortic dissection. The depictions highlight changes observed during the study period, each curve representing the relative occurrence of mutually exclusive management strategies over time adding up to 100% at every time point. Figure A demonstrates the change in utilization of preoperative

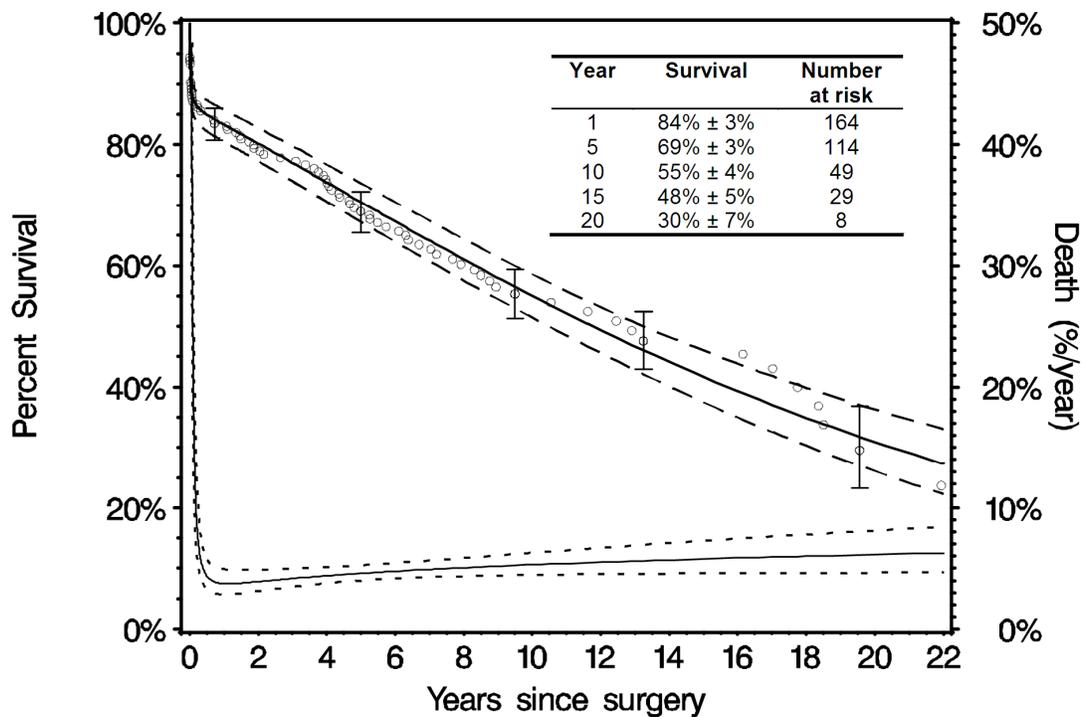
diagnostic imaging modalities. Angiography is replaced gradually with the incorporation of computed tomography and echocardiography. Figure B depicts the increasing use of circulatory arrest and antegrade/retrograde cerebral perfusion.

Figure E1 Survival after surgery for all hospital survivors compared to Massachusetts' general population matched for age, gender and race. Actuarial survival estimates with 1 SD (68% confidence limits) and size of risk sets at 1, 5, 10, 15 and 20 years after surgery are given in the embedded table. The predicted survival from the parametric multiphase model is overlaid as a bold solid line in the same illustration with a bold dashed line indicating 1 SD. The bold dotted line represents the matched survival for the Massachusetts' general population.

Figure E2 Freedom from aortic reoperation on aorta or aortic valve for all hospital survivors. Actuarial survival estimates with 1 SD (68% confidence limits) and size of risk sets at 1, 5, 10, 15 and 20 years after surgery are given in the embedded table.

Figure 1.

A



B

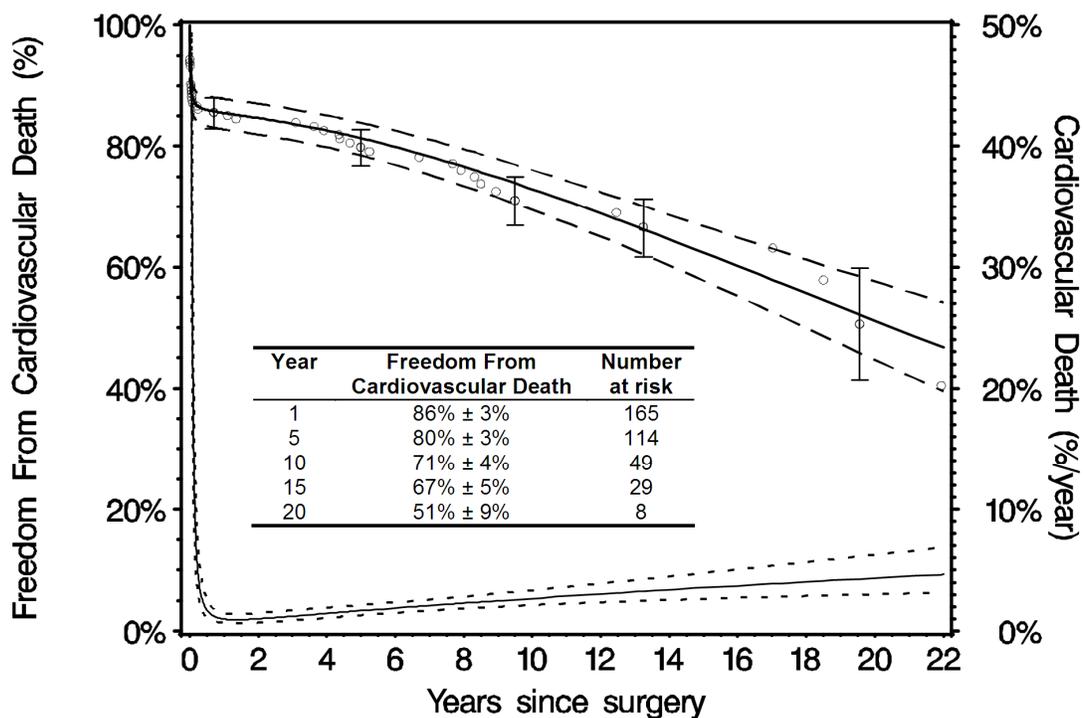


Figure 2.

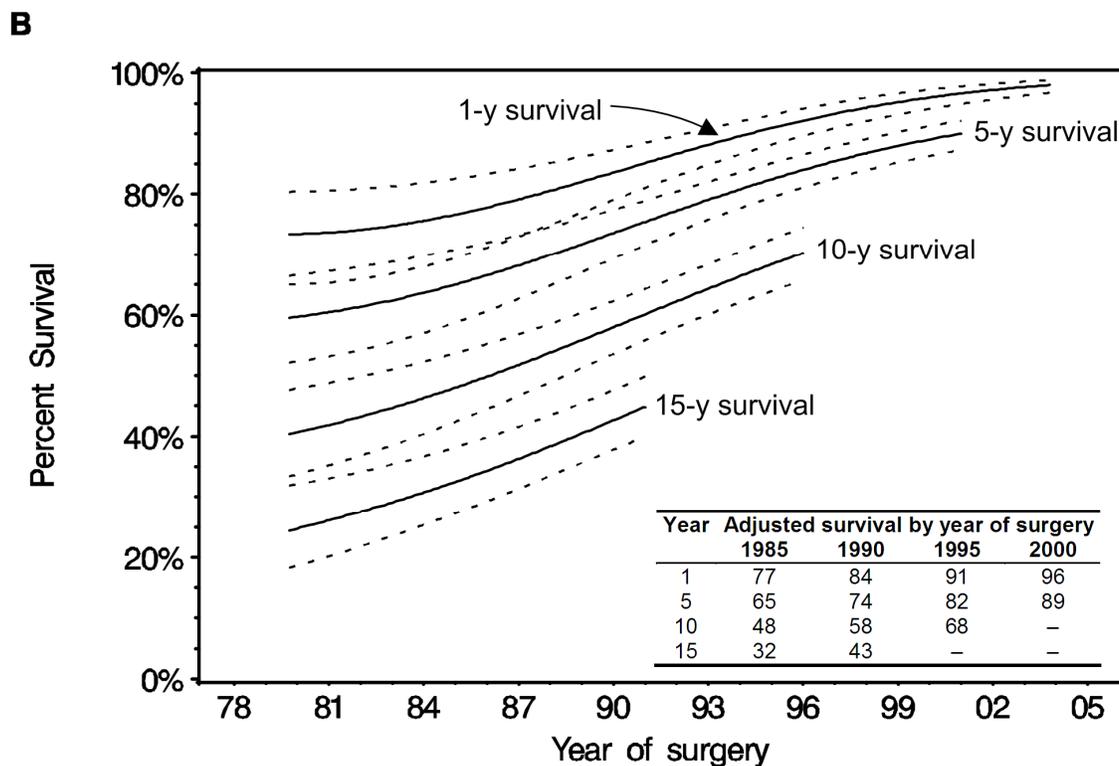
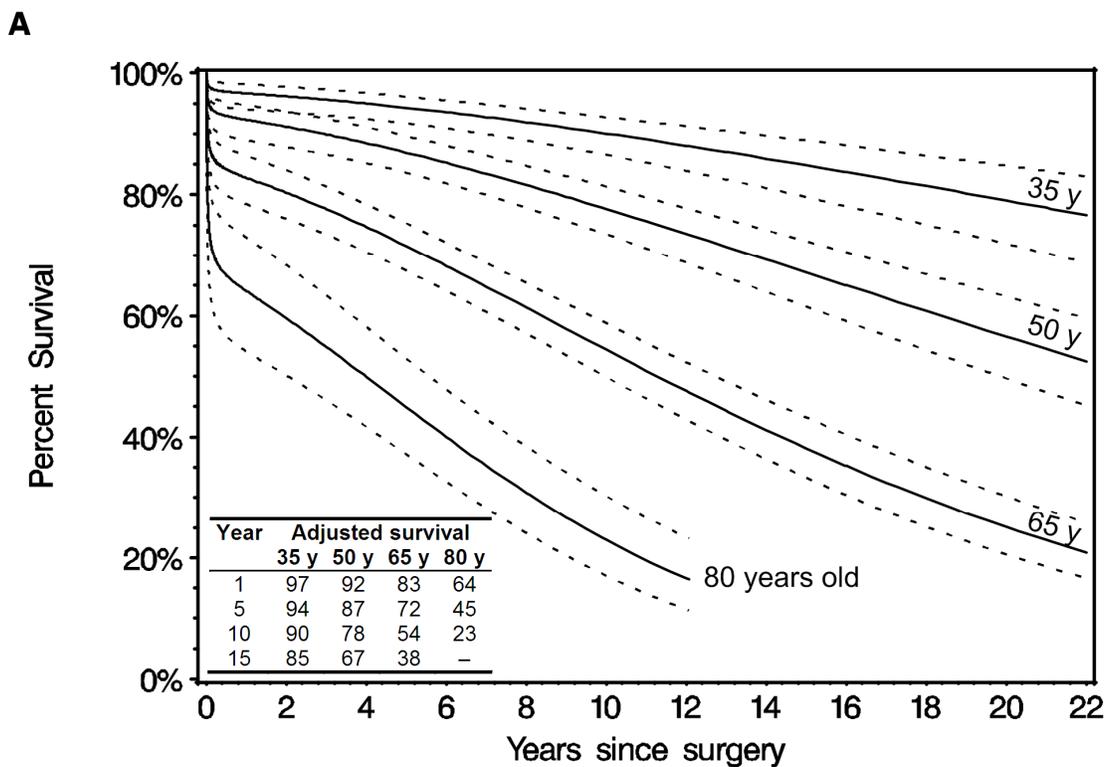
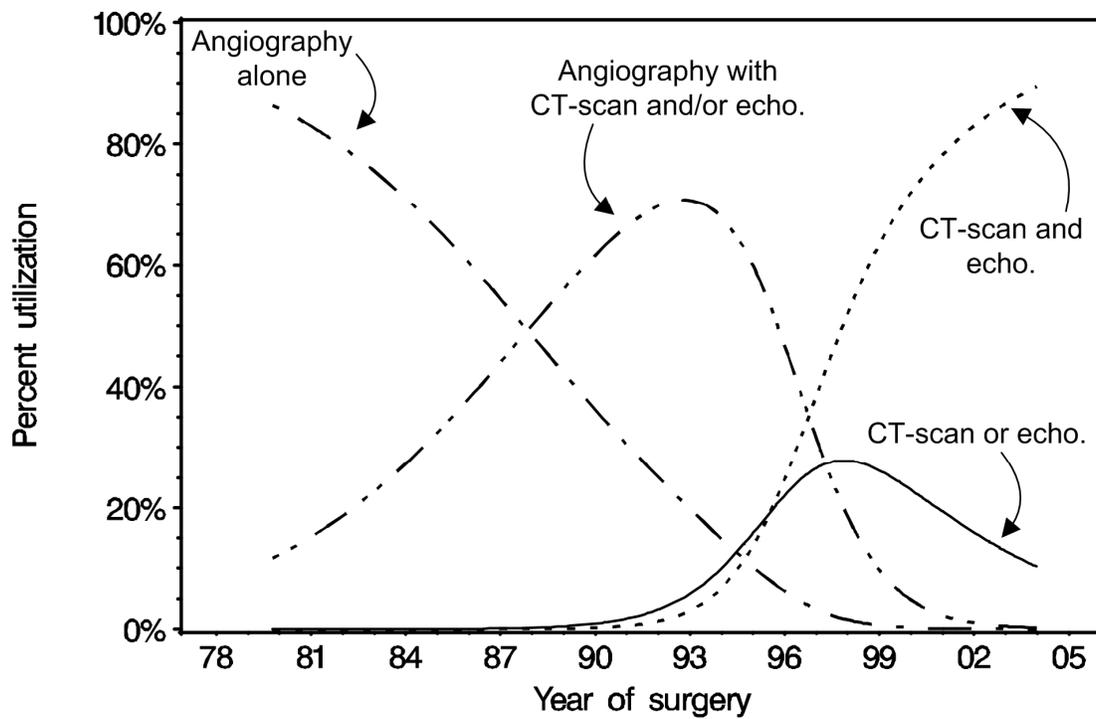


Figure 3.

A



B

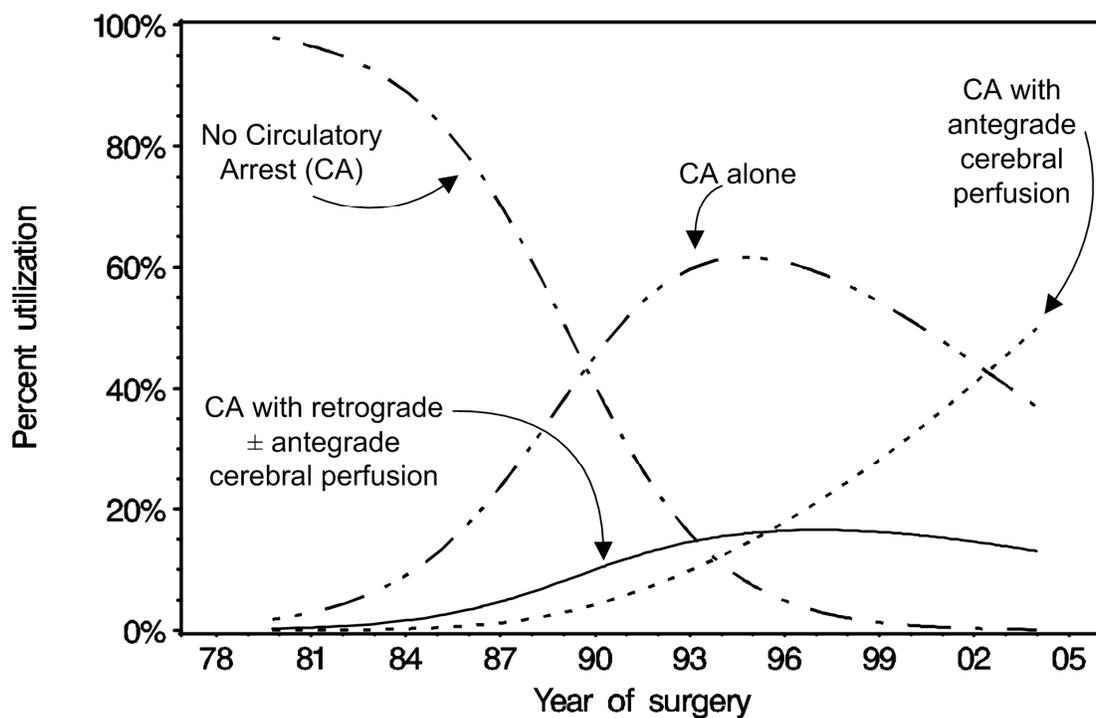


Figure E1.

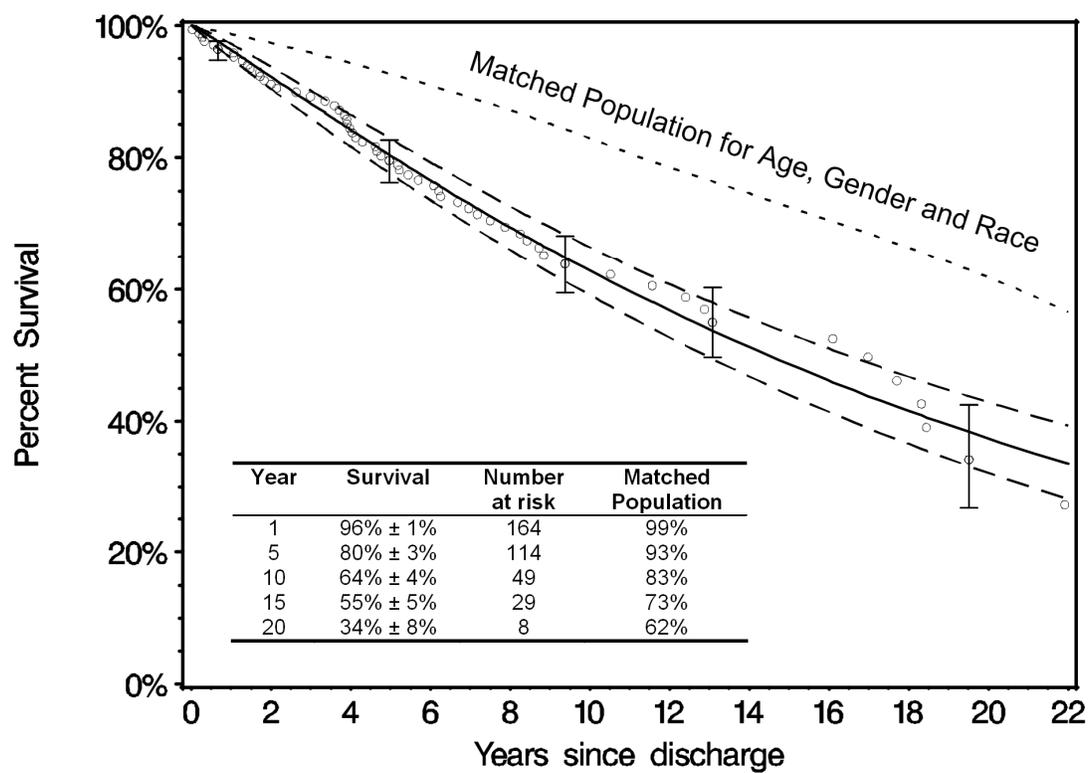
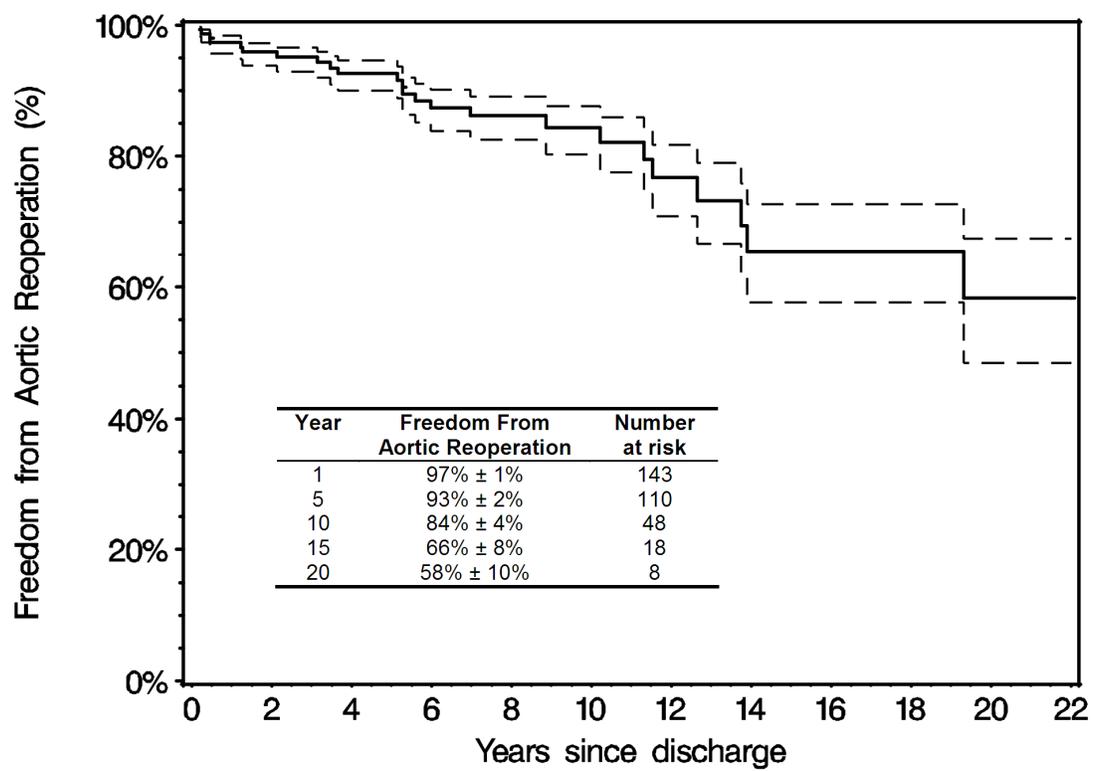


Figure E2.



- SUPPLEMENTAL MATERIAL -

Evolution in Surgical Management and Outcome for Acute Ascending Aortic

Dissection

Supplemental Methods. Variable examined for association with outcomes

Patient variables

Demography : age, gender

Presentation : interval between symptom onset and surgery, hemodynamic state at operation (stable, unstable, cardiogenic shock), characteristic of pain (sudden) and location (anterior, posterior, neck, abdominal, extremities), altered mental status, focal neurologic deficit, paraplegia, syncope

Status of aorta : known extend of dissection (ascending, descending coronary artery, carotid arteries), aortic valve regurgitation, bicuspid valve disease, aortic rupture into pericardium

Comorbidities : hypertension, coronary artery disease, previous myocardial infarction, EKG St-elevation or depression, dyslipidemia, chronic heart failure, chronic obstructive pulmonary disease, smoking, peripheral vascular disease, Marfan disease, previous cardiac surgery, previous vascular surgery, serum creatinine level

Procedure

Cardiopulmonary bypass (CPB) : CPB and cross-clamp time, use of circulatory arrest and its duration

Surgery : type of procedure performed (ascending aorta replacement (AAR) without valve procedure, AAR with either composite root replacement or aortic valve replacement, AAR with aortic valve resuspension), hemiarch or total arch repair, open distal anastomosis, use of biological glue, use of antifibrinolytic

Year of surgery : years since Oct. 1979

CHAPITRE III

**CORONARY ARTERY BYPASS GRAFTING AFTER RECENT
OR REMOTE PERCUTANEOUS CORONARY INTERVENTION
IN THE COMMONWEALTH OF MASSACHUSETTS**

ABSTRACT

Background : In this study, we sought to characterize the outcomes following isolated coronary artery bypass grafting (CABG) in patients with a history of remote (≥ 14 days), and recent (< 14 days), percutaneous coronary intervention (PCI).

Methods and Results : Patients with PCI within 5 years of CABG were identified among 12,591 primary isolated CABG reported in the mandatory Massachusetts Adult Cardiac Surgery Database. Patients were excluded if they were out-of-State ($n=1,043$, 8%), had undergone primary PCI for acute myocardial infarction ($n=401$, 3%), had a PCI-CABG interval > 5 years or unknown ($n=136$ and $n=673$, 1% and 5%). Patients with a history of remote and recent PCI were analyzed separately. Each CABG patient with PCI was matched to three patients without PCI using a propensity score. Outcomes were analyzed using generalized estimating equations and stratified proportional hazards models, with a mean follow-up of 4.1 ± 1.2 years.

There were 1,117 CABG patients (9%) with prior PCI ($n_{\text{remote}}=823$; $n_{\text{recent}}=294$). In matched CABG patients with remote prior PCI, no differences were found in 30-day mortality (1.1% vs. 1.5% ; $P=0.432$), hospital morbidity (41% vs. 40% ; $P=0.385$) and overall survival (hazard ratio [95% confidence interval] for death for prior PCI 0.93 [0.74-1.18] ; $P=0.555$). In matched CABG patients with recent prior PCI, hospital morbidity was higher (59% vs. 45% ; $P<0.001$), but no difference were found in 30-day mortality (3.5% vs. 3.1% ; $P=0.754$) and overall survival (HR 1.18 [0.83-1.69] ; $P=0.353$).

Conclusions : In patients undergoing CABG, remote prior PCI (≥ 14 days) was not associated with adverse outcomes at 30 days or during long-term follow-up.

The last two decades have witnessed a dramatic increase in the rates of percutaneous coronary interventions (PCI), with this technology expanded to more complex patient subgroups including diabetes⁸¹, severe left ventricular dysfunction⁸², left main⁸³⁻⁸⁵ and triple vessel disease⁸⁶⁻⁹⁰. A recent meta-analysis of randomized controlled trials of PCI stenting vs. coronary artery bypass grafting (CABG) for multivessel disease demonstrated a similar cumulative incidence of death, stroke and myocardial infarction at 5 years, but an increase in repeat revascularization in the PCI population⁸⁷. Conversely, large real-world registries comparing PCI stenting with bare-metal or drug-eluting stents vs. CABG demonstrated a lower incidence of death or myocardial infarction with CABG^{86,88,90}. In randomized trials and real-world registries, PCI is associated with an increased incidence of symptom recurrence and repeat revascularization compared to CABG.

Several studies have reported worse outcomes for patients undergoing CABG following acute PCI associated with complications¹⁷⁶⁻¹⁸⁰, but few studies have assessed outcomes following CABG procedures after initially successful and uneventful PCI¹⁸¹⁻¹⁸³. Moreover, no study has explicitly excluded patients undergoing primary PCI for acute myocardial infarction, where patients are more likely to be treated solely for their culprit lesion, with incomplete revascularization.

We hypothesized that patients with a history of prior PCI undergoing primary isolated CABG experience poorer survival compared to CABG patients with no prior PCI, with a greater difference in patients with shorter PCI-to-CABG interval compared to those with longer PCI-to-CABG interval. Our objectives were to describe patient characteristics and outcomes of real world patients who underwent primary isolated

CABG after uneventful PCI and contrast them to a matched group of CABG patients with no prior PCI. The primary end-point was overall survival and secondary end-points were 30-day mortality, hospital morbidity and length of stay.

Material and Methods

Patients

Ethics approval for this study was obtained from the Partners HealthCare Institutional Review Board. We linked clinical information from the mandatory Massachusetts Adult Cardiac Surgery Database (MASSDAC) to the Massachusetts Inpatient Acute Hospital Case Mix and Charge Database in order to analyze outcomes of 12,591 adults undergoing primary isolated CABG between 2002 and 2004. All fourteen state licensed cardiac surgery programs in Massachusetts submitted patient and surgeon specific data to the MASSDAC coordinating center at Harvard Medical School. Hospital data submission were cleansed and verified using a variety of procedures: continuous feedback via ongoing data quality reports, meetings and communication, review of concordance with administrative datasets, and review of concordance with medical chart audits. However, variables describing whether or not the patient had PCI in the past were not part of the auditing process. Details of PCI interventions, including dates, were not specified. Therefore, inpatient discharge data for all Massachusetts' acute care hospitals (n=84) were used to identify any occurrence of prior PCI in the 5 year interval prior to the primary

CABG (*ICD-9-CM codes* 36.01-36.02, 36.05-36.07 and 36.09). There were 61 patients (6%) with prior PCI in the administrative database who were not coded as prior PCI in MASSDAC. These patients were considered to have an unknown PCI-CABG interval.

We excluded 402 (3%) patients with a Unique Health Information Number not linkable to the inpatient discharge database, 1,043 (8%) patients residing outside of Massachusetts at time of primary CABG, 401 (3%) patients with primary PCI for acute myocardial infarction, 136 (1%) patients with PCI >5 years prior to CABG and 673 (5%) patients with unknown PCI-CABG intervals. *Primary PCI for acute myocardial infarction* was defined as PCI during hospitalization with a primary diagnosis of acute myocardial infarction (*ICD-9-CM codes* 410.X0-410.X1, excluding subendocardial infarction 410.70-410.71) or a primary diagnosis of a complication of an acute myocardial infarction (*ICD-9-CM codes* 785.51, 785.59, 429.5, 429.6 and 429.71) with a secondary diagnosis of acute myocardial infarction.

In this large cohort of 9,934 patients, we identified 1,117 CABG patients (9%) with prior PCI, including 294 patients with recent PCI and 823 patients with remote PCI. *Recent PCI* was defined as prior PCI within 14 days of the CABG surgery and *remote PCI* as prior PCI performed 14 days or greater prior to CABG surgery. Patients with a history of recent and remote PCI were analyzed separately.

Thirty-day mortality was defined as any death within 30 days of surgery. *Morbidity* outcomes adhered to STS database definitions (version 2.52.1)¹⁸⁴ and included 30-day or hospital mortality, any early postoperative reoperation,

perioperative myocardial infarction, deep sternal wound infection, neurologic deficit (transient ischemic attack, reversible ischemic neurologic deficit, stroke, coma), prolonged ventilation (>24 hours postoperatively), pneumonia, renal failure (serum creatinine >2.0 mg/dl or >2 times preoperative value or requiring dialysis), acute limb ischemia, aortic/iliac/femoral dissection, arrhythmia (atrial fibrillation, heart block, cardiac arrest), anticoagulation-related complication, cardiac tamponade, gastro-intestinal complications, and multi-organ failure. Data was linked to the Massachusetts Mortality Index Database with complete follow-up until September 30, 2007. The mean follow-up was 4.1 ± 1.2 years (range : 1 to 2129 days).

Statistical analysis

Data are expressed as mean \pm standard deviation or median and range for continuous variables, and as frequencies (percentages) for categorical variables. Univariable and multivariable regression models were constructed. Variables associated with P-values < 0.2 in univariable analyses were considered in multivariable models. As an additional effort to control for systematic differences between prior PCI and no prior PCI groups, a propensity score was constructed. Propensity for prior PCI was determined by logistic regression analysis where each patient with prior PCI was matched to three patient with no prior PCI, using a 1:N matching macro based on greedy algorithms¹⁸⁵. Variables used in the propensity score model were determined prior to obtaining the data (see Supplemental Methods). A different propensity score was calculated for patients with a history of

recent and remote PCI. The 1:N greedy matching algorithm starts with a 1:1 match where it determines the "best" match based on the propensity score, with "next-best" matches listed in hierarchical order. Cases are then matched to the set of remaining controls in N-1 additional iterations. If a case does not have N matched controls, it is removed from the set of matches at the time it fails to receive a matched control. The corresponding control is also removed from the set of matches and added back to the pool of un-matched controls for subsequent matching.

Patient characteristics before and after matching were compared using standardized differences. For 30-day mortality and hospital morbidity, logit-link generalized estimating equations were used, accounting for clustering among hospitals in the univariable and multivariable non-matched analyses (i.e. without propensity score) and for matching in univariable matched analyses. Similarly, identity-link normal-distribution generalized estimating equations were used for cardiopulmonary bypass time, cross-clamp time, number of grafts performed, and log-transformed length of stay. After verification of proportionality assumptions, survival was compared using stratified proportional hazards models to account for clustering among hospitals in the non-matched analysis and to account for matching in the matched analysis. We assumed that observations were conditionally independent within hospital or matched clusters. Survival curves were obtained using Kaplan-Meier methods ; however, group comparisons were performed using stratified proportional hazards models. P-values less than 0.05 were considered statistically significant. All statistical analyses were performed using SAS release 9.1 (SAS Institute Inc., Cary, NC, USA).

Results

CABG patients with remote prior PCI (≥ 14 days)

In CABG patient with remote PCI, the median interval between PCI and CABG was 236 days [15 to 1825 days] (Figure 1). Patient characteristics for all patients and matched groups with remote prior PCI are reported in Table 1. Compared to non-matched CABG patients with no prior PCI, non-matched CABG patients with remote prior PCI were younger, more likely to be female and present with diabetes, angina, dyslipidemia and larger body surface area. They were less likely to present with arrhythmia, left main disease, three-vessel disease, low left ventricular ejection fraction, an urgent status, or require an intraaortic balloon pump preoperatively. In the 809 CABG patients with remote prior PCI that were successfully matched to 2427 controls, no differences in patient characteristics were noted, except for a higher prevalence of angina, NYHA functional class III-IV and myocardial infarction >24 hours prior to CABG surgery in the prior PCI group.

Operative data and postoperative outcomes for the matched groups with remote prior PCI are shown in Table 2. Patient with prior PCI received fewer coronary grafts, less frequently had a left internal thoracic artery graft, and had lower cross-clamp times. There was no clustering for lower use of left internal thoracic artery graft in specific hospitals ; the decreased use of the LIMA was observed in all hospital with a sufficiently large number of patients to assess this observation.

Among non-matched CABG patients, remote prior PCI was not identified as a risk factor for 30-day mortality in univariable (OR [95% CI] 0.57 [0.26 - 1.22] ; P=0.145) or multivariable (OR=0.82 [0.34 - 2.02] ; P=0.672) analyses adjusting for age, sex, peripheral vascular disease, renal failure, arrhythmia, left ventricular ejection fraction, cardiogenic shock, and left main disease. The 30-day mortality was no different for non-matched (1.1% vs. 1.9% ; P=0.145) vs matched (1.1% vs. 1.5% ; P=0.432) CABG patients with or without remote prior PCI.

Overall hospital morbidity was also no different for non-matched (41% vs. 43% ; P=0.106) and matched (Table 2) CABG patients with or without remote prior PCI. Matched CABG patients with remote prior PCI received more blood product transfusions, without a difference in rate of reoperation for bleeding. The median length of stay was similar for CABG patients with or without remote prior PCI.

Three hundred twenty-two patients (39% of CABG patients with remote prior PCI) had 2 or more admissions for PCI in the 5-year interval preceding CABG surgery. The 30-day mortality was no different for patients with one vs. multiple admissions for PCI before CABG surgery (1.2% vs. 0.9% ; P=0.776).

Among non-matched patients, remote prior PCI was identified as a predictor of improved long-term survival in univariable (HR 0.74 [0.60-0.91] ; P=0.004) but not multivariable regression models (Supplemental Table 1). As portrayed in Figure 2A, survival was better for non-matched CABG patients with remote prior PCI compared to CABG patients with no prior PCI. However, no difference in survival was noted for matched groups (Figure 2B).

CABG patients with recent prior PCI (<14 days)

In the recent prior PCI group, 131 patients (45%) had a myocardial infarction at the time of PCI or following the procedure. The indication for CABG was unstable angina with referral to CABG within 6 hours of PCI (n=86 ; 29%), 6-24 hours after PCI (n=111 ; 38%) or greater than 24 hours after PCI (n=24 ; 8%) ; stable angina after PCI (n=55 ; 19%) ; or no documented angina but either two- or three-vessel disease (n=18 ; 6%). Ninety-seven patients with recent prior PCI (33%) required an intra-aortic balloon pump prior to CABG, including 23 patients (8%) who did not have unstable angina. Compared to non-matched patients with no prior PCI, non-matched patients with recent prior PCI were younger (standardized difference, SD=20%) and more likely to be female (SD=12%), have a prior myocardial infarction (SD=34%), angina (SD=15%), cardiogenic shock (SD=39%), require an intraaortic balloon pump preoperatively (SD=58%), and present in an emergent/salvage status (SD=59%) (Supplemental Table 2). They were less likely to present with left main (SD=-25%) or three-vessel disease (SD=-44%). In the 260 CABG patients with recent prior PCI that were successfully matched to 780 CABG patients with no prior PCI, no differences in patient characteristics were noted except for a higher prevalence of prior myocardial infarction (SD=25%) and intraaortic balloon pump use preoperatively (SD=12%).

Fewer grafts were received and the left internal thoracic artery and radial artery grafts were less frequently used by matched patients with recent prior PCI. The 30-day mortality was higher in non-matched patients with recent prior PCI

(4.1% vs. 1.9% ; P=0.033) but no different in matched patients with or without prior PCI (3.5% vs. 3.1% ; P=0.754). Overall hospital morbidity was higher for both the non-matched (58% vs. 43% ; P<0.001) and matched (59% vs. 45% ; P<0.001) CABG patients with prior PCI. Matched CABG patients with recent prior PCI received more blood product transfusions (69% vs. 54% ; P<0.001), had a higher rate of reoperation for bleeding (4.2% vs. 1.4% ; P=0.010) or graft revision (2.3% vs. 0.1% ; P=0.007), perioperative myocardial infarction (13% vs. 4% ; P<0.001) and prolonged ventilation greater than 24 hours (25% vs. 13% ; P<0.001). The incidence of postoperative neurologic deficit, renal failure, atrial fibrillation or gastro-intestinal complication was similar for the two matched groups. Matched patients with recent prior PCI had longer length of stay (median of 10 days vs. 9 days ; P<0.001) and higher readmission rate at 30 days (15% vs. 10% ; P=0.021).

Survival was no different in non-matched CABG patients with recent prior PCI and in matched CABG patients with or without prior PCI (Figure 3A and 3B).

Discussion

In this large contemporary real world multi-institutional study, we present long-term results of CABG patients with prior PCI and contrast them with a matched group of patients with no prior PCI. In order to avoid bias resulting from poorer outcomes in patients requiring CABG on an urgent basis following incomplete or complicated PCI revascularization, we excluded patients requiring primary PCI for acute myocardial infarction and analyzed separately patients with recent (<14 days) or remote (≥ 14 days) prior PCI. Our matched analyses based on a propensity score revealed higher overall morbidity for patients with recent PCI, but no difference in early and long-term survival among CABG patients with and without a prior PCI.

Prior studies have addressed outcomes of CABG patients with prior successful/uneventful PCI^{181-183,186,187}. However, these studies (1) differed in their design : retrospective cohort or case-control, with different control groups : CABG patients with no prior PCI^{181-183,186,187} or PCI patients with no subsequent CABG ; (2) did not take into consideration the PCI to CABG interval, pooling all patients with a history of prior PCI together ; (3) included patients with prior PCI in the setting of acute myocardial infarction and shock where the PCI goal may have been to treat the culprit lesion expeditiously instead of providing complete revascularization ; and (3) with few exceptions, these studies were limited to short-term outcomes^{181,183}.

In a sub-study of the IMAGINE trial¹⁸¹ of 2489 patients with a median follow-up of 2.95 years, CABG patients with prior PCI had a worse outcome than those with no prior PCI (HR=1.53 [1.17–1.98]), with a higher occurrence of the

composite outcome of cardiovascular death or resuscitated cardiac arrest, nonfatal myocardial infarction (MI), coronary revascularization, unstable angina requiring hospitalization, documented angina not requiring hospitalization, stroke, or congestive heart failure requiring hospitalization. Taken individually, only coronary revascularization (HR=1.80 [1.13–2.87]) and unstable angina requiring hospitalization (HR=2.43 [1.52–3.89]) were significantly higher in patients with prior PCI, whereas cardiovascular death was not (HR=1.91 [0.88-4.13]). Similar to our study, fewer distal anastomosis were performed in CABG patients with prior PCI (3.0 ± 1.1 vs 3.3 ± 1.1 ; $P<0.001$). However, completeness of revascularization appeared to be equivalent between groups. Kalaycioglu et al.¹⁸³ compared the outcome of 40 CABG patients with prior successful PTCA to 40 case-matched CABG patients with no prior PCI and reported decreased freedom from angina (82.5% vs. 87.5% at 3 years) and survival (92.5% vs. 95.0% at 3 years) in CABG patients with prior PCI.

Using both multivariable and propensity score matching techniques in 6,032 CABG patients, Hassan et al.¹⁸⁶ showed that prior PCI was an independent predictor of in-hospital mortality after CABG (OR=1.93 [1.26-2.96]). Thielmann et al.¹⁸⁸ compared short-term outcomes of 2,626 CABG patients with no prior PCI to CABG patients with single (n=360) or multiple (n=289) prior PCI sessions. Their analyses revealed that multiple but not single prior PCI sessions were associated with increased in-hospital mortality (OR=2.24 [1.52-3.21]) and major adverse cardiac events (OR=2.28 [1.38-3.59]). They suggested that the worse postoperative outcomes may be a consequence of increased endothelial injury with multiple stenting attempts and coronary side-branch obstruction due to overlapping stents. In

a study of 749 diabetic patients,¹⁸⁷ the same group demonstrated that prior PCI was an independent predictor of in-hospital mortality (OR=2.87 [1.29-6.37]) and major adverse cardiac events (OR=2.54 [1.39-4.62]) in patients with triple vessel disease.

In patients treated between 1981 and 1997, Barakate et al.¹⁸² demonstrated similar 30-day mortality in 361 CABG patients with prior PCI (median time from initial PCI to CABG of 4 months) compared to 11,909 CABG patients with no prior PCI. They excluded patients requiring emergent CABG within 24 hours of PCI or those with unsuccessful PCI, defined by residual luminal narrowing >40% or a PCI complication. Patients with prior PCI were less likely to present with triple-vessel or left main disease and received fewer distal anastomoses (2.9 vs. 4.1 ; P<0.05). After adjusting for the number of diseased coronary vessels and left main disease, we have shown that patients with prior PCI received, on average, less grafts and the left internal thoracic artery graft was used less frequently. Lesser use of the left internal thoracic artery is of some concern and may reflect (1) absence of significant left anterior descending artery (LAD) disease, or other suitable coronary artery with large runoff, at the time of CABG, (2) successful PCI or borderline restenosis (<70%) with adequate flow in the LAD where a graft would be subject to competition of flow, or (3) the need to graft the LAD at a more distal site with a smaller luminal diameter and suboptimal runoff, or at a point where the left internal thoracic artery would not reach in-situ or as a free graft. The left internal thoracic artery is associated with better long-term patency and survival at 10 and 20-years after CABG compared to saphenous vein grafts¹⁸⁹ and this may translate to lower late survival in CABG patients with prior PCI.

What is the impact of the time interval between PCI and CABG?

The time interval between PCI and CABG is rarely reported and no prior study has reported how varying intervals may affect outcomes after CABG. In matched CABG patients with vs. without prior PCI within 14 days of their CABG, we found a significant increase in hospital morbidity, but no difference in 30-day mortality or long-term survival. Such patients are more likely to have a prior history of myocardial infarction or angina and present in cardiogenic shock or in an emergent/salvage status compared to patients with no prior PCI or PCI more than 14 days prior to CABG. Some characteristics may be a consequence of the PCI, such as early restenosis, thrombosis or coronary dissection with unstable angina. If this were the case in a sizeable proportion of patients, propensity score matching based on these characteristics would lead to the selection of a sicker control group and bias the results towards no difference between groups. Ideally, propensity score matching is best performed using patients characteristics prior to PCI, or by identifying characteristics that are a consequence of PCI, e.g. whether cardiogenic shock developed prior to or after PCI. Nevertheless, we found an increase in hospital morbidity in this higher risk subgroup compared to their propensity score matched controls, as they were more likely to receive blood product transfusions, requiring re-exploration for bleeding, present with perioperative myocardial infarction, or require prolonged ventilation. In the early period following PCI, patients may have received loading doses and/or combinations of antiplatelet and anticoagulant agents (e.g., clopidogrel, aspirin, glycoprotein IIb/IIIa inhibitors, and/or heparin) associated with increased risk of bleeding.

“Recent” and “remote” prior PCI were arbitrarily defined a priori with a cut-off at 14 days. The objective was to assess separately patients who underwent PCI and CABG during the same hospitalization or a time frame equivalent to an average hospitalization. Analyses were repeated using a cut-off at 30 days and all the conclusions remained the same, since there were fewer patients in the 14-30 day interval (n=34) compared to the 0-14 day interval (n=294).

Limitations

As a retrospective study using CABG as the mechanism for entry into analysis, this study does not directly address the question of treatment at the time of initial intervention (PCI vs. CABG). Since we did not review the total PCI population, but only the subgroup who came to surgery, we cannot account for the fact that some patients may have died in the time interval between PCI and CABG, and sicker patients with prior PCI may have declined CABG surgery. It is also possible that in some patients there was an increase in risk between the PCI and CABG (for example, reduction in LVEF due to MI).

This study lacked detailed cardiac catheterization data at the time of PCI and subsequent CABG. Furthermore, it was not possible to determine precisely whether PCI was performed in the setting of single-vessel or multiple-vessel disease and whether balloon dilatation and/or stent(s) were used. In this study, we expected to capture mostly the patients with restenosis post-PCI, as opposed to native disease

progression, by including only patients who had undergone PCI less than 5 years before CABG¹⁹⁰. This study primarily focuses on the remote prior PCI group, as the recent prior PCI group is more likely to represent a heterogeneous group with acute PCI complications. In the available databases, it was not possible to restrict analyses to patients with a single admission for PCI prior to CABG, since some prior PCI could have been performed before 1997. However, our study found no difference in early results according to the number of prior admissions for PCI.

Conclusion

In conclusion, thirty-day mortality, hospital morbidity, hospital length of stay, and survival of CABG patients with remote prior PCI are no different than a matched CABG population with no prior PCI. Patients presenting for CABG early (within 14 days) after PCI represent a sicker subgroup in whom hospital morbidity is significantly increased. Prospective follow-up of CABG patients with prior PCI is warranted, as their late survival may be affected by the decreased use of the left internal thoracic artery and coronary grafting at distal sites with poor runoff.

Acknowledgement

The authors would like to thank the MASSDAC team, and especially Dr. Sharon-Lise T. Normand, Robert E. Wolf, Ann Lovett and Katya Zelevinsky, for kindly providing the data and initial advice.

Funding Sources

Salary support (to LMS) was provided by a fellowship award from the Canadian Institutes of Health Research's Clinical Research Initiative and the Rossetti Fund (Massachusetts General Hospital, Boston, Massachusetts).

Disclosures

The authors have no conflicts to disclose.

Table 1. Patient Characteristics in CABG Patients With Remote Prior PCI (≥ 14 days) or With No Prior PCI

	<i>All CABG patients</i>			<i>CABG patients matched 1:3</i>		
	<i>Prior-PCI (n=823)</i>	<i>No-PCI (n=8819)</i>	<i>SD</i>	<i>Prior-PCI (n=809)</i>	<i>No-PCI (n=2427)</i>	<i>SD</i>
Age, y	63±11	67±11		63±11	64±10	
• < 65	455 (55%)	3324 (38%)	35.8	441 (55%)	1300 (54%)	1.9
• 65-74	229 (28%)	2939 (33%)	-12.0	229 (28%)	709 (29%)	-2.0
• > 74	139 (17%)	2556 (29%)	-29.1	139 (17%)	418 (17%)	-0.1
Sex (female), n(%)	255 (31%)	2305 (26%)	10.7	245 (30%)	704 (29%)	2.8
Race/ethnicity						
• Non-hispanic white, n(%)	757 (92%)	8037 (91%)	3.0	745 (92%)	2196 (90%)	5.7
• Non-hispanic black, n(%)	18 (2%)	172 (2%)	1.7	17 (2%)	56 (2%)	-1.4
• Hispanic, n(%)	25 (3%)	272 (3%)	-0.3	24 (3%)	79 (3%)	-1.7
• Other, n(%)	23 (3%)	338 (4%)	-5.8	23 (3%)	96 (4%)	-6.1
Diabetes, n(%)	355 (43%)	3297 (37%)	11.7	347 (43%)	1055 (43%)	-1.2
Hypertension, n(%)	679 (83%)	6981 (79%)	8.5	666 (82%)	2024 (83%)	-2.8
Dyslipidemia, n(%)	741 (90%)	6937 (79%)	31.7	727 (90%)	2210 (91%)	-4.1

	<i>All CABG patients</i>			<i>CABG patients matched 1:3</i>		
	<i>Prior-PCI (n=823)</i>	<i>No-PCI (n=8819)</i>	<i>SD</i>	<i>Prior-PCI (n=809)</i>	<i>No-PCI (n=2427)</i>	<i>SD</i>
Body surface area, m ²	2.0±0.2	2.0±0.2		2.0±0.2	2.0±0.2	
• < 1.86 m ² , n(%)	232 (28%)	2979 (34%)	-12.1	232 (29%)	674 (28%)	2.0
• 1.86-2.05 m ² , n(%)	281 (34%)	3004 (34%)	0.2	273 (34%)	835 (34%)	-1.4
• > 2.05 m ² , n(%)	310 (38%)	2836 (32%)	11.6	304 (38%)	918 (38%)	-0.5
Angina, n(%)	771 (94%)	7664 (87%)	23.0	758 (94%)	2101 (87%)	24.1
Prior myocardial infarction						
• None, n(%)	400 (49%)	4641 (53%)	-8.1	395 (49%)	1418 (58%)	-19.3
• ≤ 24 hours prior to CABG, n(%)	11 (1%)	261 (3%)	-11.2	11 (1%)	47 (2%)	-4.5
• > 24 hours prior to CABG, n(%)	412 (50%)	3917 (44%)	11.3	403 (50%)	962 (40%)	20.6
NYHA Functional Class III-IV, n(%)	575 (70%)	5906 (67%)	6.2	567 (70%)	1515 (62%)	16.3
Left ventricular ejection fraction, %	52±12	49±14		52±12	52±12	
• < 30%, n(%)	31 (4%)	749 (8%)	-19.8	31 (4%)	76 (3%)	3.8
• 30%-39%, n(%)	77 (9%)	1025 (12%)	-7.4	76 (9%)	216 (9%)	1.7
• > 39%, n(%)	683 (83%)	6664 (76%)	18.4	671 (83%)	2042 (84%)	-3.2
• Missing, n(%)	32 (4%)	381 (4%)	-2.2	31 (4%)	93 (4%)	0.0

	<i>All CABG patients</i>			<i>CABG patients matched 1:3</i>		
	<i>Prior-PCI</i> <i>(n=823)</i>	<i>No-PCI</i> <i>(n=8819)</i>	<i>SD</i>	<i>Prior-PCI</i> <i>(n=809)</i>	<i>No-PCI</i> <i>(n=2427)</i>	<i>SD</i>
Arrhythmia*, n(%)	39 (5%)	843 (10%)	-18.8	39 (5%)	99 (4%)	3.6
Peripheral vascular disease, n(%)	122 (15%)	1574 (18%)	-8.2	119 (15%)	356 (15%)	0.1
Renal failure, n(%)	59 (7%)	586 (7%)	2.1	57 (7%)	169 (7%)	0.3
Dialysis, n(%)	22 (3%)	108 (1%)	10.5	20 (2%)	60 (2%)	0.0
Chronic lung disease, n(%)	107 (13%)	1159 (13%)	-0.4	104 (13%)	300 (12%)	1.5
Current smoker, n(%)	116 (14%)	1495 (17%)	-7.9	116 (14%)	360 (15%)	-1.4
<i>Status of surgery</i>						
• Elective, n(%)	313 (38%)	2833 (32%)	12.4	306 (38%)	943 (39%)	-2.1
• Urgent, n(%)	496 (60%)	5730 (65%)	-9.7	489 (60%)	1452 (60%)	1.3
• Emergent/salvage, n(%)	14 (2%)	256 (3%)	-8.0	14 (2%)	32 (1%)	3.4
Cardiogenic shock, n(%)	6 (1%)	118 (1%)	-6.0	6 (1%)	11 (0%)	3.7
Preoperative intra-aortic balloon pump need, n(%)	53 (6%)	902 (10%)	-13.7	52 (6%)	140 (6%)	2.8
Left main stenosis (≥ 50%)	216 (26%)	3132 (36%)	-20.2	215 (27%)	675 (28%)	-2.8

	<i>All CABG patients</i>			<i>CABG patients matched 1:3</i>		
	<i>Prior-PCI (n=823)</i>	<i>No-PCI (n=8819)</i>	<i>SD</i>	<i>Prior-PCI (n=809)</i>	<i>No-PCI (n=2427)</i>	<i>SD</i>
Number of diseased vessels, n	2.6±0.6	2.8±0.5		2.6±0.6	2.6±0.6	
• One, n(%)	63 (8%)	260 (3%)	21.1	58 (7%)	148 (6%)	4.3
• Two, n(%)	221 (27%)	1534 (17%)	22.9	213 (26%)	627 (26%)	1.1
• Three, n(%)	539 (65%)	7025 (80%)	-32.2	538 (67%)	1652 (68%)	-3.3
Year of surgery						
• 2002, n(%)	293 (36%)	3108 (35%)	0.8	287 (35%)	870 (36%)	-0.8
• 2003, n(%)	308 (37%)	3008 (34%)	6.9	302 (37%)	900 (37%)	0.5
• 2004, n(%)	222 (27%)	2703 (31%)	-8.1	220 (27%)	657 (27%)	0.3
<i>Primary health insurance</i>						
• Private-HMO, n(%)	462 (56%)	4161 (47%)	18.0	451 (56%)	1363 (56%)	-0.8
• Medicare, n(%)	303 (37%)	4117 (47%)	-20.1	302 (37%)	933 (38%)	-2.3
• Welfare/Medicaid, n(%)	39 (5%)	234 (3%)	11.1	37 (5%)	93 (4%)	3.7
• Uninsured, n(%)	14 (2%)	204 (2%)	-4.4	14 (2%)	28 (1%)	4.8

* Arrhythmia : atrial fibrillation, heart block, cardiac arrest. SD : Standardized Difference.

Table 2. Operative Data and Postoperative Course in CABG Patients With Remote Prior PCI (≥ 14 days) or With No Prior PCI (matched pairs)*

	<i>Prior-PCI</i> (n=809)	<i>No-PCI</i> (n=2427)	<i>P</i>
Cardiopulmonary bypass time, min	102±37	104±36	0.088
Crossclamp time, min	71±32	75±32	0.004
Total distal coronary anastomosis, n	3.1±1.0	3.3±1.0	< .001
• Arterial grafts, n	1.2±0.6	1.3±0.6	0.099
• Vein grafts, n	1.9±1.1	2.0±1.0	0.001
Internal thoracic artery graft(s) used, n(%)	756 (93%)	2331 (96%)	0.002
• Single internal thoracic artery, n(%)	711 (88%)	2178 (90%)	0.139
• Bilateral internal thoracic artery, n(%)	45 (6%)	153 (6%)	0.431
Radial artery graft(s) used, n(%)	151 (19%)	483 (20%)	0.436
30-day Mortality, n(%)	9 (1.1%)	36 (1.5%)	0.432
Hospital Morbidity, n(%)	335 (41%)	963 (40%)	0.385
- Perioperative myocardial infarction, n(%)	21 (2.6%)	76 (3.1%)	0.444
- Reoperation for bleeding, n(%)	20 (2.5%)	39 (1.6%)	0.121
- Blood product transfusion, n(%)	456 (56%)	1255 (52%)	0.020
- Neurologic event [†] , n(%)	12 (1.5%)	38 (1.6%)	0.870

- Atrial fibrillation, n(%)	187 (23%)	571 (24%)	0.813
- Prolonged ventilation (> 24h), n(%)	83 (10%)	221 (9%)	0.322
- Pneumonia, n(%)	18 (2.2%)	61 (2.5%)	0.649
- Renal failure, n(%)	20 (2.5%)	61 (2.5%)	0.947
Total length of stay, days	10±7 (median=9)	10±7 (median=8)	0.743
Readmission within 30 days	104/769 (14%)	265/2255 (12%)	0.201

* P-values obtained from univariable generalized estimating equation models that accounted for matching. † Including stroke, transient ischemic attack, reversible ischemic neurologic deficit or coma.

Figure Legends

Figure 1 Time interval between PCI and CABG for patients with remote prior PCI (≥ 14 days).

Figure 2 Survival after CABG surgery for (A) patients with remote (≥ 14 days) prior PCI vs. patients with no prior PCI, and (B) propensity score matched patients with remote prior PCI vs. patients with no prior PCI. Yearly survival estimates and number at risk are provided in the embedded table. Groups were compared using a stratified proportional hazards model accounting (A) for clustering among hospitals or (B) for matching. We report the death hazard ratio from the stratified proportional hazards model, with 95% confidence interval and p-value, for remote prior PCI (a HR <1 indicates a lower risk for death for patients with remote prior PCI vs. those with no prior PCI).

Figure 3 Survival after CABG surgery for (A) patients with recent (<14 days) prior PCI vs. patients with no prior PCI, and (B) propensity score matched patients with recent prior PCI vs. patients with no prior PCI. See Figure 2 legend.

Figure 1.

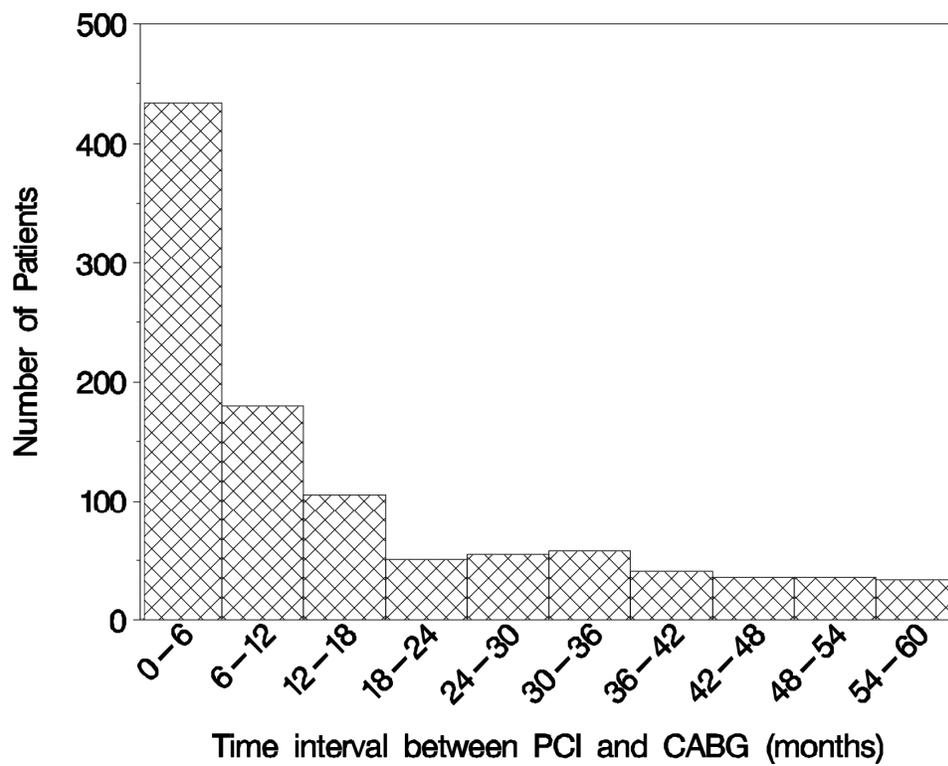


Figure 2.

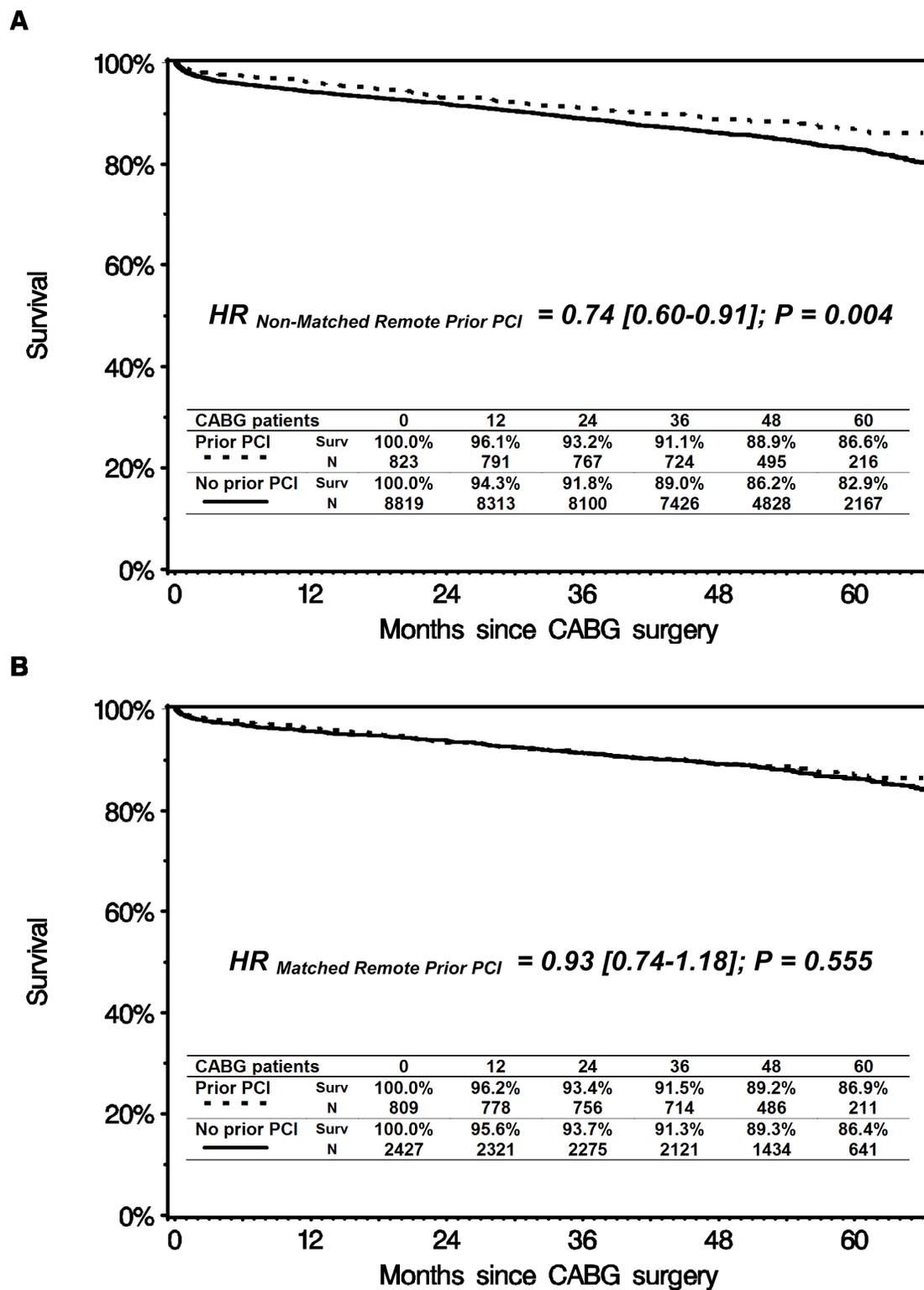
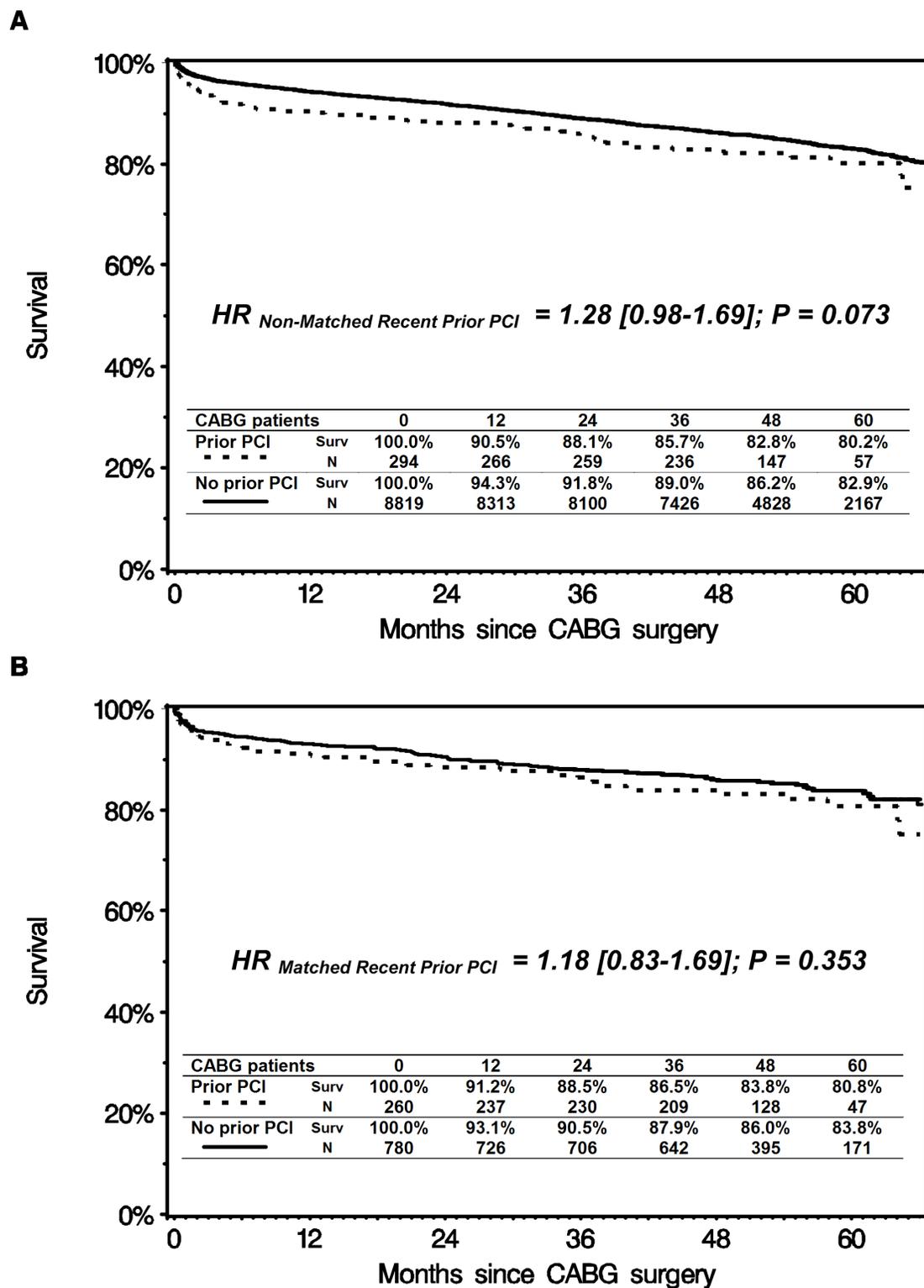


Figure 3.



- SUPPLEMENTAL MATERIAL -

*Coronary Artery Bypass Grafting After Recent or Remote Percutaneous Coronary
Intervention in the Commonwealth of Massachusetts*

Supplemental Methods

Prior to obtaining data, the following covariates were identified and later used in the propensity score : age (<65 years, 65-74 years, >74 years), female gender, body surface area (<1.86 m², 1.86-2.05 m², >2.05 m²), diabetes, hypertension, dyslipidemia, peripheral vascular disease, left ventricular ejection fraction (<30%, 30-39%, >39%), renal failure (none, with dialysis, without dialysis), chronic lung disease, current smoking, arrhythmia, status of surgery (elective, urgent, emergent/salvage), cardiogenic shock, preoperative intra-aortic balloon pump need, left main disease, number of diseased vessels, primary health insurance (private-HMO, Medicare, welfare/Medicaid, uninsured, missing), and year of CABG surgery (Table 1). Angina and prior myocardial infarction were not used in the propensity score since they could have been a consequence of the PCI.

Supplemental Table 1. Independent Predictors of Long-Term Mortality*

	<i>Hazard Ratio [95% CI]</i>	<i>P</i>
Age		< .001
• < 65 y	0.63 [0.53-0.74]	
• 65-74 y (ref.)	1	
• > 74 y	1.63 [1.45-1.84]	
Sex (female)	1.15 [1.03-1.29]	0.013
Diabetes	1.40 [1.25-1.55]	< .001
Hypertension	1.20 [1.03-1.40]	0.017
Dyslipidemia	0.80 [0.71-0.90]	< .001
NYHA Functional Class III-IV	1.31 [1.14-1.50]	< .001
Left ventricular ejection fraction		< .001
• < 30%	2.04 [1.76-2.37]	
• 30%-39%	1.34 [1.16-1.55]	
• > 39% (ref.)	1	
Arrhythmia	1.41 [1.22-1.62]	< .001
Cerebrovascular disease	1.39 [1.22-1.58]	< .001
Peripheral vascular disease	1.64 [1.46-1.84]	< .001
Renal failure		< .001
• With dialysis	5.06 [3.98-6.43]	
• Without dialysis	2.35 [2.02-2.74]	
• None (ref.)	1	
Chronic lung disease,	1.81 [1.59-2.05]	< .001
Cardiogenic shock,	1.54 [1.14-2.10]	0.005
Left main stenosis (\geq 50%)	1.13 [1.02-1.26]	0.023
Health Insurance		0.030
• Private-HMO (ref.)	1	
• Medicare	1.17 [1.03-1.33]	
• Welfare/Medicaid	1.36 [0.99-1.89]	
• Uninsured	0.70 [0.40-1.23]	

	<i>Hazard Ratio [95% CI]</i>	<i>P</i>
Internal thoracic artery graft(s) used	0.63 [0.52-0.75]	< .001
Radial artery graft(s) used	0.57 [0.46-0.71]	< .001
Remote Prior PCI (≥ 14 days)	0.93 [0.76-1.15]	0.517

*Hazard ratio and p-value obtained from stratified proportional hazard multivariable model. Missing data were handled by dummy coding for multicategory covariates (hazard ratios for dummy variables are not shown – 4% missing data for left ventricular ejection fraction and 1% missing data for primary health insurance).

Supplemental Table 2. Patient Characteristics in CABG Patients With Recent Prior PCI (<14 days) or With No Prior PCI

	<i>All CABG patients</i>			<i>CABG patients matched 1:3</i>		
	<i>Prior-PCI (n=294)</i>	<i>No-PCI (n=8819)</i>	<i>SD</i>	<i>Prior-PCI (n=260)</i>	<i>No-PCI (n=780)</i>	<i>SD</i>
Age, y	65±11	67±11		65±11	66±11	
• < 65	140 (48%)	3324 (38%)	20.2	123 (47%)	345 (44%)	6.2
• 65-74	87 (30%)	2939 (33%)	-8.0	75 (29%)	252 (32%)	-7.5
• > 74	67 (23%)	2556 (29%)	-14.2	62 (24%)	183 (23%)	0.9
Sex (female), n(%)	93 (32%)	2305 (26%)	12.1	78 (30%)	203 (26%)	8.9
Race/ethnicity						
• Non-hispanic white, n(%)	268 (91%)	8037 (91%)	0.1	236 (91%)	712 (91%)	-1.8
• Non-hispanic black, n(%)	4 (1%)	172 (2%)	-4.6	4 (2%)	14 (2%)	-2.0
• Hispanic, n(%)	12 (4%)	272 (3%)	5.4	12 (5%)	29 (4%)	4.5
• Other, n(%)	10 (3%)	338 (4%)	-2.3	8 (3%)	25 (3%)	-0.7
Diabetes, n(%)	103 (35%)	3297 (37%)	-4.9	97 (37%)	256 (33%)	9.4
Hypertension, n(%)	236 (80%)	6981 (79%)	2.8	212 (82%)	625 (80%)	3.6
Dyslipidemia, n(%)	243 (83%)	6937 (79%)	10.1	213 (82%)	653 (84%)	-4.8

	<i>All CABG patients</i>			<i>CABG patients matched 1:3</i>		
	<i>Prior-PCI (n=294)</i>	<i>No-PCI (n=8819)</i>	<i>SD</i>	<i>Prior-PCI (n=260)</i>	<i>No-PCI (n=780)</i>	<i>SD</i>
Body surface area, m ²	1.9±0.2	2.0±0.2		1.9±0.2	1.9±0.2	
• < 1.86 m ² , n(%)	111 (38%)	2979 (34%)	8.3	95 (37%)	280 (36%)	1.3
• 1.86-2.05 m ² , n(%)	96 (33%)	3004 (34%)	-3.0	84 (32%)	277 (36%)	-6.8
• > 2.05 m ² , n(%)	87 (29%)	2836 (32%)	-5.6	81 (31%)	223 (29%)	5.6
Angina, n(%)	269 (91%)	7664 (87%)	14.8	238 (92%)	698 (89%)	7.0
Prior myocardial infarction						
• None, n(%)	106 (36%)	4641 (53%)	-33.8	96 (37%)	384 (49%)	-25.0
• ≤ 24 hours prior to CABG, n(%)	35 (12%)	261 (3%)	34.6	23 (9%)	63 (8%)	2.8
• > 24 hours prior to CABG, n(%)	153 (52%)	3917 (44%)	15.3	141 (54%)	333 (43%)	23.2
NYHA Functional Class III-IV, n(%)	208 (71%)	5906 (67%)	8.2	180 (69%)	555 (71%)	-4.2
Left ventricular ejection fraction, %	49±13	49±14		49±13	50±13	
• < 30%, n(%)	19 (6%)	749 (8%)	-7.7	18 (7%)	48 (6%)	3.1
• 30%-39%, n(%)	32 (11%)	1025 (12%)	-2.3	26 (10%)	77 (10%)	0.4
• > 39%, n(%)	271 (74%)	6664 (76%)	-4.0	200 (77%)	602 (77%)	-0.6
• Missing, n(%)	26 (9%)	381 (4%)	18.3	16 (6%)	53 (7%)	-2.6

	<i>All CABG patients</i>			<i>CABG patients matched 1:3</i>		
	<i>Prior-PCI (n=294)</i>	<i>No-PCI (n=8819)</i>	<i>SD</i>	<i>Prior-PCI (n=260)</i>	<i>No-PCI (n=780)</i>	<i>SD</i>
Arrhythmia*, n(%)	28 (10%)	843 (10%)	-0.1	26 (10%)	69 (9%)	4.0
Peripheral vascular disease, n(%)	54 (18%)	1574 (18%)	1.3	48 (18%)	115 (15%)	10.0
Renal failure, n(%)	18 (6%)	586 (7%)	-2.1	17 (7%)	43 (6%)	4.3
Dialysis, n(%)	8 (3%)	108 (1%)	10.8	7 (3%)	18 (2%)	2.57
Chronic lung disease, n(%)	38 (13%)	1159 (13%)	-0.6	34 (13%)	96 (12%)	2.3
Current smoker, n(%)	61 (21%)	1495 (17%)	9.7	53 (20%)	149 (19%)	3.2
<i>Status of surgery</i>						
• Elective, n(%)	48 (16%)	2833 (32%)	-37.5	48 (18%)	141 (18%)	1.0
• Urgent, n(%)	183 (62%)	5730 (65%)	-5.7	182 (70%)	543 (70%)	0.8
• Emergent/salvage, n(%)	63 (21%)	256 (3%)	59.1	30 (12%)	96 (12%)	-2.4
Cardiogenic shock, n(%)	30 (10%)	118 (1%)	38.7	14 (5%)	42 (5%)	0
Preoperative intra-aortic balloon pump need, n(%)	97 (33%)	902 (10%)	57.6	71 (27%)	202 (26%)	3.2
Left main stenosis (≥ 50%)	71 (24%)	3132 (36%)	-25.0	69 (27%)	227 (29%)	-5.7

	<i>All CABG patients</i>			<i>CABG patients matched 1:3</i>		
	<i>Prior-PCI (n=294)</i>	<i>No-PCI (n=8819)</i>	<i>SD</i>	<i>Prior-PCI (n=260)</i>	<i>No-PCI (n=780)</i>	<i>SD</i>
Number of diseased vessels, n	2.5±0.7	2.8±0.5		2.6±0.6	2.5±0.7	
• One, n(%)	30 (10%)	260 (3%)	29.6	17 (7%)	73 (9%)	-10.4
• Two, n(%)	88 (30%)	1534 (17%)	29.8	75 (29%)	227 (29%)	-0.6
• Three, n(%)	176 (60%)	7025 (80%)	-44.1	168 (65%)	480 (62%)	6.4
Year of surgery						
• 2002, n(%)	88 (30%)	3108 (35%)	-11.3	73 (28%)	232 (30%)	-3.7
• 2003, n(%)	100 (34%)	3008 (34%)	-0.2	90 (35%)	270 (35%)	0
• 2004, n(%)	106 (36%)	2703 (31%)	11.5	97 (37%)	278 (36%)	3.5
<i>Primary health insurance</i>						
• Private-HMO, n(%)	158 (54%)	4161 (47%)	13.1	139 (53%)	406 (52%)	2.8
• Medicare, n(%)	126 (43%)	4117 (47%)	-7.7	112 (43%)	343 (44%)	-1.8
• Welfare/Medicaid, n(%)	4 (1%)	234 (3%)	-9.2	4 (2%)	7 (1%)	5.8
• Uninsured, n(%)	4 (1%)	204 (2%)	-7.1	4 (2%)	13 (2%)	-1.0

* Arrhythmia : atrial fibrillation, heart block, cardiac arrest. SD : Standardized Difference.

CHAPITRE III

FINANCIAL CONSEQUENCES OF IMPLEMENTING A PARTNER IN CARE IN CARDIAC SURGERY

ABSTRACT

Background : In 2003, a satellite cardiac surgery program (SAT) was implemented at an affiliated community hospital located in an area historically served by an academic medical center (AMC). The aim of this study was to assess the financial consequences and the changes in case-mix that occurred at the AMC subsequent to SAT implementation.

Methods : From June 2002 through December 2005, 4,593 adult patients underwent cardiac surgery at the AMC. We excluded 400 patients operated during the 4-month transition period following SAT implementation as well as 1,210 patients living >35 miles from the AMC. Changes in case-mix and propensity-score adjusted costs were compared for AMC patients referred from SAT area ($N_{\text{before/after}}=328/291$) vs. other patients ($N_{\text{before/after}}=897/1,467$) using multivariable regression.

Results : The referral rate from SAT area decreased by 55%. Compared to other patients, AMC patients referred from SAT area showed a greater increase in age in the second period ($P=0.013$). The nursing workload and adjusted mean costs increased more for patients from SAT area ($P=0.015$ and 0.014 , respectively). The hospital margin decreased in the second period for both referral areas ($P<0.001$). For the CABG patient subgroup, this hospital margin decrease was greater for SAT area patients ($P=0.017$).

Conclusions : After implementation of SAT program, fewer patients of lower complexity came to the AMC from the SAT area and there was a significant increase in nursing workload and costs. During this interval, hospital margin for cardiac surgery decreased from both referral areas but decreased significantly more for CABG patients from the SAT area.

The evidence that percutaneous coronary interventions (PCI) improves survival for patients with acute myocardial infarction has led to increased demand for access to this technology, particularly in community hospitals¹⁹¹. Given the difficulty of maintaining financial viability as well as a skilled team with acute myocardial infarction volume alone, community PCI programs have a strong incentive to include elective PCI procedures as well^{191,192}. In most states, regulations require on-site cardiac surgery as a prerequisite for elective PCI, with the circuitous result that new community cardiac surgery programs (SAT) have been formed to support an elective PCI program that exists to enable a PCI program for acute myocardial infarction^{99,193}. While cardiac services remain an important and profitable service line for hospitals, they are associated with significant start-up and fixed costs. Cost constraints amplified by the extension of health insurance through national health reform will undoubtedly put pressure on the competing goals of consolidation and regionalization to save cost vs. geographic distribution to make life saving technology more accessible¹⁹⁴.

In Massachusetts, where there is a state determination of need process, legislation generated by a coalition of community political leaders, community hospitals and interventional cardiologists resulted in the approval in 2002 of three new SAT programs with mandatory affiliation to an existing academic medical center (AMC). One of these programs was implemented in an area historically served by our AMC at an institution that had a long standing relationship with the AMC. In a collaborative planning process, both hospitals agreed upon a list of high risk patient characteristics for transfer to the AMC (Supplemental Methods 1),

although relatively few patients were transferred for these reasons during the study period (N=7). In general, transfer of SAT area patients to the AMC was not explicitly managed but driven by patient/provider preferences on a case by case basis. The aim of this study was to assess changes in volume, case-mix and financial margin at the AMC before and after implementation of SAT.

Patients and Methods

Patients

Approval for this study was obtained from the Institutional Review Board in September 2006. From June 2002 through December 2005, 4,593 adult patients underwent cardiac surgery at the AMC, excluding heart transplantation and assist devices. We decided to exclude patients operated at the AMC during the initial 4 months following SAT implementation in March 2003 (N=400) on the assumption that this transition interval would have an intermediate mixture of patient characteristics and potentially obscure comparisons. The SAT hospital is located 17.8 miles from the AMC. The *SAT hospital service area* was identified using ZIP codes to define its primary and secondary service area, including Essex and eastern part of Middlesex County in Massachusetts. Patients were considered *referred from SAT area* if they were initially treated by a SAT area provider or were residing in the SAT hospital service area. Otherwise, patients were considered to be *referred from other areas*. We observed that the population of patients referred from a greater

distance was more heterogeneous with referral of unusual cases or need for particular expertise, and decided, *a priori*, to restrict the comparator patient population to those living within approximately the same distance from the AMC as patients referred from the SAT area. Patients living >35 miles from the AMC according to their home ZIP code address (N=1,210) were excluded, resulting in a final cohort of 2,983 patients. Changes in case-mix and costs per encounter following SAT implementation on March 2003 were compared for AMC patients from SAT area (N_{before/after}=328/291) vs. all other AMC patients (N_{before/after}=897/1,467). Also, the clinical characteristics of patients treated at the SAT hospital were compared to patients referred from the SAT area following implementation.

Baseline patient demographics, comorbidities, operative details and postoperative outcomes were prospectively collected in an institutional database of cardiac surgery procedures based on the STS database¹⁹⁵. *Major morbidity and mortality* was a combined outcome measure including any of the following postoperative variables: operative mortality (30-day), reoperation for any cause, permanent stroke, prolonged ventilation support (>24h), deep sternal wound infection or renal failure. *Health insurance status* was defined as the primary health insurance coverage. *Patient acuity* is hourly measure of intensity of care as assessed by a computerized nursing questionnaire linked to the TSI cost accounting system (Transition System Incorporated, Boston, Inc., Boston, MA). *Total nursing workload* is obtained by summing up the time spent in each acuity level multiplied by their respective weights. *Average (daily) acuity level* is obtained by dividing total nursing workload by total length of stay.

Costs

Costs were measured using the institutional general ledger and activity-based cost-accounting system TSI. Patient costs were recorded daily and summed to obtain the *total cost* for each hospitalization. Physician fees were not included in cost figures. Costs were classified into three general categories. *Variable direct (patient-related) costs* measured expenditures directly related to patient care on a particular day, such as laboratory tests, radiographs, and disposable supplies. *Fixed direct (department-related) costs* were defined as expenditures associated with departments that provide direct patient care but not to a particular patient, such as equipment and medical devices. *Indirect (institution-related) costs* referred to expenditures outside individual departments, such as administration and other support services departments. Taken together, department- and institution-related costs were collectively referred to as “hospital overhead”.

Expected payments per encounter, adjusted for denials, free care, and/or settlements if relevant, were recorded at the end of hospitalization and are referred to as *net revenue*. Net revenue was based on previously agreed reimbursement from each insurer, reconciled at the end of each fiscal year to ensure accuracy. All costs and revenues were converted to January 2002 US\$ using the *Medical Consumer Price Index* for Boston, Brockton and Nashua (MA, NH, ME, CT) from the U.S. Bureau of Labor Statistics. *Hospital margin* was defined as net revenue minus total costs. Only costs and revenue from the index admission were considered. The all-hospital 30-day readmission rate was constant through the time period of the

study (between 11% and 13%). Cost analyses were performed only at the AMC. Cost analysis using data generated by TSI is only accurate when comparing patients in the same institution and does not allow comparison of different institutions' financial performance.

Statistical methods

Data are expressed as number (percentage), mean, standard deviation or 95% confidence interval. Multivariable regression models (i.e., logistic, linear or ordinal) included a main effect for time period and referral area to provide two-way ANOVA statistics. Similar multivariable regression models included an interaction term between time period and referral area to assess whether changes experienced by SAT area patients and patients from other referral area were different. Identity-link normal-distribution linear models were used to assess log-transformed length of stay, nursing workload, costs and revenue. These analyses were restricted to cases with values within 3 SD of the mean because of skewed data. AMC patients referred from SAT area were compared to patients treated at SAT hospital using χ^2 and Student's t-tests.

In order to adjust for differences in case-mix between referral areas in cost modeling, we generated a propensity score derived from a nonparsimonious logistic model using referral from SAT area as the outcome variable. Twenty categorical variables and interaction terms were included in the propensity score (Supplemental

Methods 2). Separate propensity score was fitted for isolated CABG subgroup. The propensity score was considered as a continuous variable in all cost and revenue models after verification of the linearity assumption. Propensity-score adjusted mean cost and revenue for each cohort in original units were obtained by exponential transformation of the predicted value and application of Duan smearing estimate to adjust for the bias resulting from retransformation of log-transformed data¹⁹⁶. All statistical analyses were performed using SAS release 9.1 (SAS Institute Inc., Cary, NC, USA) and p-values <0.05 were considered statistically significant.

Results

Patient characteristics, operative data and postoperative course for the four cohorts are shown in Tables 1-3. Two general observations were apparent. First, the patients referred from SAT area were more complex than patients referred from other areas throughout the study period. Secondly, the overall patient population became more complex over time. Thirty-day mortality and major morbidity remained unchanged throughout the study period and was not different between referral areas. For both time periods, patients referred from SAT area required more transfusions and experienced longer hospital length of stay. Globally, there was an increase in Medicare-insured patients and a corresponding decrease in private/HMO-insured patients after SAT implementation, with no difference between referral areas.

Differences between referral areas changes

The increase in patient age, peripheral vascular disease, arrhythmia and use of immunosuppressant in the second period was greater for patients referred from SAT area following SAT implementation compared to other patients. The increase in preoperative length of stay and nursing workload in the second period was more pronounced for SAT area patients.

SAT patients treated at the SAT hospital were directly compared to AMC patients referred from SAT area during the second period (Tables 1-2, 5th column and last 2 columns). For patients where a STS predicted mortality risk was obtainable,

the median STS predicted mortality was lower for SAT patients treated at SAT (1.9% interquartile range [1.0%, 4.2%] vs. 3.9% [1.4%, 6.9%]; $P<.001$). Correspondingly, observed mortality was lower for SAT patients treated at SAT (4/289 (1.0%) vs. 10/213 (3.4%); $P=0.029$).

Surgical Volume

Time trends for cardiac surgical volumes of AMC hospital (referred from SAT area or others) and SAT hospital are illustrated in Figure 1. AMC hospital cardiac surgery volume from SAT area decreased by 55% (264 to 120 patients/year) compared to a decrease of 15% for other AMC referral areas (721 to 611 patients/year). Looking at isolated CABG only, AMC referral volume from SAT area decreased by 61% (187 to 73 patients/year) compared to a decrease of 26% for other AMC referral areas (412 to 305 patients/year).

Costs, Revenue and Hospital Margin

Propensity-score adjusted cost data are reported in Table 4. For both referral areas, cost increased more than revenue in the second period leading to a decrease in hospital margin (Figure 2). For both time periods, revenue and hospital margin were lower for SAT area patients. In the second period, adjusted mean costs increased more for SAT area (\$7,436 or +26%) compared to other patients (\$4,476 or +15% ;

$P=0.014$). The hospital margin decreased by \$5,334 for the SAT area vs. \$2,947 for other patients but the differences did not reach statistical significance ($P=0.110$).

Cost and revenue were lower for patients who underwent isolated CABG compared to patients who underwent other types of cardiac surgery. Cost increased more than revenue in the second period leading to a significant hospital margin decrease in subgroup analyses. In this more homogenous group of patients, the hospital margin decreased more for the SAT area compared to other referral areas (\$5,552 vs. \$1,160 ; $P=0.017$).

Comment

This article evaluates the impact of implementing a satellite cardiac surgery program on the local referral pattern, costs and revenue for an AMC. To our knowledge, this is the first study to assess this common issue.

Change in Referral Volume and Case-Mix

In the last decade, adult cardiac surgery case volume declined nationally while the number of surgical centers paradoxically increased in conjunction with the increase in community PCI programs⁹⁹. The decrease in volume was driven by a reduction in isolated CABG surgery, which overshadowed smaller increases in valve surgery, arrhythmia surgery and heart failure surgery^{197,198}. The overall decreased referral for isolated CABG surgery is partly explained by declining admissions for acute coronary symptoms from improved preventive (e.g., smoking cessation) and medical (e.g., ACE inhibitors and statins) care⁷⁷, but also by the growth of PCI¹⁰⁰. In the current study, AMC referral for isolated CABG declined between 2002 and 2003, stabilized afterwards for patients referred from other areas, but further decreased for those referred from SAT area (Figure 1). Between 1995 and 2003, the opening of specialty cardiac surgery hospitals was associated with increasing population-based rates of coronary revascularization (CABG and/or PCI) in Medicare beneficiaries compared to hospital referral regions where new cardiac programs opened at general hospital or no new program opened¹⁹⁹. Correspondingly, it is probable that this SAT

implementation at a general hospital led to a simple redistribution of cases within the health care market shared by the AMC.

Cardiac surgery mortality has continued to decline in last decade despite greater patient complexity. According to the STS database, the population of patients undergoing cardiac surgery is becoming older, with an increasing incidence of prior catheter-based interventions¹⁰¹, and CABG patients are increasingly more likely to be of female gender, have a history of smoking, diabetes mellitus, renal failure, hypertension, stroke, chronic lung disease, NYHA functional class IV, and three-vessel disease¹⁰². Although most of these trends were also apparent in our study population, they were more pronounced for patients referred from the SAT area after SAT implementation. Total nursing workload was increased for SAT area patients after SAT implementation reflecting a sicker patient population with a more complicated perioperative course. In spite of an overall increase in the case-mix severity, we observed no increase in 30-day mortality, major morbidity or daily acuity level for both referral areas in the period after SAT implementation. This is different than the findings of Omoigui et al.²⁰⁰ where the results of patients referred from New York State to the Cleveland Clinic deteriorated after public dissemination of surgeon-specific outcome data. In many regards, public reporting and the implementation of a start-up program under public scrutiny have similar referral incentives.

Financial Consequence

The AMC margin for SAT area patients declined after SAT implementation. This observation was partially confounded by the fact that the population studied was heterogeneous in term of surgical procedures and that the margin for cardiac surgery as a whole declined during this time interval. The magnitude of reduction was greater for patients from the SAT region, with a non-significant difference for the two time intervals for the non-SAT region. The difference in reduction in the AMC margin for the second period was significantly greater in the CABG group, a more homogenous patient sub-population.

These findings were not surprising. A cardiac surgery satellite program should be expected to impact the residual regional case-mix referred to the AMC unfavorably. This is not a prescription for AMCs to shun satellite programs but more of a balanced message : the relationship carries reciprocal positives and negatives. Given the considerable fixed costs in cardiac surgery, residual SAT area volume that does not fully cover total costs may nonetheless make a positive “contribution margin” at the AMC by adding revenue in excess of direct costs and contributing to institutional overhead. Moreover, in spite of the generally lower costs of the community hospital setting, SAT cardiac surgery programs are costly to establish and maintain because of higher personnel ratios and other fixed program costs relative to patient volume so that SAT programs probably could not survive without this tacit subsidy. Highly complex patients that place a heavy burden on ancillary hospital services like the blood bank and subspecialty services are more costly to care for but more appropriately matched to AMC’s resources.

There is not much prior literature that looks at the relationships of SAT programs and AMC's or the impact of SAT programs on regional healthcare delivery, and the existing literature is mostly focused on volume. The 1997 American College of Surgeons Guidelines for Standards in Cardiac Surgery suggested at least 200 CABG procedures per program per year to function efficiently based on studies showing a decrease in costs with higher volume programs²⁰¹. A study in Massachusetts, performed before SAT implantation, showed no volume-cost relationship for programs above 270 CABG cases per year²⁰¹; whether this holds true for lower volume programs is unclear (Figure 2). National CABG volumes have fallen markedly since both of these studies. Insofar as the volume/quality relationship is concerned, states with certificate-of-need regulations have higher hospital CABG procedural volumes which have been associated with decreased mortality rates in some reports²⁰². More recent data showed similar outcomes suggesting that certificate of need regulation and higher volume alone do not result in improved quality of care²⁰³. In Massachusetts, all AMC and SAT hospitals participate in a rigorous collaborative effort to monitor and improve clinical quality and there have been no concerns of quality outliers among any of the SAT hospitals since their inception.

The larger health policy question remains as to whether PCI for acute myocardial infarction remains an appropriate rationale to drive SAT surgery program development. There is a growing appreciation that PCI can be performed safely in hospitals without on-site cardiac surgery, and the practice is becoming more commonplace nationally¹⁹¹. Long et al. have shown similar clinical outcomes for

low-risk elective PCI at a community hospital without on-site cardiac surgery compared to matched patients treated at a tertiary care hospital, albeit at increased direct medical cost¹⁹². Based on the current state of the art of therapy for acute myocardial infarction, PCI is both necessary and appropriate in a community setting. With a nationwide reduction in coronary surgery volume, a threatened profound decline in the cardiac surgical workforce and the intensity of cost pressures in healthcare, allowing independent PCI without on-site cardiac surgery with a resultant trend towards greater regionalization of cardiac surgery is probably inevitable.

Limitations

In this study, we assumed that the impact of increased PCI utilization was similar for patients referred from SAT area or other areas; concomitant implementation of the SAT PCI program alone could have had an impact on cardiac surgery patients' case-mix and referral to SAT hospital.

In order to capture the change in case-mix for the overall population of patients referred from the SAT area, all adult cardiac surgeries were included except for heart transplantation, assistance device or adult congenital cardiac surgery, which are not available at the SAT hospital. The study population remains heterogeneous by nature and multivariable modeling and a propensity score may be limited in its ability to fully adjust for case-mix differences. To address this issue, subgroup analyses were performed for isolated CABG patients. Other homogeneous

subgroups were too small to generate meaningful analyses. During the study period, there was no major change in AMC financial policies and insurer reimbursement policies. It is possible, however, that minor changes had an impact on our inferences. Looking forward, recent approval by the Center for Medicare and Medicaid Services of a cardiothoracic surgical reimbursement codes valuation process using STS data instead of the former Relative Value Update Committee's survey estimation process may lead to more accurate compensation for high-risk patients which could change the impact that we observed¹⁰¹.

Conclusions

Implementation of a satellite cardiac surgery program was associated with a decrease in regional AMC referral volume and an increase in case-mix severity, length of stay, nursing workload, and total costs per encounter for the remaining cases. The AMC margin decreased in the second period for all patients and decreased significantly more for CABG patients referred from the satellite hospital referral area. While targeted referral of complex patients or procedures to an AMC may be desirable from a clinical and academic perspective, it is evident that establishing a SAT program results in compromised financial performance at the AMC.

Acknowledgements

We are grateful for the research assistance provided by Karen E. Lynch and the financial expertise from Benjamin S. Parson and Robert Seger. Salary support (to LMS) was provided by a fellowship award from the Canadian Institutes of Health Research's Clinical Research Initiative and the Rossetti Fund (Massachusetts General Hospital, Boston, Massachusetts).

Table 1. Characteristics of AMC patients from SAT or other referral areas, before (01/2002-03/2003) or after (08/2003-12/2005) SAT implementation, and characteristics of SAT patients treated at SAT hospital (08/2003-12/2005) ^a

	AMC patients				<i>Two-way ANOVA</i>		<i>Interaction</i>	SAT patients	
	Other referral areas		SAT referral area		<i>Time Period</i>	<i>Referral Area</i>	<i>Time Period*</i>	Treated at SAT	
	Before	After	Before	After ^c	<i>P</i> ^b	<i>P</i> ^b	<i>Referral Area</i>	<i>P</i> ^b	After ^c
	(n=897)	(n=1467)	(n=328)	(n=291)					(n=402)
Age (years)	67±12	66±13	69±11	71±12	0.931	<.001	0.013	69±11	0.007
Gender (Male)	633 (71%)	1024 (70%)	223 (68%)	188 (65%)	0.441	0.063	0.554	291 (72%)	0.031
Diabetes	256 (29%)	423 (29%)	106 (32%)	108 (37%)	0.459	0.004	0.307	133 (33%)	0.294
Hypertension	637 (71%)	1115 (76%)	234 (71%)	236 (81%)	<.001	0.172	0.183	285 (71%)	0.002
Prior MI	253 (28%)	470 (32%)	131 (40%)	143 (49%)	0.005	<.001	0.307	184 (46%)	0.397
CHF	202 (23%)	365 (25%)	94 (29%)	109 (37%)	0.022	<.001	0.176	92 (23%)	<.001
Renal failure	75 (8%)	165 (11%)	28 (9%)	32 (11%)	0.014	0.977	0.877	23 (6%)	0.015
PVD	130 (14%)	226 (15%)	47 (14%)	66 (23%)	0.065	0.030	0.043	50 (12%)	<.001
Arrhythmia	148 (16%)	305 (21%)	53 (16%)	84 (29%)	<.001	0.030	0.042	62 (15%)	<.001

	AMC patients				Two-way ANOVA		Interaction	SAT patients	
	Other referral areas		SAT referral area		Time Period	Referral Area	Time Period*	Treated at SAT	
	Before	After	Before	After ^c			Referral Area	After ^c	
	(n=897)	(n=1467)	(n=328)	(n=291)	P ^b	P ^b	P ^b	(n=402)	P ^c
Any valvulopathy	486 (54%)	914 (62%)	153 (47%)	165 (57%)	<.001	0.003	0.706	214 (53%)	0.395
COPD	68 (8%)	178 (12%)	32 (10%)	43 (15%)	<.001	0.082	0.867	32 (8%)	0.006
Immunosuppressant	54 (6%)	83 (6%)	15 (5%)	31 (11%)	0.251	0.100	0.009	10 (2%)	<.001

^a Plus-minus values are means \pm SD. ^b *p* values were calculated using multivariable logistic, linear, log-normal or ordinal regression models where appropriate. Two-way ANOVA refers to the global effect of time period (before vs. after) or referral area (SAT referral area vs. other referral areas). Interaction refers to the divergence in time period effect for the two referral area groups. ^c *p* values using Student's *t*-test or χ^2 comparing patients treated at SAT vs. patients treated at AMC but referred from SAT area after SAT implementation. AMC : academic medical center ; SAT : satellite hospital.

Table 2. Operative data of AMC patients from SAT or other referral areas, before (01/2002-03/2003) or after (08/2003-12/2005) SAT implementation, and operative data of SAT patients treated at SAT hospital (08/2003-12/2005) ^a

	AMC patients				Two-way ANOVA		Interaction	SAT patients	
	Other referral areas		SAT referral area		Time	Referral	Time Period*	Treated at SAT	
	Before	After	Before	After ^c	Period	Area	Referral Area	After ^c	P ^c
	(n=897)	(n=1467)	(n=328)	(n=291)	P ^b	P ^b	P ^b	(n=402)	
Prior sternotomy	75 (8%)	152 (10%)	22 (7%)	33 (11%)	0.019	0.802	0.294	27 (7%)	0.053
Non-elective	549 (61%)	845 (58%)	275 (84%)	254 (87%)	0.234	<.001	0.082	149 (37%)	<.001
Left main disease (if CABG)	131/627 (19%)	310/926 (30%)	72/277 (25%)	90/240 (38%)	<.001	0.004	0.909	76/311 (24%)	0.001
# of diseased vessels (if CABG)	2.4±0.8	2.6±0.7	2.4±0.8	2.6±0.7	<.001	0.226	0.777	2.7±0.6	0.103
Type of surgery					<.001	<.001	0.521		0.191

	AMC patients				<i>Two-way ANOVA</i>		<i>Interaction</i>	SAT patients	
	Other referral areas		SAT referral area		<i>Time</i>	<i>Referral</i>	<i>Time Period*</i>	Treated at SAT	
	Before	After	Before	After ^c	<i>Period</i>	<i>Area</i>	<i>Referral Area</i>	After ^c	<i>P^c</i>
	(n=897)	(n=1467)	(n=328)	(n=291)	<i>P^b</i>	<i>P^b</i>	<i>P^b</i>	(n=402)	
- Isolated CABG	513 (57%)	745 (51%)	232 (71%)	181 (62%)				220 (55%)	
- Isolated valve	136 (15%)	200 (14%)	20 (6%)	26 (9%)				50 (12%)	
- CABG-valve	114 (13%)	181 (12%)	45 (14%)	38 (13%)				56 (14%)	
- Other	134 (15%)	341 (23%)	31 (9%)	46 (16%)				76 (19%)	

See Table 1 legend.

Table 3. Postoperative course and insurance status of AMC patients from SAT or other referral areas, before (01/2002-03/2003) or after (08/2003-12/2005) SAT implementation ^a

	AMC patients				Two-way ANOVA		Interaction
	Other referral areas		SAT referral area		Time Period	Referral Area	Time Period*
	Before	After	Before	After			Referral Area
	(n=897)	(n=1467)	(n=328)	(n=291)	<i>P</i> ^b	<i>P</i> ^b	<i>P</i> ^b
<i>Postoperative course</i>							
30-day mortality	24 (2.7%)	44 (3.0%)	7 (2.1%)	10 (3.4%)	0.391	0.947	0.508
Major morbidity	206 (23%)	292 (20%)	84 (26%)	71 (24%)	0.086	0.059	0.581
Transfusion	492 (55%)	917 (63%)	197 (60%)	203 (70%)	<.001	0.005	0.560
Preop LOS (d) ^d	1 (1-2)	2 (1-2)	2 (2-3)	3 (3-4)	0.012	<.001	0.013
ICU LOS (h)	67±128	74±137	75±140	90±154	0.077	0.061	0.510
Postop LOS (d) ^d	7 (7-7)	7 (7-8)	7 (7-8)	8 (8-9)	0.002	0.022	0.094
Acuity level, average	2.2±0.4	2.1±0.4	2.2±0.4	2.1±0.4	0.235	0.570	0.973
Nursing workload ^d	21 (20-22)	22 (21-23)	22 (21-24)	27 (25-29)	<.001	<.001	0.015

	AMC patients				Two-way ANOVA		Interaction
	Other referral areas		SAT referral area		Time Period	Referral Area	Time Period*
	Before	After	Before	After			Referral Area
	(n=897)	(n=1467)	(n=328)	(n=291)	<i>P</i> ^b	<i>P</i> ^b	<i>P</i> ^b
Insurance status					<.001	0.220	0.424
Private	269 (30%)	378 (26%)	118 (36%)	76 (26%)			
HMO	169 (19%)	242 (17%)	55 (17%)	38 (13%)			
Medicare	402 (45%)	751 (51%)	137 (42%)	164 (57%)			
Medicaid	41 (5%)	70 (5%)	14 (4%)	8 (3%)			
No insurance	14 (2%)	25 (2%)	4 (1%)	4 (1%)			

See Table 1 legend. ^d: Median and 95% confidence interval are reported in original units (days, nursing workload) after exponential transformation.

Table 4. Adjusted mean costs, revenue, hospital margin (in thousand\$) for AMC patients from SAT or other referral areas, before (01/2002-03/2003) or after (08/2003-12/2005) SAT implementation. ^a

	AMC Patients				Two-way ANOVA		Interaction
	SAT Area		Other Areas		Time	Referral	Time Period*
	Before	After	Before	After	Period	Area	Referral Area
	(n=328)	(n=291)	(n=897)	(n=1467)	<i>P</i> ^b	<i>P</i> ^b	<i>P</i> ^b
<i>All patients</i>							
Total cost, K\$ ^e	28.1 (26.7-29.5)	35.6 (33.8-37.4)	30.5 (29.6-31.4)	35.0 (34.1-35.8)	<.001	0.145	0.014
Net revenue, K\$ ^e	32.7 (31.2-34.2)	35.9 (34.2-37.5)	37.0 (36.0-38.0)	39.4 (38.6-40.3)	<.001	<.001	0.426
Hospital margin, K\$	5.27±1.00	-0.06±0.99	7.17±0.56	4.22±0.43	<.001	<.001	0.110
Negative hospital margin	32%±3%	47%±3%	28%±1%	36%±1%	<.001	0.001	0.135
<i>Isolated CABG</i>	(n=232)	(n=180)	(n=512)	(n=744)	<i>P</i> ^b	<i>P</i> ^b	<i>P</i> ^b
Total cost, K\$ ^e	24.9 (23.6-26.1)	30.8 (29.1-32.5)	27.5 (26.6-28.4)	31.1 (30.2-32.0)	<.001	0.014	0.033
Net revenue, K\$ ^e	28.9 (27.5-30.3)	31.1 (29.5-32.7)	32.2 (31.3-33.2)	34.6 (33.6-35.5)	<.001	<.001	0.931
Hospital margin, K\$	5.05±1.03	-0.50±1.07	5.23±0.64	3.57±0.53	<.001	0.018	0.017
Negative hospital margin	32%±3%	46%±4%	31%±2%	37%±2%	0.003	0.125	0.124

^a Plus-minus values are means \pm standard error. ^b *p* values were calculated using multivariable logistic, linear or log-normal regression models including propensity score. Two-way ANOVA refers to the global effect of time period (before vs. after) or referral area (SAT referral area vs. other referral areas). Interaction refers to the divergence in time period effect for the two referral area groups. ^c Means and 95% confidence interval, adjusted for the propensity score, are reported in original units (thousand\$) after exponential transformation and application of Duan smearing estimate ; medians were obtained by exponential transformation only.¹⁹⁶
AMC : academic medical center ; SAT : satellite hospital ; Hospital margin : net revenue minus total costs ; Negative hospital margin : percent of patients with a negative hospital margin (costs>revenue).

Figure Legends

Figure 1 Annualized cardiac surgery rate non-parametrically smoothed with a one-year moving window for all cardiac surgeries (A) or isolated CABG surgeries (B). The 4-month transition period following SAT implementation is illustrated with a shaded area. AMC : academic medical center ; SAT : satellite hospital.

Figure 2 Median contribution margin expressed as revenue minus variable costs (patient-related) for AMC patients referred or not from the SAT area, before and after SAT implementation. Median overhead costs are also shown, including fixed direct (department-related) and indirect (institution-related) hospitalization costs. Estimates adjusted for the propensity score and reported in original units (thousand\$) after exponential transformation and application of Duan smearing estimate.

Figure 1.

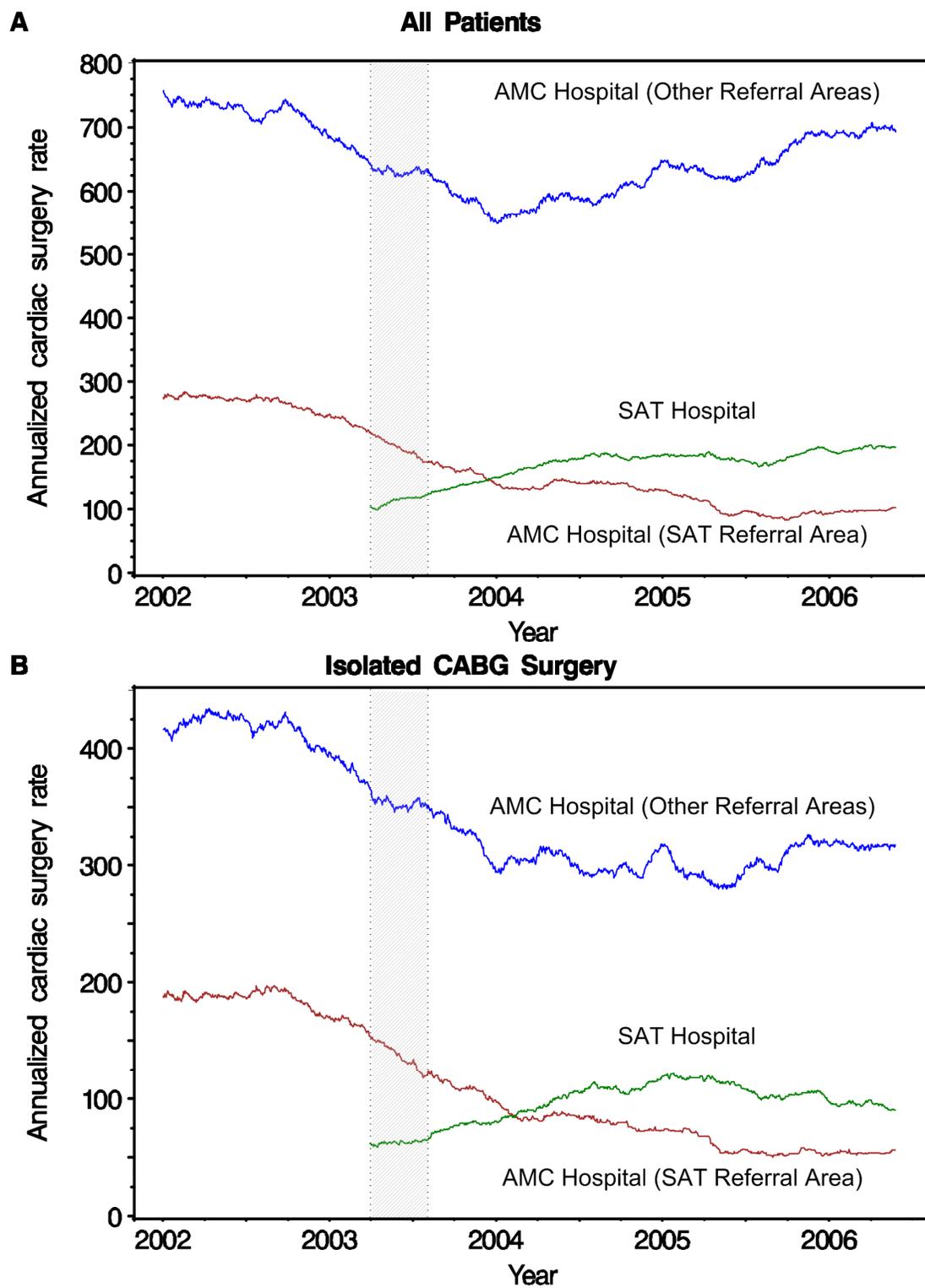
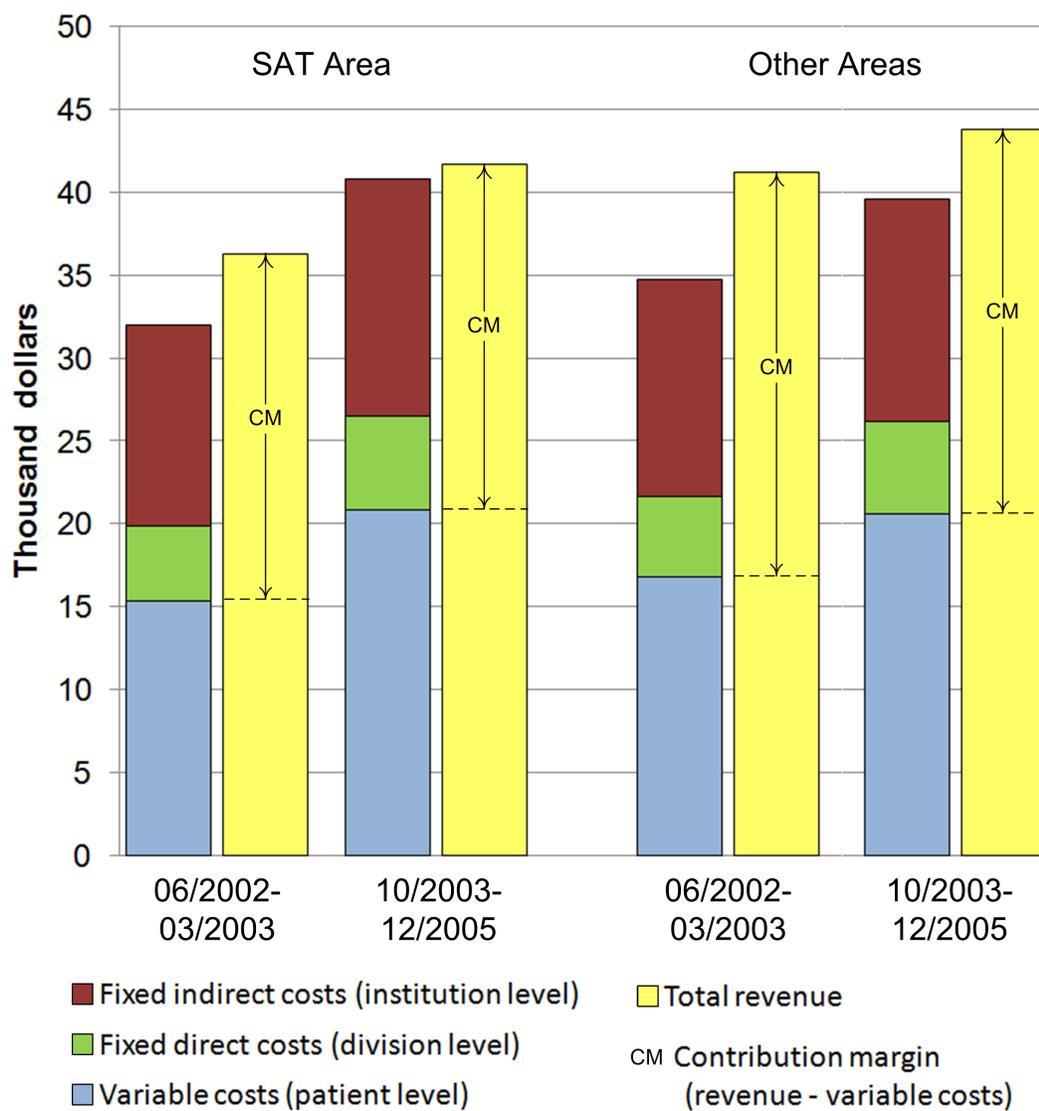


Figure 2.



- SUPPLEMENTAL MATERIAL -

Financial Consequences of Implementing a Partner in Care in Cardiac Surgery

Supplemental Methods 1. High-risk conditions leading to a transfer from the SAT hospital to the AMC.

1. Respiratory insufficiencies as defined by FEV1 <1.5 L
2. Generalized arteriosclerosis
 - Combined carotid and coronary artery disease
 - Coronary disease and lower extremity arteriosclerosis
 - Severe arteriosclerosis of aorta
3. Ventricular Assistant Devices (VAD)
4. Right heart failure
5. Morbid obesity (>300 lbs.)
6. Complex Medical issues, e.g. :
 - Long term treatment with high doses of steroids
 - Long term treatment with amiodarone
 - Coagulopathies, including heparin induced thrombocytopenia

Supplemental Methods 2. Variables included in the propensity score :

- Age : <65 years, 65-75 years, >75 years ;
- Gender ;
- Interaction between age and gender ;
- Diabetes ;
- Angina : none, stable, unstable ;
- Prior myocardial infarction : none, ≤ 1 week prior to surgery, >1 week ;
- NYHA functional class III or IV ;
- Renal disease : none, renal failure without dialysis, with dialysis ;
- Cerebrovascular disease ;
- Peripheral vascular disease ;
- Arrhythmia ;
- Infective endocarditis ;
- Smoking ;
- Status of the surgery : elective, urgent, emergent ;
- Left main disease ;
- Number of coronary vessels involved : none, 1 or 2, 3 ;
- Type of surgery performed : isolated CABG, isolated valve, CABG-valve, other ;
- Health insurance : private, HMA, Medicare, Medicaid-welfare, uninsured ;
- Distance from the AMC : 0-7.5 miles, 7.6-16.9 miles, 17-35 miles ;
- Year of surgery.

CHAPITRE IV

DISCUSSION DES RÉSULTATS

BIostatISTIQUES APPLIQUÉES EN CHIRURGIE CARDIAQUE

Tel que décrit en introduction ainsi que dans la section « hypothèses et but » :
« L'objectif principal de ce travail est d'étudier en profondeur certaines techniques biostatistiques avancées en recherche évaluative en chirurgie cardiaque adulte ».

Cette section sur la discussion des résultats suit la même logique et se veut une extension des sections de méthodologie statistique résumées dans les articles avec mes commentaires et réflexions. Je vais donc discuter des aspects méthodologiques des techniques biostatistiques exposées en introduction et du choix des types d'analyse en relation avec les données analysées, avec des exemples d'analyses tirés de ces articles. En tant que clinicien d'abord et avant tout, j'estime que l'aspect médical des trouvailles cliniques de ces travaux et leur impact demeurent primordial. Ces conclusions sont soulignées point par point avant de décrire la méthodologie utilisée et ses alternatives pour bien démontrer la motivation clinique derrière ces analyses.

DISSECTION AIGÛE DE L'AORTE ASCENDANTE

Les facteurs de risque indépendant de mortalité hospitalière étaient l'âge des patients ($P < 0,001$), la valeur de la créatinine sérique ($P = 0,03$), la syncope ($P = 0,003$) et l'année de la chirurgie ($P = 0,002$).

Pour l'analyse des facteurs prédictifs de mortalité à court terme, j'ai utilisé des modèles de régression logistique univariés ; soit un modèle pour chaque variable prise individuellement. L'intérêt de cette approche est qu'elle permet ensuite la régression multivariée et l'identification des facteurs de risque indépendant. Pour ces modèles, j'ai analysé la plupart des données préopératoires sauf si celles-ci n'avaient pas d'intérêt clinique ou étaient présente chez un trop faible pourcentage de patients (<5%). Le but ultime du modèle de régression logistique multivarié est d'indiquer quels sont les facteurs de risque généraux de mortalité et non d'identifier un petit sous-groupe de patients sur lequel nos données sont limitées voire insuffisantes.

J'ai sélectionné les variables ayant une valeur de P de moins de 0,2 pour le modèle multivarié. La valeur limite de P à moins de 0,2 est communément choisie dans les analyses univariées pour établir une première sélection de variables candidates pour le modèle multivarié et de s'assurer que les variables confondantes importantes n'ont pas été mises de côté (voir ci-dessous pour une discussion des variables confondantes en statistique). Je conserve parfois des variables ayant une valeur de P de plus de 0,2 si cette variable a déjà démontré un effet dans la littérature ou s'il s'agit d'une variable d'intérêt principale, auquel cas cette variable est « forcée », i.e. conservée, dans le modèle jusqu'à la toute fin.

Plusieurs variables étaient fortement associés à l'année de chirurgie (*experience*) et ont été, par nécessité, exclues de l'analyse multivariée. La variable *experience* tient compte de l'expérience acquise au cours des ans et consiste à l'intervalle de temps entre le début de l'étude (1979) et la chirurgie. Les variables exclues se rapportent à des modalités, traitements ou techniques qui sont devenu disponible durant la période d'étude, et dont l'effet n'est pas estimable pour toute la durée de l'étude. Celles-ci incluent des modalités d'imagerie diagnostique plus sophistiquées comme la tomodensitométrie (corrélation Pearson $r = 0,56$) et l'échocardiographie transoesophagienne peropératoire ($r = 0,57$), les techniques d'anastomose distale ouverte ($r = 0,73$) et de perfusion cérébrale antérograde ($r = 0,41$), l'utilisation d'antifibrinolytiques ($r = 0,64$) et de la colle biologique ($r = 0,51$) (tous $P < 0,001$).

Pour mes modèles multivariés, j'évalue la colinéarité entre chaque facteurs de risque pris deux par deux à l'aide de tests de corrélation puis j'effectue un choix parmi les variables présentant une forte colinéarité (i.e. fortement corrélées entre-elles, redondantes, qui expliquent le même concept). Par exemple, *tamponade* et *choc_cardio* étaient colinéaires car le patient en tamponnade cardiaque se présente souvent en choc cardiogénique. Pour l'analyse, j'ai opté pour la variable *tamponade* car elle ne présente pas de données manquantes et elle est plus facilement définie. La variable *experience* est un autre exemple (voir encadré)

J'ai analysé attentivement les variables présentant un effet contraire au bon sens (« hypothèse biologique »), mettant en doute leur validité, parfois en allant jusqu'à revoir tous les dossiers où cette variable est notée en s'assurant que la bonne définition de variable a été utilisée ou que le codage dans la base de donnée est adéquat. Par exemple, je me suis rendu compte que la variable *tamponade* avait été mal rapportée pour certains dossiers. En tant que clinicien, j'ai su dépister ces erreurs de codage et ainsi amélioré le modèle.

Pour les modèles d'analyse des facteurs de risque de mortalité à court terme, la transformation logarithmique a été utilisée pour les variables créatinine sérique, temps de clampage aortique et année de la chirurgie (*experience*). Pour les facteurs de risque de mortalité dans la phase tardive de suivi, la transformation « au carré » a été employée pour la variable *experience*.

J'ai évalué les variables continues dans le but de déterminer si elles prédisent l'évènement (décès) de façon linéaire (supposition de linéarité). Pour ce faire, j'ai :

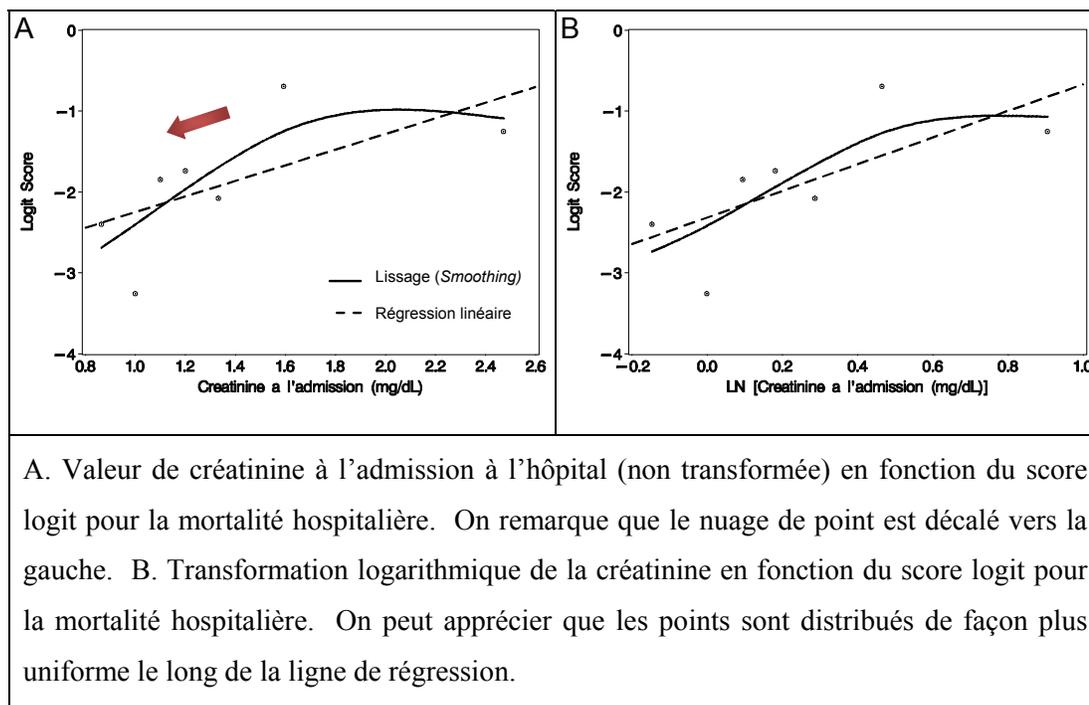
- 1) trié les patients en ordre ascendant de la valeur de la variable continue et divisé le groupe en cinq à huit sous-groupes ;
- 2) calculé la probabilité « *prob* » de décès dans chaque sous-groupe (nombre de décès divisé par nombre de personnes dans le sous-groupe) et le score logit

$$(\textit{logit score}) : \ln \left[\frac{\textit{prob}}{(1 - \textit{prob})} \right] \text{ (ce score logit peut être calculé seulement pour}$$

les sous-groupes où *prob* est plus grand que 0, i.e. au moins un évènement) ;

- 3) effectué quelques transformations pour la variable continue (logarithme naturel, inverse (1/x), exposant, puissance carrée) ;
- 4) créé des graphiques « nuage de point » (*scatter plot*) pour le score logit et la variable continue (ou une de ses transformations) ;
- 5) ajouté aux graphiques une ligne de régression linéaire et une ligne obtenue à l'aide d'un algorithme de lissage (*smoothing*), dans le but de m'aider à apprécier la linéarité (Figure 1). L'algorithme de lissage est un outil graphique du logiciel statistique qui permet de constater à quel point cette ligne courbée représentant le reste des points se rapproche de la ligne droite de régression obtenue par le technique des moindres carrés (voir Chapitre I).

Figure 1. Transformation de variables continues et score logit.



Pour les variables binaires présentant des données manquantes, j'ai utilisé la technique du "dummy coding" en assignant un troisième niveau pour ces variables (soit « 0 » si absent ; « 1 » si présent ; « 2 » si donnée manquante). Ceci permet de conserver les patients qui ont des données manquantes et évite que le logiciel statistique ne les élimine d'emblée. L'utilisation de cette méthode suppose que les patients avec données manquantes sont représentatifs des patients sans données manquantes (« non informatif »). Je n'utilise cette méthode que si le pourcentage de données manquantes pour cette variable est bas (<5%) et qu'il ne s'agit pas d'une variable d'intérêt principale (ex. : si le but du travail est de comparer les diabétiques et non-diabétiques, les patients avec données manquantes pour le diabète sont éliminés). Le coefficient pour le niveau « 2 » n'est pas en soit pertinent, mais permet d'estimer directement l'effet de la variable d'intérêt (niveau « 1 ») par rapport aux

patients qui ne présentent pas cette variable (niveau « 0 ») sans que les patients avec donnée manquante ne soient considérés comme des « 1 » ou des « 0 ». Par contre, lorsque le coefficient pour le niveau « 2 » est statistiquement significatif, je m'interroge si ces patients avec données manquantes sont véritablement représentatifs des autres patients. Par exemple, il est possible que les patients qui se présentent en urgence extrême, où le risque de mortalité est plus grand, aient plus de chances de présenter des données manquantes (car le temps a manqué pour recueillir les données préopératoires) ; ces patients instables ne sont pas représentatifs des autres patients. Lorsque le nombre de données manquantes est plus grand, que je crois que les patients avec données manquantes sont représentatifs des autres patients et que la variable demeure importante, je prends l'ensemble des patients sans données manquante, je bâti un modèle pour prédire la valeur de la variable d'intérêt à l'aide des autres variables connues et j'utilise ce modèle prédictif pour attribuer une valeur à la variable d'intérêt lorsqu'elle est manquante (en anglais "*multiple imputation*", i.e. substitution ou redressement multiple).

J'ai ensuite regroupé les variables sélectionnées dans le modèle univarié dans un modèle multivarié. La capacité du modèle multivarié à identifier les différences de résultats entre les sous-groupes de patients définis par ces variables est davantage influencée par le nombre d'évènements (ici le nombre de décès hospitaliers) que la taille de l'échantillon analysé. Règle générale, le ratio entre les évènements et les facteurs de risque dans le modèle multivarié doit être de 10 pour 1 ^{9,204,205}. Par exemple, s'il y a 30 décès rapportés dans ce groupe de patients, le modèle de régression logistique multivarié pourra compter jusqu'à trois variables. Il est parfois

possible d'inclure davantage de variables (jusqu'à un ratio de 5:1), surtout s'il s'agit de variables continues ou catégorielles à plus de deux niveaux. Je remarque par contre que les modèles tendent à devenir instable à ces extrêmes, ce que je note par une faible amélioration du degré de précision du modèle (C-statistique^m) et des coefficients imprécis pour les variables ajoutées (i.e. larges intervalles de confiance des rapports de cotes).

J'ai alors éliminé une à une les variables du modèle multivarié en commençant par celles où la valeur de P était la plus élevée (*backward stepwise selection*). Après l'élimination de chacune de ces variables, j'ai noté l'impact que cette élimination a eu sur les coefficients des autres variables dans le but de dépister la présence de variables confondantes (*confounding*). Une variable confondante est une variable qui est associée/corrélée à la fois avec la variable dépendante (résultat) et une des autres variables indépendantes. Il est important de bien tenir compte des variables confondantes pour éviter une erreur de type Iⁿ au risque de nuire à la validité interne de l'étude et de la possibilité d'établir une relation de cause à effet. Par exemple, si on assume que le poids d'un enfant et le produit national brut (PNB) augmentent dans le temps, un chercheur pourrait en venir à la conclusion que l'augmentation du poids des enfants est associé avec l'augmentation du PNB ou

^m C-statistique est une mesure de la concordance, ou accord, entre les résultats prédits et observés. Une valeur de 0.5 est équivalente à la prédiction d'un événement basé sur le hasard (pile ou face). Une valeur de plus de 0.5 indique un modèle qui prédit mieux que le hasard et une valeur de 1.0 une capacité de prédiction de résultat parfaite. Pour le modèle de régression logistique, la c-statistique est équivalente à l'aire sous la courbe ROC (*receiver operator characteristic*) décrite par Hanley et McNeil²⁰⁶.

ⁿ L'erreur de type I consiste à rejeter l'hypothèse nulle alors que celle-ci est vraie (faux positif). La probabilité d'erreur de Type I (alpha; α) est habituellement fixée à 0,05, i.e. que l'on accepte le risque que la différence observée soit due au hasard 5% du temps. L'erreur de type II consiste à ne pas rejeter l'hypothèse nulle alors que celle-ci est fautive (faux négatif). La probabilité d'erreur de type II (béta; β) est reliée au concept de puissance ($1 - \beta$).

inversement que l'augmentation du PNB est associée avec une augmentation du poids des enfants. Par contre, cet effet disparaît lorsque la variable confondante « temps » est incluse dans le modèle. Lorsqu'une variable confondante est identifiée, il est important de la conserver dans le modèle avec les autres variables qu'elle influence. Je n'ai pas encore rencontré d'exemple concret de variable confondante dans les modèles statistiques auxquels j'ai été exposé en chirurgie cardiaque.

J'ai poursuivi ce processus d'élimination de variable jusqu'à ce que la valeur de P pour les variables restant dans le modèle soit de moins de 0,05. J'ai alors reconsidéré chaque variable éliminée pour sa ré-inclusion dans le modèle (d'où le terme « *stepwise* »). Voici les lignes de code dans SAS^o pour le modèle final :

```
proc logistic data=temp ;
model h_death (event='1') = exper5 age10 CrAdmln p_syncop
/lackfit influence iplots outroc=roc parmlabel ;
run ;
```

Où la variable `h_death` désigne la présence de décès à 30 jours ou au cours de l'hospitalisation si celle-ci est de plus de 30 jours ; `exper5` est une variable continue consistant en la variable `experience` divisée par 5 (ce qui nous donne l'effet par tranche de cinq ans) ; `age10` consiste en l'âge du patient divisé par dix ; `CrAdmln` est la transformation logarithmique naturel (\ln) de la valeur de créatinine à l'admission (voir étude de linéarité plus haut) ; `p_syncop` est une variable binaire désignant la présence de syncope en préopératoire.

^o Toutes les lignes de code utilisées pour les analyses présentées dans cette thèse sont disponibles sur demande auprès de Dr. LM Stevens. Ces programmes ne sont pas reproduits ici pour éviter d'alourdir la thèse. L'intérêt ici est plutôt de présenter la rationnelle clinique derrière ces analyses.

Le syndrome de Marfan a été initialement sélectionné comme un prédicteur indépendant de mortalité hospitalière (rapport de cote 24,61 [2,08 à 291,1], $P = 0,01$), mais la variable a finalement été retirée de la liste des facteurs de risque, après avoir constaté que la stabilité de cette estimation était en question. Avec la procédure de “*bootstrapping*”, la variable a été rejetée avec une valeur P de 0,25 ; toutes les autres estimés de valeur de P sont demeurés significatifs ($P < 0,001$). Le modèle final avait eu une C-statistique de 0,85 (très bon) et une valeur de P de 0,6 pour le test de validité de l’ajustement (*Goodness-of-Fit*) de Hosmer et Lemeshow.

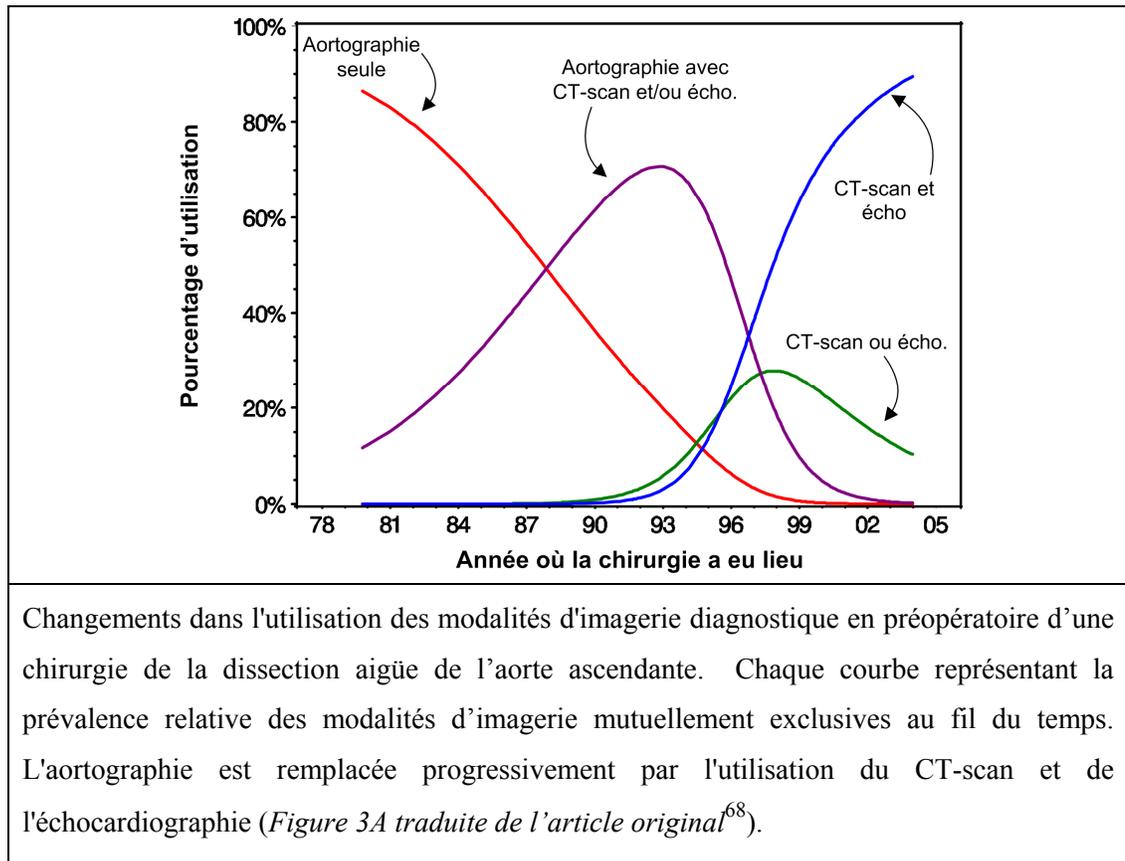
La variable Marfan avait un coefficient avec un très large intervalle de confiance dans le modèle multivarié (voir aussi le Chapitre I, section sur la régression logistique, où l’impact de la variable Marfan sur le modèle est discuté). J’ai procédé à une technique de “*bootstrapping*” pour évaluer la stabilité des estimés des coefficients des variables du modèle final en définissant 5000 échantillons de 195 patients choisis de façon aléatoire avec remplacement à partir du groupe d’origine suivant la méthode décrite dans le Chapitre I^p. La variable Marfan n’a pas conservé son effet lors de ces analyses. Ceci est probablement dû au fait que peu de patients présentaient la caractéristique Marfan. J’ai donc décidé de l’éliminer du modèle car je ne disposais pas de suffisamment d’information pour en justifier l’inclusion dans le modèle final. Le test de validité de l’ajustement (*Goodness-of-Fit*) de Hosmer et Lemeshow compare les valeurs observées à celles prédites par le modèle ; une valeur de P petite ($< 0,05$) nous indique que le modèle est mal spécifié (prédictions inadéquates).

^p Les ordinateurs sont devenus tellement performants qu’il devient maintenant envisageable de définir 5000 échantillons et de répéter l’analyse autant de fois sans que cela prenne trop de temps ; j’ai remarqué qu’avec 1000 échantillons, les résultats pouvaient légèrement varier d’une fois à l’autre et j’ai donc décidé d’augmenter le nombre d’échantillon dans le but de diminuer ces différences.

Les modalités d'imagerie diagnostique se sont améliorées au cours des deux dernières décennies avec une transition de procédures invasives nécessitant des quantités de contraste élevées (aortographie) vers les modalités avec peu de contraste (tomodensitométrie) et/ou accessibles au chevet (échocardiographie). Avec l'amélioration des technologies d'imagerie, l'aortographie était rarement demandée comme une modalité unique et a été progressivement remplacée par la tomodensitométrie, l'échographie préopératoire ou les deux modalités combinés.

Afin d'évaluer la prévalence des stratégies de diagnostic ou de technique chirurgicale à travers le temps, j'ai défini des catégories mutuellement exclusives, comme « aortographie seule », « aortographie et CT-scan », « aortographie et échocardiographie », « CT-scan et échocardiographie » « CT-scan seul », etc. Un modèle logistique multinomial a été élaboré en utilisant ces catégories comme variable dépendante (résultat) et la date de la chirurgie, ou expérience, comme variable indépendante, produisant ainsi des probabilités prédites dans le temps, chacune représentant la probabilité d'être affecté à l'une ou l'autre des catégories (avec un total de 100% à chaque instant). J'ai combiné les catégories avec trop peu de patients et des tendances similaires si cela faisait du sens et puis re-calculé le modèle (Figure 2). Dans le but de vérifier si ces « courbes de prédiction » tirées du modèle logistique multinomial représentaient bien les données, j'ai divisé la population de l'étude en cinq à huit groupes consécutifs dans le temps et obtenu des pourcentages d'utilisation pour chacune des catégories que j'ai ensuite ajouté au graphique vis-à-vis de la valeur moyenne de date de chirurgie pour chaque sous-groupe de patients. Les valeurs prédites suivaient de près les valeurs obtenues (non illustrées ici pour ne pas surcharger la figure).

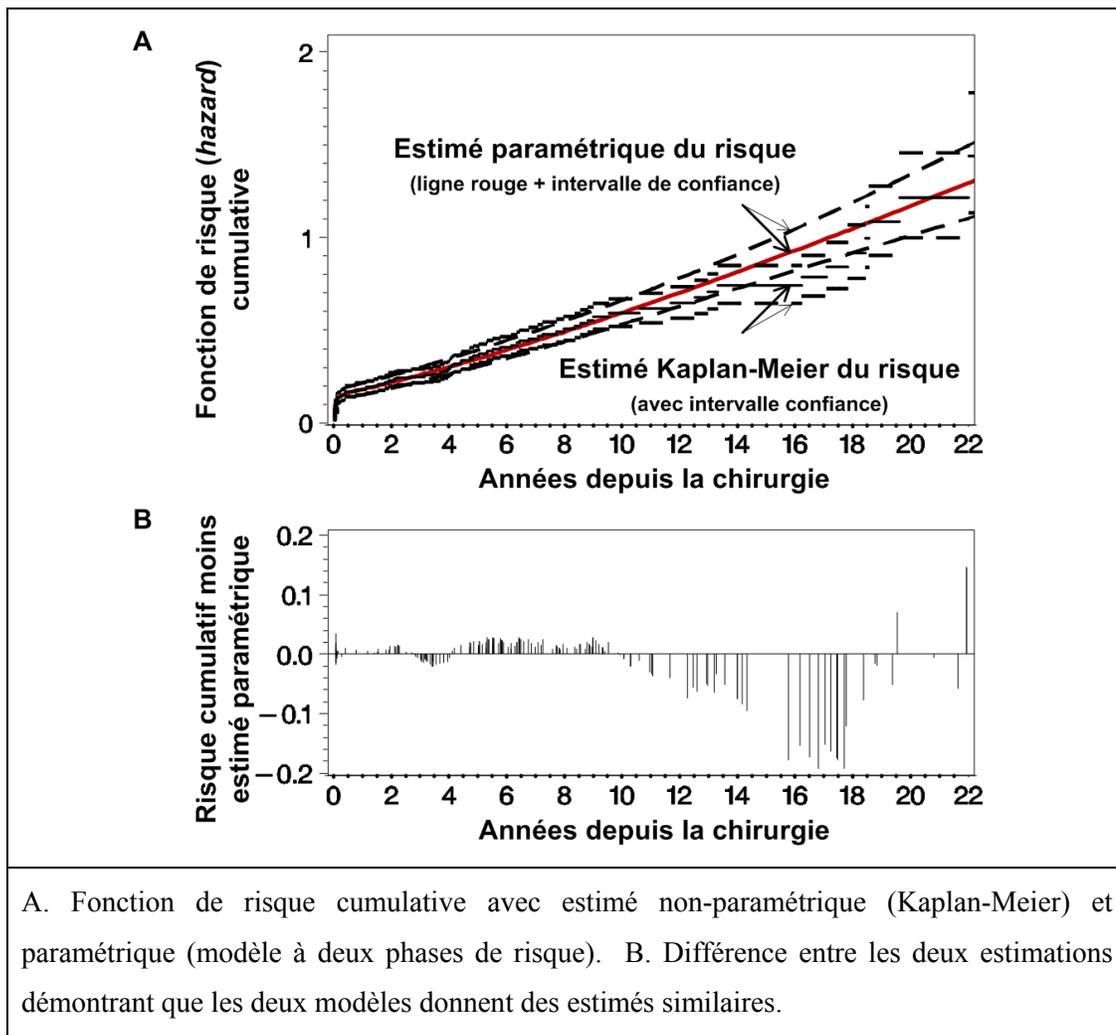
Figure 2. Modèle multinomial pour les modalités d'imagerie diagnostique⁶⁸.



Le suivi moyen était de $7,0 \pm 5,9$ ans (maximum 27 ans), soit $8,1 \pm 5,6$ ans pour les patients vivants à la fin de leur hospitalisation initiale (suivi complet à 100%). La survie était de 84%, 69%, 55% et 30% à 1, 5, 10 et 20 ans, respectivement.

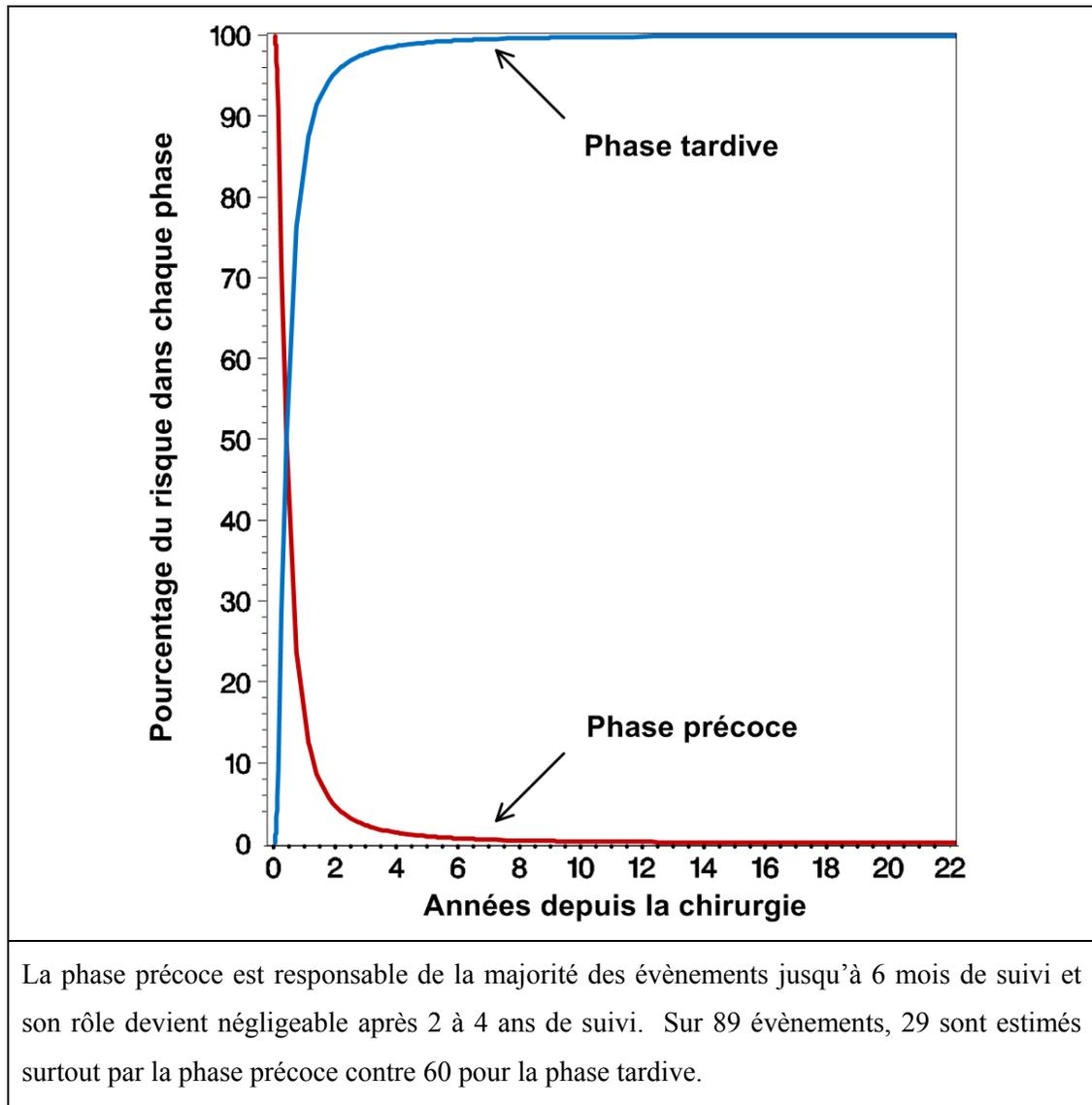
Étant donné le très long suivi disponible pour ces patients, j'ai décidé d'explorer des modèles plus complexes d'analyse de survie avec décomposition dans le temps du niveau de risque (*hazard*) en phases (voir Chapitre I). Les estimés de Kaplan-Meier m'ont guidés pour estimer une courbe de survie paramétrique avec deux phases de risque : précoce et tardive (Figure 3).

Figure 3. Fonction de risque cumulative.



Malgré un total de 89 décès au suivi pour ces 195 patients, le nombre d'évènements n'était pas suffisant pour me permettre d'estimer une troisième phase intermédiaire (appelée phase de risque constant) – cette phase est majoritairement estimée ici dans la phase tardive. Environ 29 évènements sont estimés par la phase précoce alors que les autres le sont par la phase tardive (Figure 4).

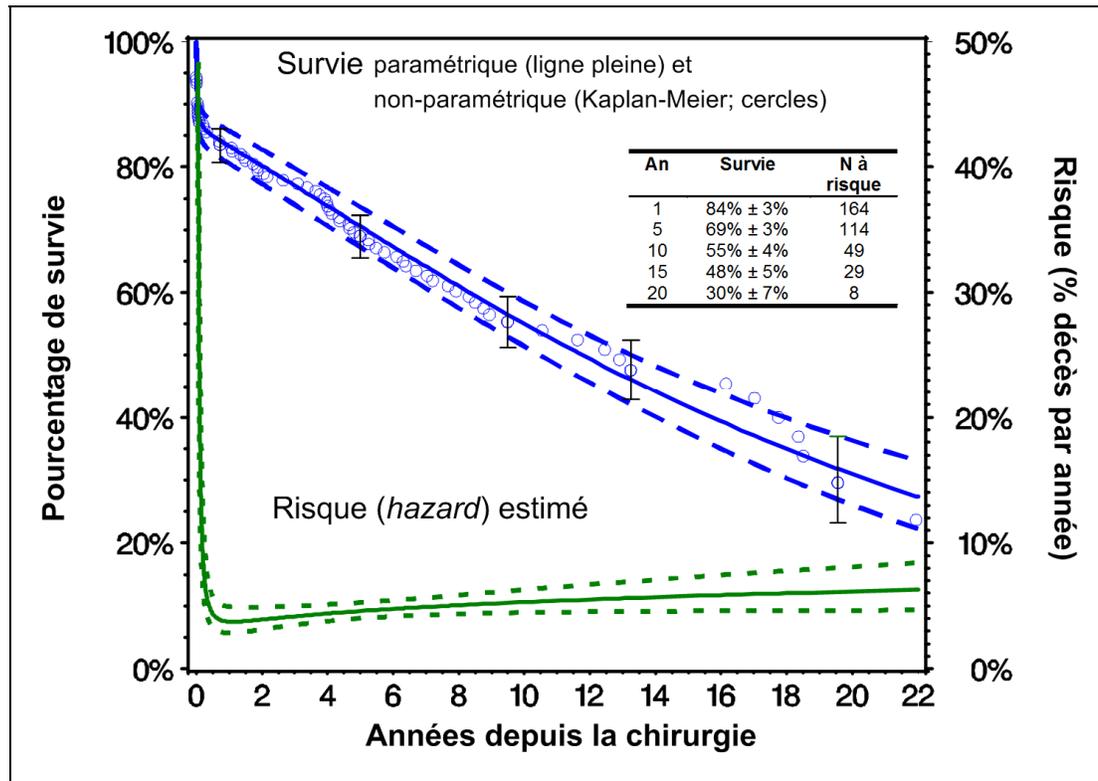
Figure 4. Importance relative des phases précoce et tardive au suivi.



Après être parvenu à estimer tous les coefficients définissant la courbe de survie paramétrique en suivant la méthodologie de décomposition du niveau de risque en phase (<http://www.clevelandclinic.org/heartcenter/hazard/default.htm>)¹¹⁹⁻¹²², j'ai été en mesure de produire une courbe de survie paramétrique (Figure 5) et d'estimer les coefficients des facteurs de risque pour la phase précoce et tardive.

La figure de survie paramétrique et de risque de décès montre un risque élevé de décès au début suivi par un risque faible mais croissant dans la phase tardive.

Figure 5. Survie et niveau de risque instantané de décès après la chirurgie.



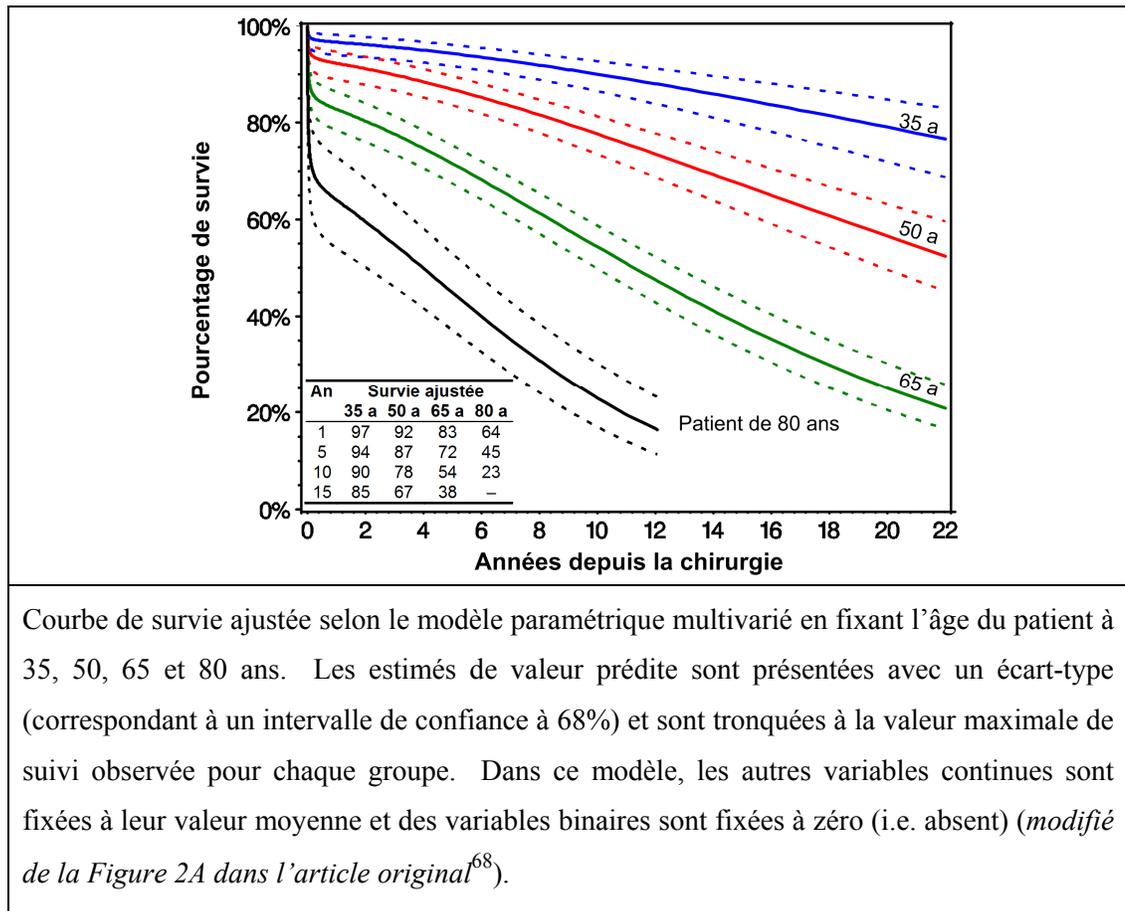
La survie non-paramétrique (actuarielle, Kaplan-Meier) est représentée avec des cercles et des intervalles de confiance à 68% à 1, 5, 10, 15 et 20 ans (correspondant à un écart-type). La survie prédite à partir du modèle paramétrique à deux phases est indiquée avec une ligne pleine bleue plus ou moins un écart-type (ligne en pointillé). La courbe de niveau de risque instantané avec un écart type est donnée en vert et se rapporte à l'axe vertical droit. Le tableau résume les estimés de survie actuarielle à 1, 5, 10, 15 et 20 ans, incluant le nombre de patient à risque (*modifié de la Figure 1A dans l'article original*⁶⁸).

Pour la phase précoce, les facteurs de risque indépendants de mortalité sont l'âge ($P = 0,002$), la créatinine sérique ($P = 0,003$), la syncope ($P = 0,007$), la chirurgie en extrême urgence ($P = 0,02$), le temps de clampage aortique ($P = 0,02$) et l'année de chirurgie ($P < 0,001$) alors que pour la phase tardive il s'agit de l'âge ($P < 0,001$), de la maladie vasculaire périphérique sévère ($P = 0,006$) et l'année de chirurgie ($P = 0,02$).

À l'instar de la sélection des variables pour le modèle de régression logistique pour la mortalité hospitalière décrit plus tôt, j'ai d'abord effectué des modèles univariés puis multivariés en y incluant les variables avec une valeur de P de moins de 0,2 et en les éliminant une à une jusqu'au modèle final où toutes les variables ont une valeur de P de moins de 0,05, tel que rapporté dans l'article en Table 4. De façon pratique, j'ai d'abord estimé les facteurs de risque pour la phase précoce puis ceux de la phase tardive. Ceci m'a également permis de réaliser des figures démontrant l'impact de `age` et `experience` sur la courbe de survie paramétrique en fixant la valeur des variables à des points précis (Figure 6).

Au cours des 25 dernières années, les résultats précoces et tardifs de dissection aortique aiguë ascendante se sont améliorés en association avec l'incorporation de nouvelles méthodes et technologies, y compris l'échocardiographie, l'imagerie CT, l'utilisation de l'anastomose distale ouverte, et l'amélioration des techniques de protection cérébrale. La mortalité hospitalière a diminuée de 21% à 4% comparant les chirurgies effectuées au début et à la fin de l'expérience (χ^2 pour la tendance $P = 0,007$).

Figure 6. Courbe de survie ajustée selon l'âge du patient⁶⁸.



Courbe de survie ajustée selon le modèle paramétrique multivarié en fixant l'âge du patient à 35, 50, 65 et 80 ans. Les estimés de valeur prédite sont présentées avec un écart-type (correspondant à un intervalle de confiance à 68%) et sont tronquées à la valeur maximale de suivi observée pour chaque groupe. Dans ce modèle, les autres variables continues sont fixées à leur valeur moyenne et des variables binaires sont fixées à zéro (i.e. absent) (*modifié de la Figure 2A dans l'article original*⁶⁸).

Les résultats de l'analyse multivariée selon le modèle paramétrique ont été comparés au modèle semi-paramétrique de Cox. Il est important de rappeler que le modèle de Cox suppose que l'effet de chaque variable est constant dans le temps (risque proportionnel). Il n'est donc pas surprenant de constater que les facteurs de risque identifiés à la phase précoce dans le modèle paramétrique et dont l'effet est faible (*salvage* et *crossclamp_time* ; voir Table 4 dans l'article⁶⁸), ne ressortent pas dans le modèle de Cox multivarié (Table 1).

Table 1. Facteurs de risque de décès au suivi, selon le modèle de Cox multivarié.

	Rapport de risque [IC 95%]	P
Âge	1,06 [1,04-1,08]	<,001
Créatinine sérique à l'admission (en mg/dL) (<i>transf. log</i>)	1,88 [1,04-3,39]	0,037
Syncope	2,10 [1,23-3,61]	0,007
Chirurgie en extrême urgence	1,50 [0,65-3,42]	0,341
Temps de clampage aortique (<i>transf. log</i>)	1,53 [0,75-3,12]	0,245
Année de la chirurgie (<i>transf. au carré</i>)	0,77 [0,65-0,91]	0,002
Maladie vasculaire périphérique	1,95 [1,11-3,45]	0,021

Par contre, si on limite le suivi maximal à 12 mois (équivalent ici de la phase précoce ; voir Figure 4) et que l'on s'en tient aux variables identifiées dans la phase précoce uniquement (i.e. excluant maladie vasculaire périphérique), l'effet des variables *salvage* et *crossclamp_time* redevient « significatif » et ces variables demeurent dans le modèle de Cox final (Table 2). La durée de la phase précoce peut varier selon les cohortes étudiées et n'est pas nécessairement synonyme de la période d'hospitalisation initiale^{119-121,132}. Les modèles de décomposition du risque en phases permettent de préciser la durée de ces périodes de risque plus élevés. Ceci me permet de mieux conseiller/informer mes patients sur ce qu'ils peuvent s'attendre de leur évolution postopératoire et des périodes où ils sont plus à risque d'une complication.

Le risque de décès au suivi est nettement réduit après la période périopératoire, bien que la survie des patients vivants à la fin de leur hospitalisation initiale reste inférieure par rapport à la population générale apparié pour l'âge, le sexe et la race.

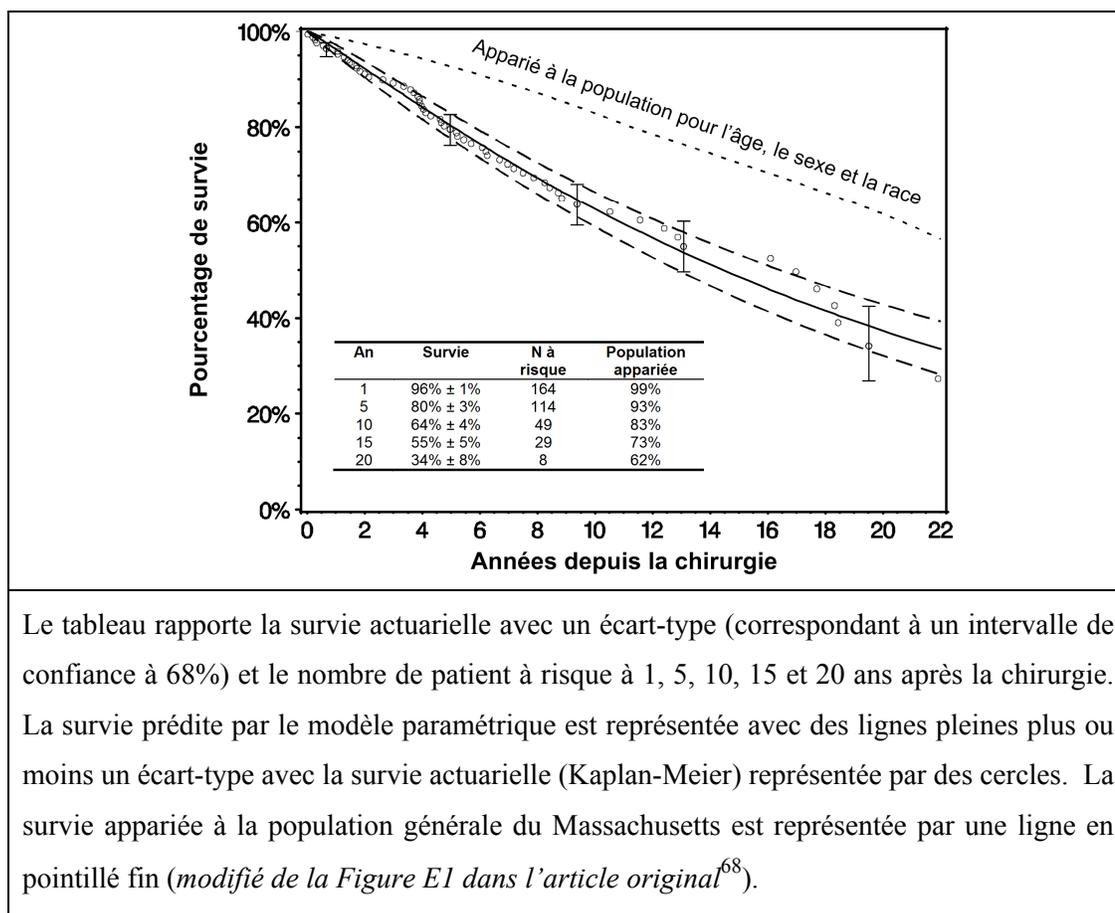
Table 2. Facteurs de risque de décès lorsque le suivi maximal est limité à 12 mois, selon le modèle de Cox multivarié.

	Rapport de risque [IC 95%]	P
Âge	1,07 [1,03-1,11]	<,001
Créatinine sérique à l'admission (en mg/dL) (<i>transf. log</i>)	5,24 [1,92-14,34]	0,001
Syncope	2,80 [1,31-5,97]	0,008
Chirurgie en extrême urgence	2,99 [1,02-8,74]	0,046
Temps de clampage aortique (<i>transf. log</i>)	4,10 [1,13-14,82]	0,031
Année de la chirurgie (<i>transf. au carré</i>)	0,65 [0,50-0,83]	<,001

La survie des patients qui ne sont pas décédés durant leur hospitalisation initiale a été modélisée avec une courbe paramétrique en utilisant comme temps « 0 » la date de leur congé de l'hôpital au lieu de la date de la chirurgie. Cette fois-ci, seulement une seule phase de risque a pu être identifiée (modèle paramétrique de type Weibull). Cette courbe de survie a été comparée à la survie de la population générale du Massachusetts appariée pour l'âge, le sexe et la race en utilisant les tables de statistique de survie de l'État du Massachusetts de 1989 à 1991¹⁴⁵ (Figure 7). Ces tables de statistiques de survie rapportent la proportion d'individus (q_x) qui décèderont au cours de l'année parmi ceux qui étaient vivant au début de l'année et ce pour chaque âge de vie et pour chaque sexe et race (caucasien, noire, autre)²⁰⁷. Ces estimés sont basés sur des recensements effectués au Massachusetts entre 1989 et 1991. Il est alors possible d'obtenir pour chacun des patients de l'étude un estimé du risque de décès annuel pour un individu de même âge, sexe et race, à partir du moment de leur chirurgie jusqu'à la fin de leur suivi. On calcule ensuite la moyenne

des risque de décès annuel pour l'ensemble de la cohorte à un an de suivi, deux ans de suivi, etc.. La courbe de survie appariée à la population est enfin calculée en multipliant le pourcentage de survie annuel (1- risque de décès annuel) par le pourcentage de survie de l'année de suivi précédente. Par exemple, la survie appariée au temps zéro est forcément de 100%, si le risque de décès annuel moyen de la première et de la deuxième année de suivi sont de 1% et 2%, respectivement, la survie appariée à un an de suivi est de (100% x [100%-1%]=99%) et à deux ans de suivi de (99% x [100%-2%] = 97,02%).

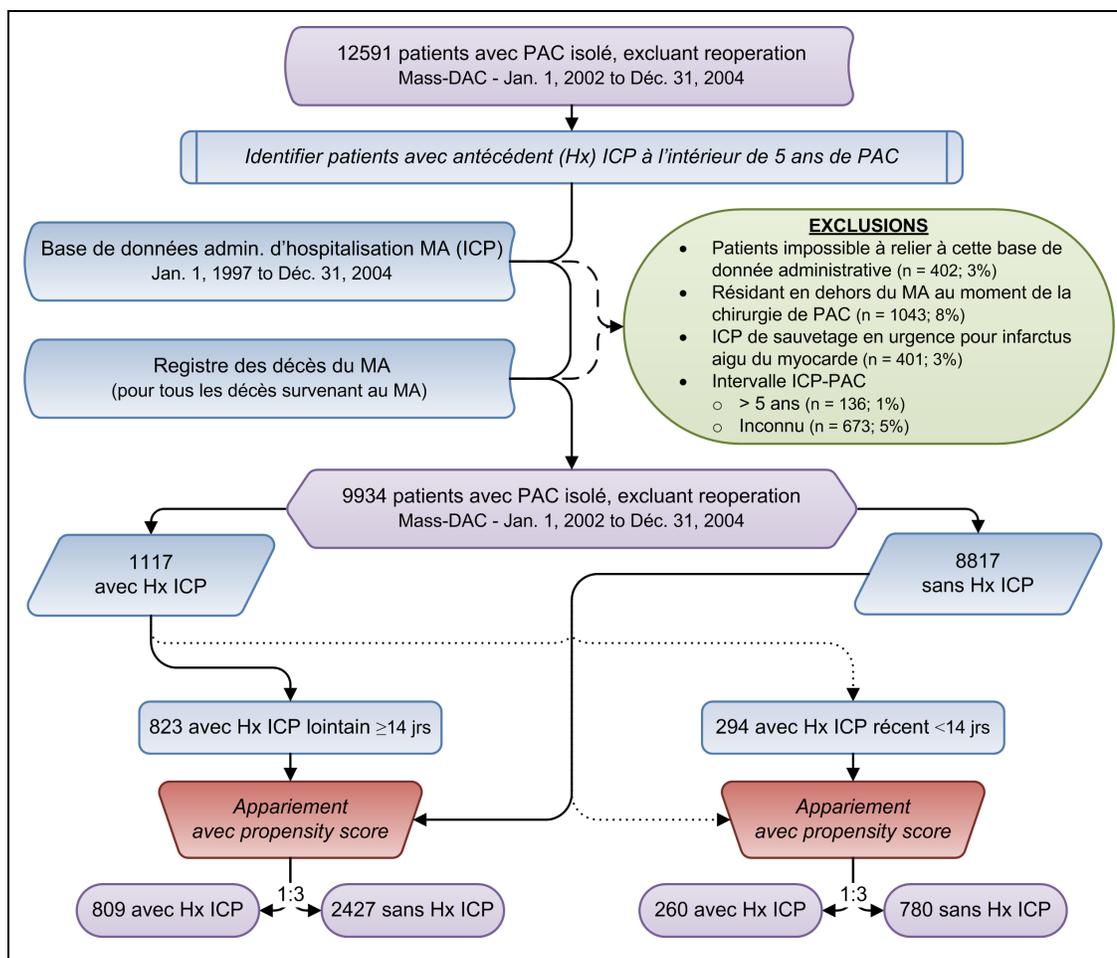
Figure 7. Courbe de survie des patients survivant à l'hospitalisation comparée à la population générale du Massachusetts appariée selon l'âge, le sexe et la race⁶⁸.



PONTAGE CORONARIENS APRÈS INTERVENTION CORONARIENNE PERCUTANÉE

L'analyse de cette étude porte sur un sous-groupe de patients référés en chirurgie cardiaque pour une première revascularisation chirurgicale par pontage coronarien (PAC) dont certains auront été exposé à une ou plusieurs interventions coronariennes percutanées (ICP). Ce travail ne se veut pas une mise en garde contre les ICP même si on remarque que le pourcentage de patients référés en chirurgie cardiaque avec antécédent d'ICP a augmenté ces dernières années. Il n'est pas réaliste de juger *a posteriori* des indications qui ont justifié l'ICP au lieu du PAC en première intention, et même s'il était possible d'obtenir cette information, il faudrait également s'intéresser à tous les autres patients qui ont bénéficié d'ICP et n'ont pas nécessité de chirurgie cardiaque par la suite soit en raison d'un résultat satisfaisant et durable, soit en raison de comorbidités rendant la chirurgie cardiaque non réaliste au moment de l'ICP initiale ou au moment d'une réintervention, ou bien en raison de décès relié ou non à l'ICP. L'analyse présentée s'intéressait à un sous groupe bien précis de patients PAC ayant bénéficié d'ICP dans les meilleures conditions en excluant les patients s'étant présenté avec un infarctus du myocarde où l'hémodynamicien est plus susceptible d'adresser en priorité l'artère coupable pour diminuer les conséquences néfastes de l'infarctus quitte à accepter une revascularisation incomplète si l'état ou l'anatomie des artères coronaire ne rends pas possible une revascularisation complète. Le devis de l'étude, décidé *a priori* avant d'avoir accès aux données, est représenté schématiquement à la Figure 8.

Figure 8. Devis d'étude pour l'analyse des résultats des patients PAC ayant ou non un antécédent d'ICP⁶⁹



La cohorte de patients opérés en chirurgie cardiaque pour une première opération de PAC isolé de 2002 à 2004 (3 ans) a été obtenue à partir de la base de données à participation obligatoire de l'État du Massachusetts (MA). J'ai lié (*linked*) cette base de donnée à la base de données administrative des patients hospitalisés du MA dans le but d'identifier les patients ayant bénéficié d'une ICP dans les 5 ans précédant les PAC. La base de données a été également liée au registre des décès du MA pour connaître la survie des patients suivant leur ICP (pour cette raison, j'ai exclu les patients résidant dans un autre État que MA). L'ovale vert rapporte les exclusions. J'ai apparié chaque patient ayant un antécédent d'ICP à 3 patients sans histoire d'ICP, en utilisant un "propensity score", selon que la dernière ICP est effectuée moins de 14 jours avant le PAC ou 14 jours ou plus avant le PAC.

Pour ce travail, j'ai utilisé des modèles de régression logistique, linéaire et de Cox et des analyses sur un sous-groupe de patients appariés selon le "*propensity score*" (PSc) utilisant des modèles de régression adaptés aux groupes appariés : "*generalized estimating equations*" et des modèles de Cox stratifiés. Tel qu'indiqué dans l'article, j'ai obtenu des résultats similaires avec les modèles de régression plus classiques comparé aux analyses avec groupes appariés.

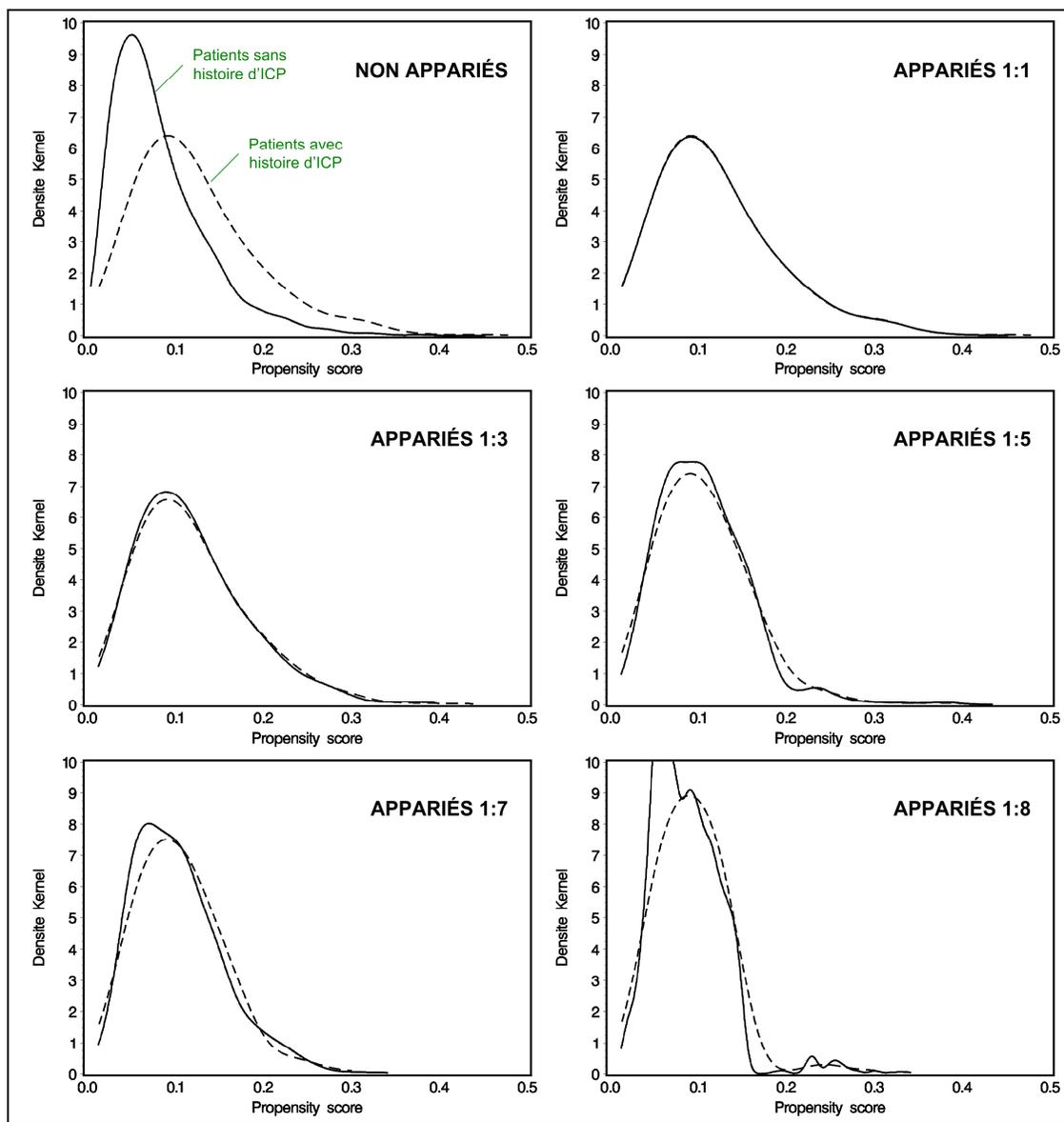
La technique d'appariement 1:3 ainsi que son effet sur la distribution du PSc des deux groupes comparés a déjà été discuté (voir Chapitre I). Plusieurs ratios d'appariement ont été examinés pour cette étude (Table 3). On peut noter dans la Table 3, l'effet bénéfique de l'augmentation du nombre de sujets contrôle (i.e. sans histoire d'ICP) sur le degré de précision (intervalle de confiance, écart-type ou variance) des estimés d'effet (rapports de cote) pour le décès à 30 jours et l'utilisation de l'artère mammaire interne gauche. En effet, les intervalles de confiance des rapports de cotes diminuent jusqu'à un ratio d'appariement de 1:3-1:4. Miettinen^{208,209} et d'autres auteurs²¹⁰⁻²¹² ont démontré qu'il y avait peu de gain statistique passé un ratio d'appariement de plus de 1:4. Dans cet exemple, on note également que passé un ratio d'appariement de 1:4, il devient difficile de trouver cinq sujets contrôles et plus par patient avec histoire d'ICP. Ceci entraîne l'exclusion de certains patients avec histoire d'ICP augmentant la valeur des intervalles de confiance et écart-types et diminuant la chance de rejeter l'hypothèse nulle (les valeurs de *P* augmentant). La distribution des PSc dans chacun des groupes devient aussi moins uniforme à mesure que les ratios d'appariement augmentent (Figure 9).

Table 3. Effet des ratios d'appariement sur le nombre de patients dans chaque groupe et sur les rapports de cote (*odds ratio* ; OR) du décès à 30 jours et de l'utilisation de l'artère mammaire interne gauche (AMIG).

Ratio d'appariement	Patients (n)		Décès à 30 jours			Utilisation de l'AMIG		
	Avec Hx ICP	Sans Hx ICP	OR [IC 95%]	OR SD	P	OR [IC 95%]	OR SD	P
Non appariés	823	8819	0,82* [0,34-2,02]	0,377	0,672	0,57* [0,44-0,75]	0,080	< 0,001
1:1	823	823	0,90 [0,36-2,24]	0,418	0,819	0,58 [0,37-0,89]	0,127	0,012
1:2	822	1644	0,69 [0,32-1,48]	0,270	0,341	0,61 [0,42-0,88]	0,114	0,008
1:3	809	2427	0,74 [0,36-1,55]	0,277	0,432	0,59 [0,42-0,83]	0,103	0,002
1:4	805	3220	0,80 [0,39-1,63]	0,290	0,535	0,59 [0,43-0,82]	0,099	0,002
1:5	723	3615	0,92 [0,45-1,87]	0,332	0,812	0,61 [0,43-0,86]	0,106	0,005
1:6	720	4320	1,02 [0,50-2,07]	0,368	0,958	0,61 [0,43-0,85]	0,104	0,004
1:7	720	5040	0,97 [0,48-1,95]	0,346	0,929	0,60 [0,43-0,84]	0,102	0,003
1:8	618	4944	0,98 [0,47-2,06]	0,370	0,967	0,66 [0,46-0,95]	0,121	0,024

* Rapports de cotes (*odds ratio*) obtenu par le modèle *generalized estimating equation* (GEE) multivarié qui tient compte de l'hôpital où la chirurgie a eu lieu sont ajustés pour l'âge, le sexe, la fraction d'éjection du ventricule gauche, le choc cardiogénique, la sténose du tronc commun, la dysfonction rénale et la présence d'arythmie ou de maladie vasculaire périphérique sévère. Les autres rapports de cotes sont obtenus par des modèles GEE qui tiennent compte de l'unité d'appariement. Hx ICP : histoire d'intervention coronarienne percutanée ; IC : intervalle de confiance ; SD : écart-type (*standard deviation*).

Figure 9. Distribution du “propensity score” et effet de divers ratio d’appariement⁶⁹.



Exemple de divers ratio d’appariement tirés de la base de données du second travail⁶⁹ (voir aussi Figure 10 du Chapitre I). On peut apprécier la distribution des “propensity score” (PSc) pour les deux groupes pour divers ratio d’appariement (non apparié, 1:1, 1:3, 1:5, 1:7 et 1:8). Le ratio d’appariement 1:1 permet d’obtenir des distributions de PSc identiques pour les deux groupes, mais le nombre de patient contrôle (sans histoire d’ICP) est plus faible. Les ratios d’appariement 1:7 et 1:8 entraînent des asymétries dans les distributions de PSc pour les deux groupes. Le ratio d’appariement 1:3 permet d’augmenter le nombre de patient contrôle sans trop perturber les distributions de PSc.

Dans le but de démontrer hors de tout doute qu'il n'y avait pas de différence de mortalité ou de morbidité entre les patients PAC avec ou sans histoire d'ICP, un "propensity score" a été obtenu pour tous par régression logistique. Chaque patient avec histoire d'ICP fut apparié à trois patients sans histoire d'ICP à l'aide d'un programme utilisant un algorithme "greedy" [voir Chapitre I].

Un des principaux avantages de la technique d'appariement est d'obtenir deux groupes directement comparables. L'évaluation du degré d'équilibre (*balance*) entre les distributions des caractéristiques préopératoires de deux groupes appariés doit se faire avec des techniques statistiques qui respectent le fait que les groupes sont non randomisés. Rappelons que le PSc est obtenu à l'aide des variables observées, il se peut toujours que l'effet soit dû en partie à une variable non-observée (par exemple, un facteur environnemental si on compare les populations de deux villes différentes). Si cette variable non-observée n'est pas incluse dans le modèle du PSc, on ne peut assumer que la distribution de cette variable est la même dans les deux groupes appariés^{36,213,214}.

Parmi les 809 patients PAC avec histoire d'ICP plus de 14 jours avant le PAC appariés avec succès à 2427 patients PAC sans histoire d'ICP, aucune différence n'a été notée entre les groupes, sauf pour une plus grande prévalence d'angine (différence standardisée [SD] = 24), de patients en classe fonctionnelle NYHA III-IV (SD = 16) et d'infarctus du myocarde > 24 heures avant les PAC (SD = 21) pour le groupe de patients avec histoire d'ICP.

Imai et al. suggèrent que le test statistique utilisé pour évaluer l'équilibre des variables de base de deux groupes devrait être une propriété de l'échantillon étudié et non d'une population hypothétique (à l'instar des tests basés sur la distribution

Normale ou Chi-carré) et la taille d'échantillon ne devrait pas influencer la valeur de la statistique obtenue²¹⁵. Ces propriétés sont respectées avec le test de différence

standardisée (*standardized difference* ; *sd*)²¹⁶ :
$$sd = \frac{100 \times |\bar{x}_{\text{Groupe}_A} - \bar{x}_{\text{Groupe}_B}|}{\sqrt{\frac{s_{\text{Groupe}_A}^2 + s_{\text{Groupe}_B}^2}{2}}}$$
, où :

$\bar{x}_{\text{Groupe}_A}$ est la moyenne du groupe A pour la variable x et $s_{\text{Groupe}_A}^2$ est l'écart-type au carré, ou variance, de la variable x pour le groupe A. La différence standardisée est la différence absolue entre les moyennes de la variable pour les deux groupes divisé par un estimé de l'écart-type commun (et non pas de l'erreur-type qui dépend de la taille de l'échantillon, i.e. $\left[\frac{\text{écart-type}}{\sqrt{n}} \right]$). C'est en fait une différence en unité

d'écart-type²¹⁷. Une différence standardisée de moins de 10% est généralement considérée comme indicative d'un équilibre entre les groupes appariés²¹⁸. Les tests classiques comme les Chi-carré, le test exact de Fisher, le t -test et le test de la somme des rangs de Wilcoxon (*Wilcoxon rank sum test*) ne sont pas adéquats pour comparer des groupes appariés³⁶. Ces tests dépendent de la taille de l'échantillon – des groupes appariés pourraient sembler bien équilibrés seulement en raison d'une réduction de la taille de l'échantillon dû à l'appariement²¹⁵.

Les variables « angine » et « infarctus du myocarde > 24h » n'ont pas été incluses dans le modèle de régression logistique utilisé pour créer le PSc puisqu'elles auraient pu être une conséquence de l'ICP et donc un des mécanismes potentiellement responsable des moins bon résultats anticipés chez ces patients référés ensuite pour chirurgie de pontage coronarien (voir hypothèse de l'étude).

Tel que noté plus haut, lorsque on compare les groupes appariés avec ou sans histoire d'ICP, on remarque aucune différence sauf pour une plus grande prévalence d'angine, de classe fonctionnelle NYHA III-IV et d'antécédent d'infarctus du myocarde dans le groupe avec histoire d'ICP. Ceci illustre bien que lorsqu'on ne tient pas compte d'une variable dans le PSc, soit volontairement (comme c'est le cas ici pour les variables « angine » et « infarctus du myocarde ») ou involontairement (tel qu'une variable non observée), on court le risque d'être en présence de différences systématiques entre les groupes appariés. Les groupes appariés étaient aussi légèrement différents pour la prévalence de classe fonctionnelle NYHA III-IV malgré que le PSc ait tenu compte de variables fortement corrélées avec la classe fonctionnelle NYHA telles que la fraction d'éjection du ventricule gauche et la présence de choc cardiogénique. Il n'est pas rare de retrouver de légères différences résiduelles malgré l'utilisation d'un PSc ; il faut par contre s'assurer que cela ne risque pas d'affecter la validité des analyses subséquentes. Puisque cette différence était possiblement reliée à la plus grande prévalence d'angine et d'antécédent d'infarctus du myocarde dans ce groupe, je n'ai pas modifié le PSc pour tenir compte de cette variable. D'ailleurs, l'inclusion de la variable NYHA III-IV dans le PSc rend difficile l'estimation des coefficients des variables corrélées « fraction d'éjection » et « choc cardiogénique », lesquelles sont alors associés avec une plus grande imprécision de ces coefficients (larges intervalles de confiances).

Parmi les patients PAC non-appariés, l'ICP plus de 14 jours avant les PAC n'a pas été identifiée comme un facteur de risque de mortalité à 30 jours dans l'analyse univariée (rapport de cote (OR) [IC 95%] 0,57 [0,26 – 1,22] ; $P = 0,145$) ou multivariée (OR = 0,82 [0,34 – 2,02] ; $P = 0,672$) qui tient compte de l'âge, le sexe, la maladie vasculaire périphérique, insuffisance rénale, les arythmies, la fraction d'éjection ventriculaire gauche, le choc cardiogénique, et la maladie sévère du tronc commun. La mortalité à 30 jours n'était pas différente entre les groupes non-appariés (1,1% vs 1,9% ; $P = 0,145$) ou appariés (1,1% vs 1,5% ; $P = 0,432$) de patients PAC avec ou sans histoire d'ICP plus de 14 jours avant les PAC.

Pour l'estimation de l'effet de traitement, il est également nécessaire de tenir compte de l'appariement. L'appariement a pour effet de sélectionner des sujets comparables (et quasi-identique au sein de chaque unité d'appariement) et d'éliminer les différences systématiques entre les groupes ; ces sujets sont en moyenne plus similaires que ne le sont des sujets randomisés à l'un ou l'autre des groupes. Les groupes comparés ne peuvent pas être considérés comme indépendant et l'analyse doit en tenir compte dans l'estimation des effets du traitement et de leurs mesures de dispersion. Pour des groupes appariés 1:1, on peut utiliser pour les variables catégorielles le test de McNemar ou la méthode des risque relatifs proposée par Agresti et Min²¹⁹ pour proportions corrélées, et pour les variables continues le test de t pour observations appariées (*paired t-test*) ou le test de Wilcoxon pour observations appariées (*Wilcoxon signed ranks test*). Pour la comparaison de courbes de survie Kaplan-Meier, le test décrit par Klein et Moeschberger²²⁰ peut alors être utilisé.

Pour des groupes définis par des ratios d'appariement autre que 1:1, on doit alors utiliser des techniques de régression qui tiennent compte ou qui stratifient sur la base de l'unité d'appariement car le test de McNemar ou autre tests pour variables

continues appariées ne sont pas directement utilisables. Dans mon travail, j'ai utilisé des modèles de régression logistique ou linéaire qui tiennent compte de l'unité d'appariement dans l'estimation des effets en utilisant la méthode "*generalized estimating equations*" (GEE) et le lien logit (*logit link*) pour les variables catégorielles ou le lien d'identité Normal (*normal identity-link*) pour les variables continues. Pour la comparaison des courbes de survie, l'utilisation du test de log-rank n'était pas appropriée car il ne tient pas compte de l'appariement. J'ai donc utilisé un modèle de régression de risque proportionnel de Cox (*proportional hazard model*) stratifié selon l'unité d'appariement²²¹ (voir Chapitre I).

Tel que discuté en introduction, plusieurs techniques peuvent être utilisées dans l'utilisation du PSc. J'ai choisi la technique d'appariement pour ce travail⁶⁹ et l'ajustement à l'aide d'un modèle de régression pour le troisième travail⁷⁰. Il est possible aussi d'utiliser une technique de stratification selon le PSc, que j'ai utilisé dans deux autres travaux non représentés ici^{106,107}. Les techniques d'ajustement par régression pondérée (*inverse-probability-of-treatment*, IPTW ; *standardized mortality/morbidity ratio*, SMR), sont peu utilisées en clinique bien qu'ils représentent d'intéressants modèles pour les épidémiologistes fondamentalistes, particulièrement dans l'analyse des relations de cause à effet. Certains auteurs suggèrent que l'appariement selon le PSc permet une plus grande réduction du biais de sélection du traitement que ne le fait la stratification selon le PSc^{213,222}.

Parmi les patients PAC non-appariés, l'ICP plus de 14 jours avant les PAC a été identifié comme un prédicteur de meilleure survie à long terme dans l'analyse univariée (HR 0,74 [0,60 – 0,91] ; $P = 0,004$), mais pas dans l'analyse multivariée. Graphiquement, la survie est meilleure pour les patients PAC non-appariés avec ICP plus de 14 jours comparé aux patients sans histoire d'ICP, alors que la survie des deux groupes est similaire pour les groupes appariés (Figure 2 du 2^e manuscrit).

Comme c'est souvent le cas pour les études cliniques, j'ai obtenu des résultats similaires, peu importe la technique d'application du PSc utilisée. Le tableau suivant compare les rapports de risque (*hazard ratio*) pour la comparaison des courbes de survie 1) des groupes appariés selon le PSc (1:3) ; 2) stratifié en quintile selon le PSc ; 3) avec ajustement par régression utilisant le PSc comme une variable continue dans le modèle de régression de Cox, et 3-4) avec ajustement par régression pondérée selon les techniques IPTW et SMR (Table 4).

Table 4. Résultats obtenus pour les rapports de risque entre les groupes en utilisant diverses techniques d'application du PSc.

Technique d'utilisation du PSc	Rapport de risque [IC 95%]	P
Appariement 1:3 selon le PSc	0,93 [0,74-1,18]	0,555
Stratification en quintile selon le PSc	0,99 [0,81-1,23]	0,955
Ajustement par régression, avec le PSc inclus comme variable continue	0,96 [0,78-1,18]	0,695
Ajustement par régression pondérée IPTW, avec le PSc inclus comme variable continue	0,96 [0,87-1,07]	0,490
Ajustement par régression pondérée SMR, avec le PSc inclus comme variable continue	0,96 [0,85-1,07]	0,458

ANALYSE DE COÛTS EN CHIRURGIE CARDIAQUE

Pour cette étude, j'étais intéressé par les changements dans les caractéristiques des patients de chirurgie cardiaque et des procédures effectuées (*case mix*) pour les patients référés d'une région où les patients étaient historiquement référés à un grand centre hospitalier universitaire (CHU) après l'implantation d'un centre de chirurgie cardiaque satellite. Mais puisqu'il est bien connu que la clientèle de patients référés en chirurgie cardiaque est de plus en plus complexe, je devais comparer ces changements avec les changements vécus par une cohorte similaire de patients référée au CHU. J'ai identifié les patients provenant de la zone où le centre de chirurgie cardiaque satellite a été implanté (région SAT) d'après leur provenance (à l'aide du code postal de leur résidence principale) ou d'après le lieu de pratique principal du médecin référant. Puisque le CHU reçoit une clientèle particulière provenant de tous les autres États référés pour les sur-spécialités du CHU, j'ai préféré limiter le groupe contrôle aux patients référés plus localement, dans le même rayon de distance du CHU selon les codes postaux des autres patients (exclusion de 1210 patients). Ceci m'a permis de rendre le groupe contrôle plus homogène et comparable. Aussi, je m'attendais à ce que les différences entre les régions SAT vs autres régions soient artificiellement augmentées dans les premiers moments de l'implantation du SAT. J'ai donc décidé *a priori* d'exclure les patients opérés dans les 4 mois suivant l'implantation du SAT (période de transition ; n=400).

Les modèles de régression multivariés (i.e. logistique, linéaire ou ordinaux) comprenaient une variable dichotomique pour décrire l'effet de temps « period » (avant vs après l'implantation du SAT) et une variable dichotomique pour décrire l'effet de la région « area » (région SAT vs autre région) permettant la modélisation avec analyses *two-way ANOVA* (analyse de variance).

Une fois ces quatre groupes de patients définis, i.e. avant ou après l'implantation du SAT, et provenant de la région SAT ou autre région, j'ai procédé à la modélisation et aux analyses *two-way ANOVA*, dont l'équation se résume ainsi.

$$\text{cost} = \beta_0 + \beta_1 * \text{area} + \beta_2 * \text{period}, \text{ où :}$$

- *cost* représente les coûts totaux d'hospitalisation (ou une transformation de ceux-ci ; voir ci-dessous) ;
- *area* représente la région et a une valeur de 1 s'il s'agit de la région SAT et de 0 s'il s'agit d'une autre région ;
- *period* représente la période de temps avant (valeur de 0) vs après (valeur de 1) l'implantation du SAT.

Lorsque le coefficient obtenu pour *area* (β_1) était différent de zéro (arbitrairement défini avec une valeur de *P* de moins de 5%), je concluais qu'il y avait une différence entre les deux régions (*area*) tenant compte des changements dans le temps (*period*). Et vice-versa pour *period*.

Dans le but d'explorer la possibilité que les changements vécus dans le temps pour les patients de la région SAT étaient différents des changements vécus pour les patients des autres régions, j'ai inclus dans d'autres modèles une variable d'interaction area*period.

Le modèle complet pour décrire les changements de coûts peut s'écrire ainsi :

$$\text{cost} = \beta_0 + \beta_1 * \text{area} + \beta_2 * \text{period} + \beta_3 * (\text{area} * \text{period})$$

Lorsque le coefficient obtenu pour cette variable d'interaction area*period (β_3) était différent de zéro, je concluais que l'un des groupes avait connu des changements différents de l'autre groupe.

Après l'implantation du programme SAT, moins de patients de faible complexité furent référés au CHU de la zone SAT et on a assisté à une augmentation significative de la charge de travail infirmier et des coûts d'hospitalisation. Pendant cet intervalle, les profits générés par l'hôpital pour la chirurgie cardiaque ont diminués pour les patients référés des deux régions (SAT et non-SAT), mais cette diminution était significativement plus marquée pour les patients PAC provenant de la région SAT.

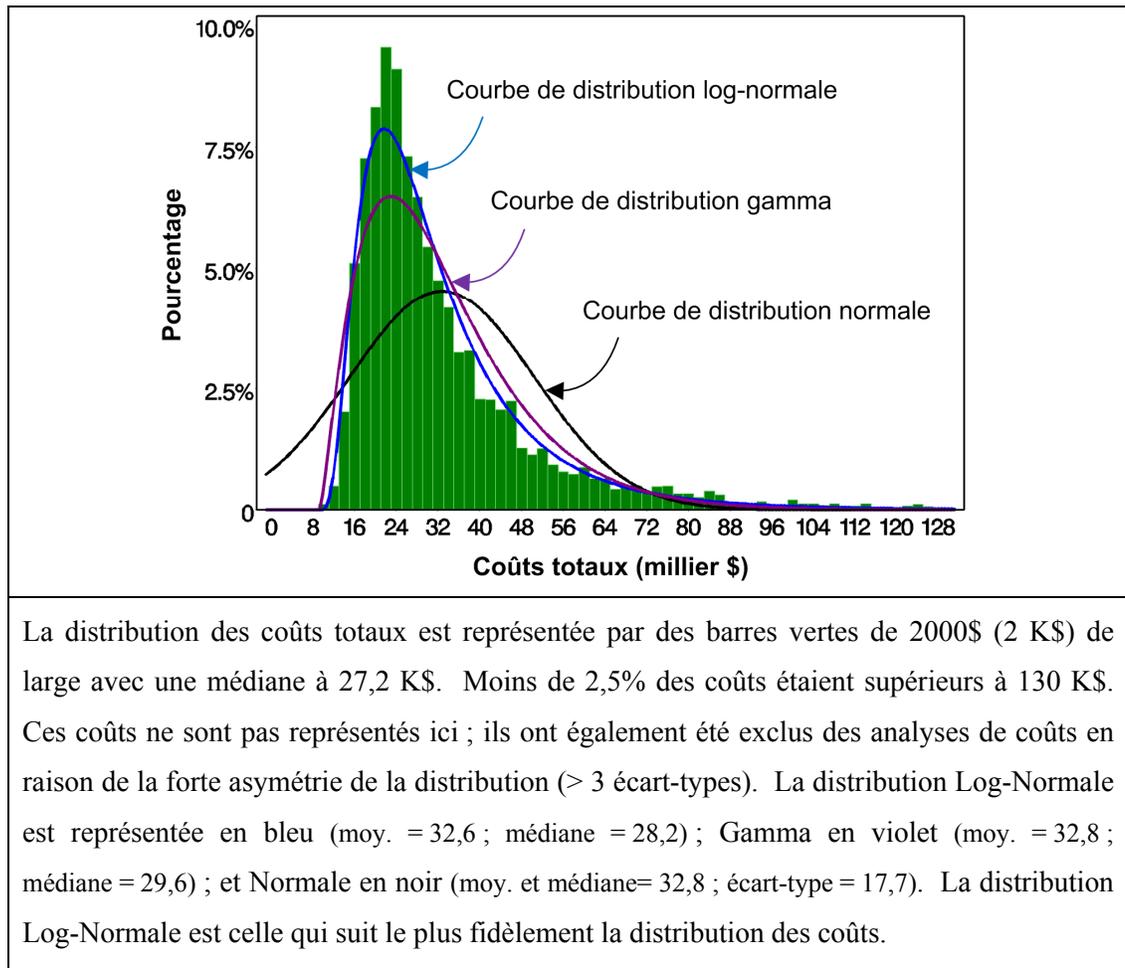
L'analyse de la démographie des quatre groupes de patients et des données chirurgicales a mise en évidence de nombreuses différences entre les groupes. Et bien que j'ai tenté d'obtenir un groupe contrôle (autre région) homogène et comparable, il n'en demeure pas moins que les caractéristiques des patients des deux groupes étaient différentes avant même l'implantation du SAT, rendant difficile l'interprétation des changements des coûts, revenus et profits d'une région par rapport à l'autre. J'ai donc décidé d'utiliser un "*propensity score*" (PSc) pour rendre

les patients des deux régions plus comparables. J'ai construit un modèle de régression logistique prédisant la chance d'un individu de provenir de la région SAT vs autre région selon ses caractéristiques pré- et peropératoires incluant vingt variables et interactions selon la méthode décrite au Chapitre I. J'ai inclus la valeur du PSc (probabilité de 0 à 1) comme une variable continue dans tous les modèles d'analyse de coûts.

Puisque les coûts varient dans le temps selon l'inflation, tous les coûts ont été convertis en \$2002 (début de la période d'étude) selon l'index des coûts à la consommation de biens médicaux (*Medical Consumer Price Index*) pour la région de Boston, Brockton et Nashua (MA, NH, ME, CT) provenant des statistiques gouvernementales disponible publiquement aux États-Unis (*U.S. Bureau of Labor Statistics*). Il est ainsi possible d'obtenir le ratio d'augmentation des coûts à chaque deux mois. J'ai utilisé une simple règle d'intrapolation pour me permettre de convertir tous les coûts en ce qu'ils auraient été en janvier 2002.

L'analyse des coûts est compliquée par la distribution asymétrique de ceux-ci. La distribution des coûts est limitée à zéro (à moins que le modèle admette des coûts négatifs) avec une médiane inférieure à la valeur moyenne et la queue de la distribution est étalée à droite (*positive or right-skewed*)¹⁴⁷. Plusieurs modèles d'analyse paramétrique et semi-paramétrique (modèle de Cox) ont été explorés pour ce travail avant de faire le choix de la distribution Log-Normale (Figure 10).

Figure 10. Distribution des coûts totaux d'hospitalisation pour la cohorte entière.



La plupart des tests statistiques et modèles de régression linéaire sont basés sur la distribution Normale ou une approximation de celle-ci (ex. : distributions de t , z ou Chi-carré). Vu la distribution asymétrique des coûts, il m'apparaissait inapproprié d'utiliser la distribution Normale (Figure 10). La distribution Log-Normale a une propriété intéressante pour la modélisation : si une variable dépendante (coûts) suit la distribution Log-Normale, le logarithme de la variable dépendante (\log -coûts) suit la distribution Normale^{147,223,224}. Les caractéristiques de la distribution Normale sont comparées à celles de la distribution Log-Normale ci-dessous (Table 5).

Table 5. Comparaison des distributions Normale et Log-Normale.

	Normale (Gaussienne)	Log-Normale (Galton)
Forme de la distribution	Symétrique	Asymétrique
Effets des variables du modèle	Additifs	Multiplicatifs
Paramètres	mu : μ sigma : σ	zeta ou paramètre d'échelle (<i>scale</i>) : μ_* sigma ou paramètre de forme (<i>shape</i>) : σ_*
Fonction de densité de probabilité (<i>probability density function</i>)	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)}$	$\frac{1}{x\sqrt{2\pi\sigma_*^2}} e^{-\left(\frac{(\ln x - \mu_*)^2}{2\sigma_*^2}\right)}$
Moyenne	μ	$e^{\left(\mu_* + \frac{\sigma_*^2}{2}\right)}$
Médiane (aussi appelée moyenne géométrique pour la distribution Log-Normale)	μ	e^{μ_*}
Mode	μ	$e^{(\mu_* - \sigma_*^2)}$
Variance	σ^2	$(e^{\sigma_*^2} - 1)e^{(2\mu_* + \sigma_*^2)}$

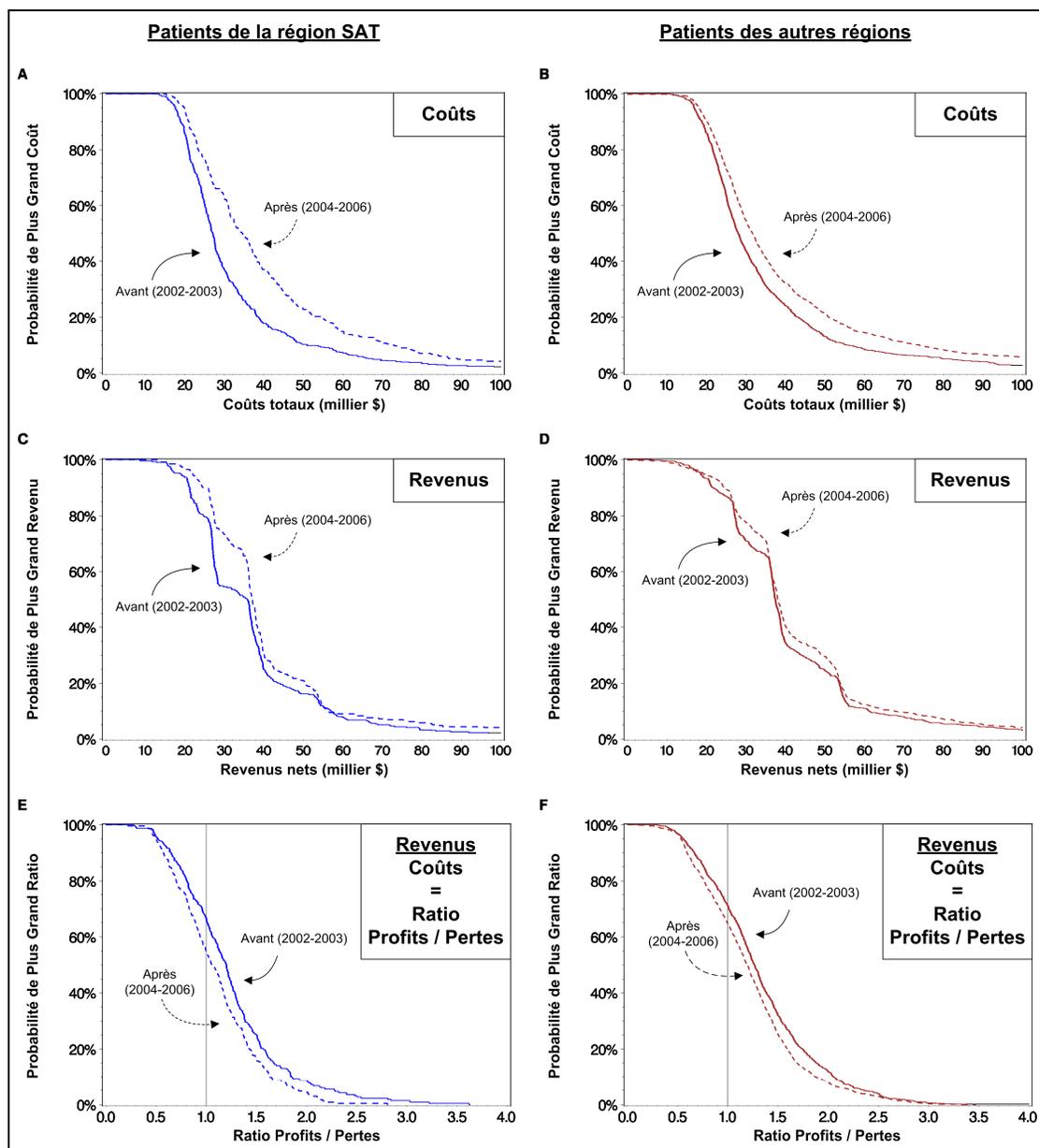
Les propriétés de la distribution Log-Normale la rendent donc intéressante pour la modélisation des coûts^{147-151,224}. En utilisant ce modèle, les inférences statistiques peuvent se faire sur l'unité d'origine (coûts) et non sur la variable transformée (*log-coûts*). Il est important de constater que l'effet des variables incluses dans modèle Log-Normal est multiplicatif au même titre que le modèle de régression logistique et les modèles d'analyse de survie tel que le modèle de risque proportionnel de Cox, et les autres modèles paramétriques exponentiel, Weibull et Gamma.

Le modèle Log-Normal peut être estimé en utilisant un modèle “*generalized estimating equations*” (GEE) linéaire où la fonction de lien logistique (logit) est choisie. Autrement, on peut effectuer simplement la régression linéaire sur la variable transformée (*log-coûts*) et utiliser une méthode non-paramétrique ou de “*bootstrapping*” pour retransformer la valeur prédite par le modèle en unité non transformée (coûts). La retransformation avec la fonction exposant (e^β) permet d’avoir une idée de la valeur prédite mais ne compense pas pour le biais qui survient avec la retransformation. Duan a proposé un simple ajustement pour le biais qui résulte de cette retransformation^{196,225}. Cet ajustement, appelé le “*smearing estimate*” est un estimé non-paramétrique qui ne nécessite pas que la distribution sous-jacente ait une forme particulière. Bien que la valeur retransformée par la transformation exposant (e^β) soit un estimé raisonnable de la médiane de la distribution originale, ce n'est pas un très bon estimé de la moyenne de celle-ci qui, en raison de la longue queue à droite de la distribution, pourrait être sensiblement plus grande. Il y a plusieurs façons de retransformer en utilisant les équations de la distribution Log-Normale, mais le “*smearing estimate*” fonctionne bien sans que l’on ait à assumer une distribution spécifique. Pour appliquer le “*smearing estimate*”, on calcule la valeur moyenne des résidus retransformés, c'est-à-dire la moyenne des $e^{\text{résidu}}$ à travers tous les sujets. Un résidu est simplement la différence entre la valeur prédite et obtenue calculée pour chaque sujet. Bien que pour un modèle de régression linéaire, la méthode des moindres carrés fait en sorte que la moyenne des résidus est de zéro, il n’est pas nécessairement le cas que la moyenne des résidus retransformés selon la distribution Log-Normale est de un ($e^0 = 1$). En

fait, même quand une transformation logarithmique est appropriée, la valeur des résidus retransformés va presque toujours être supérieure à un, et peut-être beaucoup plus élevée. La valeur prédite avec “*smearing estimate*” est ainsi calculée en multipliant les valeurs retransformées, e^{β} , par la valeur moyenne des résidus retransformés, $e^{\text{résidu}}$.

Tel que mentionné plus tôt, les modèles d’analyse de survie se prêtent bien à l’analyse des coûts. Au lieu de modéliser des temps de survie avec ou sans censure on modélise les coûts sans censure, à moins que le modèle n’admettent des sujets qui sont toujours hospitalisés et pour lesquels on a seulement une idée partielle des coûts de l’hospitalisation. Au lieu de « courbe de survie » ou de « probabilité de survivre plus que t lorsque l’on a survécu jusqu’à t », on parle de « probabilité que les coûts d’hospitalisation soient plus que $c\$$ lorsqu’on a déjà dépensé $c\$$ ». Le modèle de Cox est particulièrement intéressant dans ce cas ci car on n’a pas à assumer de forme particulière pour la « fonction de risque de coûts » de base. L’inclusion des variables dans le modèle doit par contre respecter la règle du « risque proportionnel » et l’effet de chaque variable est multiplicatif, comme c’est le cas pour la distribution Log-Normale. La Figure 11 présente des « courbes de survie » selon la méthode de Kaplan-Meier adaptées aux coûts, revenus et ratio profits/bénéfices. Ce sont les « courbes de probabilité de plus grand coût, revenu ou ratio profits/pertes ».

Figure 11. Analyse des coûts, revenus et ratio profits/pertes selon le modèle de Kaplan-Meier.



Chaque point sur la « courbe de probabilité de coûts » donne un estimé du pourcentage de sujets dont les coûts d'hospitalisation sont plus élevés qu'une certaine valeur. Par exemple, environ 80% des sujets ont des coûts d'hospitalisation de plus de 20 K\$. La médiane des coûts (à 50%) peut aussi être estimée à partir de ces courbes, mais il est plus difficile d'estimer la moyenne des coûts. La distribution des coûts pour les périodes (2002-03 vs 2004-05) et les régions (SAT vs autres) peuvent être comparées entre-elles visuellement.

Dans le cadre de l'article présenté ici, j'ai décidé de ne pas rapporter les résultats des analyses selon le modèle de Cox étant donné que :

- Les conclusions de l'étude basées sur le modèle de régression de Cox étaient similaires à celles obtenues avec le modèle Log-Normal, et ce même si la règle du risque proportionnel n'était pas strictement respectée dans la comparaison des courbes de probabilités de coûts, revenus et ratio profits/pertes ;
- Les modèles de Kaplan-Meier et de Cox sont difficile à comprendre et interpréter pour un clinicien dans le contexte de l'analyse de coûts (concept de « probabilité de plus grand coût ») avec des rapports de cotes (*hazard ratio*) ;
- Le modèle Log-Normal permettait d'obtenir des estimés de valeur des différences de coûts moyen (et non seulement médian), prédits en \$, et ce modèle est utilisé plus couramment en économétrie.

CHAPITRE V

CONCLUSION ET PERSPECTIVES

La recherche clinique en sciences cardiovasculaire est très active. Alors que les champs d'études cliniques en cardiologie sont fort variés, qu'ils soient ou non reliés aux compagnies pharmaceutiques et de matériel thérapeutique (tuteurs et cathéters, cardiostimulateurs et défibrillateurs, ...), la recherche clinique en chirurgie cardiaque s'intéresse majoritairement à la recherche évaluative sur les résultats des procédures chirurgicales. Pour assurer une recherche de haute qualité et bénéficier des progrès récents en biostatistique appliquée, le chirurgien cardiaque doit savoir s'entourer de collaborateurs variés. En effet, le clinicien-chercheur n'a pas nécessairement besoin de savoir comment calculer les statistiques à partir des données cliniques, mais il doit pouvoir choisir la bonne méthode et interpréter les résultats avec justesse. Il bénéficie de l'apport critique de l'épidémiologiste, de l'éthicien, d'investigateurs expérimentés, biostatisticiens, secrétaires médicales, coordonnateurs de recherche, programmeurs, cardiologues, radiologistes, pathologistes, chimistes, etc. Dans mon cheminement de doctorat, j'ai été appelé à jouer quelque uns de ces rôles de collaborateur dont celui d'épidémiologiste, de biostatisticien, de programmeur et de coordonnateur de recherche.

De fait, je suis maintenant mieux en mesure de faire le pont entre le clinicien et ses collaborateurs épidémiologistes et biostatisticiens. Suite à mes collaborations en recherche clinique et biostatistique avec divers autres cliniciens, je note que le langage statistique apparaît très aride pour la plupart d'entre eux. Ceux-ci ont souvent de la difficulté à exprimer l'hypothèse de travail sur laquelle ils désirent consacrer un projet de recherche. Il n'est donc pas surprenant qu'ils ne parviennent pas à dialoguer avec l'épidémiologiste pour le choix du devis d'étude approprié pour la question d'intérêt, ou le biostatisticien dont les méthodes d'analyse sont basées sur des séries d'hypothèses et de suppositions.

Je constate qu'il devient facile de nos jours pour le clinicien de s'improviser « biostatisticien » avec les logiciels statistique mis à notre disposition et leurs interfaces “*user-friendly*” où un test de t ou une analyse de survie relativement complexe peuvent être complétés en trois à quatre clics de la souris. On devient rapidement tenté de partir à la pêche aux valeurs « statistiquement significatives » (*fishing expedition, data mining*)²²⁶, sans s'assurer si toutes les suppositions des modèles utilisés sont adressées ou si on a tenu compte des comparaisons multiples pour toutes les analyses effectuées. Il devient parfois même difficile de dépister les chercheurs qui « torturent les données » de la sorte et pour lesquels les résultats probants d'analyses peuvent n'être que le fruit du hasard, particulièrement lorsque les groupes à l'étude semblent avoir été redéfinis dans le but d'obtenir un résultat favorable ou lorsque l'on présente le résultat de sous-analyses (positives) alors que le résultat principal de l'étude est négatif. Pour mes travaux, j'ai donc tâché de définir tous mes protocoles de recherche, devis d'étude et mes hypothèses *a priori*.

Je remarque aussi que bien souvent ces analyses rapidement disponibles en quelques clics sont inappropriées en fonction des problèmes biostatistiques et des distributions des variables dépendantes et indépendantes. L'utilisation d'un programme puissant comme SAS m'a permis de construire mes propres programmes et de les adapter en fonction de mes besoins pour chacun de mes travaux. Le clinicien ne devrait pas se limiter à un test facilement accessible sans avoir vérifié auprès d'un expert si cette analyse est appropriée, surtout lorsque l'on prévoit présenter ou publier ce résultat.

Il y a souvent plus d'une façon d'analyser les données en fonction des questions qui nous intéressent ; chacune avec ses avantages et ses inconvénients qu'il faut considérer soigneusement. Il est parfois avantageux de s'en tenir à la méthode la plus classique et comprise par la plupart des cliniciens si le gain associé à une technique plus complexe est négligeable et ce dans le but de favoriser une meilleure compréhension et diffusion des résultats présentés. Par contre, il ne faut pas hésiter à sortir des sentiers battus et présenter des analyses plus complexes si les tests classiques sont inadéquats, que la nouvelle méthode permet de répondre à d'autres questions d'intérêts ou d'optimiser l'utilisation de toutes les données disponibles. Parfois, comme c'est le cas avec les analyses longitudinales, l'utilisation d'une analyse plus appropriée qui considère toute l'information disponible permet même d'obtenir des résultats positifs alors qu'une analyse plus conventionnelle (analyse d'un seul point dans le temps – analyses transversales) n'aurait pas permis de mettre en évidence ces différences.

Dans cet ouvrage, j'ai présenté trois applications avancées des analyses de survie et deux applications du "*propensity score*" (PSc) appliquées à la chirurgie cardiaque : courbes de survie paramétrique et la décomposition dans le temps du niveau de risque en phase, courbes de survies adaptées aux groupes appariés selon le PSc et courbes de survies paramétriques et semi-paramétrique pour les analyses de coûts conjointement avec le PSc. Les analyses utilisées été décortiquées pour en expliquer la rationnelle et les résultats de modèles alternatifs ont été présentés.

Le modèle d'analyse de survie avec décomposition des phases de risque dans le temps nécessite un suivi assez long et le plus complet possible pour permettre de bien définir les diverses phases de suivi et de spécifier adéquatement le modèle paramétrique. Les programmes actuellement disponibles pour effectuer ces analyses sont plutôt difficile à maîtriser, mais l'analyse est plus rapide une fois que l'on a bien compris chacune des étapes. Dans le cadre de mon premier travail, même avec un suivi moyen de $7,0 \pm 5,9$ ans et maximal de plus de 25 ans, je n'ai été en mesure de définir que les phases de risque précoce et tardive (en fait cette phase tardive comprends à la fois une phase constante et tardive de par le risque qui demeure à peu près le même pour toute la période de suivi et augmente surtout vers la fin). Il est envisageable que j'aie pu définir plus précisément une troisième phase avec une cohorte plus grosse et un suivi plus long, ou bien que le modèle que j'ai défini demeure suffisamment robuste pour un suivi plus long. Je recommande donc ce genre de modèle lorsque le suivi disponible est suffisamment long (suivi moyen de plus de 5 ans) et surtout lorsque l'on s'attend à ce que le risque varie dans le temps de façon prédictible.

L'intervention chirurgicale représente une bonne situation pour ce genre de modèle avec son risque plus élevé durant et immédiatement après l'intervention, suivi d'une phase d'amélioration de la condition du patient lorsque l'intervention est curative et d'une phase de risque augmenté après plusieurs années de suivi, soit par échec de la procédure avec le temps (comme le remplacement valvulaire avec prothèse biologique qui se détériorent progressivement) ou par décès du patient relié ou non avec la pathologie de base. Une fois le modèle défini, on peut s'en servir pour produire d'intéressantes courbes de survie prédites pour des sous-groupes de patients définis par les variables incluses dans le modèle. Il devient aussi envisageable de produire des courbes de survie « sur mesure » pour les patients rencontrés en préopératoire pour leur exposer l'évolution postopératoire attendue au plus long terme en fonction de leur facteurs de risque préopératoire.

La technique d'appariement selon le PSc est conceptuellement plus simple à saisir que les autres applications du PSc et les résultats sont plus facile à interpréter que des modèles de régression multivariés plus classiques. Elle permet de produire des groupes appariés directement comparables chez lesquels il est relativement simple de démontrer que les groupes sont biens équilibrés (différence standardisée). Lors de son application, on doit s'assurer de bien décrire : 1) le choix des variables pour le PSc et sa rationnelle ; 2) le type de modèle utilisé pour le PSc ; 3) l'étendue de PSc utilisé pour former les paires ; 4) la technique d'appariement utilisée : “*greedy*” vs optimale ou autre, avec ou sans remplacement ; et enfin 5) utilisation de méthodes d'analyse des résultats pour les groupes appariés qui tiennent compte de l'appariement.

Dans le cadre du troisième travail, le PSc a été utilisé comme variable continue dans les modèles de régression des analyses de coûts. Le choix de cette technique d'application du PSc est justifiée par le fait que les groupes à l'étude étaient plus petits en nombre et qu'une technique d'appariement aurait entraîné l'élimination de plusieurs patients et ainsi limité la capacité de mettre en évidence des différences entre les groupes à l'étude (diminution de la puissance). Dans cet exemple, les distributions du PSc pour les groupes à l'étude n'étaient pas grossièrement asymétriques permettant inclusion du PSc en tant que variable continue dans les modèles (et les résidus du modèle d'ajustement par régression avaient une distribution approximativement Normale – voir Chapitre I) et elles présentaient un chevauchement suffisant pour justifier la comparaison des groupes.

La comparaison des coûts d'hospitalisation en chirurgie cardiaque est un problème complexe pour lequel plusieurs solutions sont envisageables. L'utilisation d'un modèle d'analyse de survie semi-paramétrique de Cox est élégante car celui-ci est très robuste et s'adapte bien aux distributions asymétriques des coûts, au même titre que les distributions des temps de suivi pour l'analyse de survie sont asymétriques. Par contre, la présentation des résultats sous forme de courbes de survie et de rapports de risque (*Hazard Ratio*) est plutôt abstraite et il est plus difficile de résumer l'information par des moyennes de coûts avec écarts-types. Le modèle Log-Normal représente une solution intéressante lorsque la distribution des coûts respecte cette distribution paramétrique. Ce modèle est parfois utilisé comme modèle d'analyse de survie paramétrique et peut être utile dans cette situation lorsque l'on dispose de suffisamment d'information pour s'assurer que le modèle est bien défini.

Ce modèle paramétrique est peu utilisé dans la littérature médicale contemporaine vu la popularité du modèle de Cox. Lorsqu'il est bien spécifié, le modèle paramétrique Log-Normal peut être plus puissant que le modèle de Cox. Comme nous devons travailler avec des données de suivi souvent limitées et imparfaites en recherche évaluative en chirurgie cardiaque, le modèle d'analyse de survie de Cox est probablement un meilleur choix dans la plupart des situations. Pour l'analyse de coûts où les données sont plus complètes avec pas ou très peu de « censure », le modèle Log-Normal s'avère particulièrement utile et probablement préférable au modèle de Cox.

Pour évaluer si les différences de coûts entre deux groupes sont « significatives », on doit comparer ces différences en fonction de la variabilité de la mesure. Pour l'analyse des coûts, le degré de variabilité limite la capacité des analyses à démontrer des différences entre les groupes malgré des différences de coûts qui semblent importantes à première vue. L'utilisation du modèle Log-Normal a permis la transformation des coûts sur une échelle logarithmique où la variabilité était moins grande. De plus, j'ai exclu de l'analyse les valeurs extrêmes qui avaient trop de poids dans l'analyse. Pour mes analyses, j'ai défini *a priori* un seuil arbitraire de 0,05 pour la valeur de P jugée « statistiquement significative ». En rétrospective, une valeur de P moins conservatrice aurait pu être utilisée (0,1?) vu les différences « administrativement significatives » notées entre les groupes et l'importance de ces données d'un point de vue d'allocation des coûts de santé. J'ai pu retrouver relativement peu d'article de recherche statistique en analyses de coûts médicaux utilisant le modèle Log-Normal. Je crois qu'il s'agit là d'une intéressante avenue de recherche en statistique où les normes restent à définir.

Je débute ma carrière de chercheur-clinicien au sein du Département de Chirurgie de l'Université de Montréal et j'ai bien l'intention de continuer à pousser plus loin les limites de mes connaissances en biostatistiques appliquées et d'en faire profiter mes collègues et amis. Dans le cadre de mon programme de recherche évaluative en chirurgie cardiaque (clinicien-chercheur-boursier junior 1 au Fonds de la Recherche en Santé du Québec), je m'intéresse aux analyses de larges bases de données et plus particulièrement aux analyses de survie ainsi qu'aux analyses longitudinales appliquées à la chirurgie cardiaque (modèles « *mixed effects* » avec variable résultat répétée dans le temps).

CHAPITRE VI

SOURCES DOCUMENTAIRES

1. Faculté des études supérieures et postdoctorales de l'Université de Montréal: Guide de présentation et d'évaluation des mémoires et des thèses de doctorat. 2009;
http://www.fesp.umontreal.ca/fileadmin/Documents/Cheminement/GuidePresentationMemoireThese_versionDS.pdf. Accessed January 23, 2012.
2. Kurichi JE, Sonnad SS. Statistical methods in the surgical literature. *J Am Coll Surg*. Mar 2006;202(3):476-484.
3. Agarwal S, Colditz GA, Emerson JD. Chapter 3. Use of Statistical Analysis in the New England Journal of Medicine. In: Bailar JC, Hoaglin DC, eds. *Medical uses of statistics*. 3rd ed. Hoboken, N.J.: John Wiley & Sons; 2009:xxix, 491 p.
4. Emerson JD, Colditz GA. Use of statistical analysis in the New England Journal of Medicine. *N Engl J Med*. Sep 22 1983;309(12):709-713.
5. Horton NJ, Switzer SS. Statistical methods in the journal. *N Engl J Med*. Nov 3 2005;353(18):1977-1979.

6. Robinson PM, Menakuru S, Reed MW, Balasubramanian SP. Description and reporting of surgical data--scope for improvement? *Surgeon*. Feb 2009;7(1):6-9.
7. Manterola C, Pineda V, Vial M, Losada H. What is the methodologic quality of human therapy studies in ISI surgical publications? *Ann Surg*. Nov 2006;244(5):827-832.
8. Bodnar E, Blackstone EH. Editorial: An 'actual' problem: Another issue of apples and oranges. *The Journal of heart valve disease*. Nov 2005;14(6):706-708.
9. Blackstone EH, Rice TW. Clinical-pathologic conference: use and choice of statistical methods for the clinical study, "superficial adenocarcinoma of the esophagus". *J Thorac Cardiovasc Surg*. Dec 2001;122(6):1063-1076.
10. Blackstone EH. Comparing apples and oranges. *J Thorac Cardiovasc Surg*. Jan 2002;123(1):8-15.
11. Grunkemeier GL, Wu Y. Interpretation of nonfatal events after cardiac surgery: actual versus actuarial reporting. *J Thorac Cardiovasc Surg*. Aug 2001;122(2):216-219.
12. Grunkemeier GL, Jin R, Eijkemans MJ, Takkenberg JJ. Actual and actuarial probabilities of competing risks: apples and lemons. *Ann Thorac Surg*. May 2007;83(5):1586-1592.
13. Wechsler AS. Statistics for the rest of us. *J Thorac Cardiovasc Surg*. Aug 2001;122(2):207-208.

14. Blackstone EH. Breaking down barriers: helpful breakthrough statistical methods you need to understand better. *J Thorac Cardiovasc Surg.* Sep 2001;122(3):430-439.
15. Anderson WN. Statistical techniques for validating logistic regression models. *Ann Thorac Surg.* Oct 2005;80(4):1169.
16. Anderson RP, Jin R, Grunkemeier GL. Understanding logistic regression analysis in clinical reports: an introduction. *Ann Thorac Surg.* Mar 2003;75(3):753-757.
17. Grunkemeier GL, Wu Y. Actual versus actuarial event-free percentages. *Ann Thorac Surg.* Sep 2001;72(3):677-678.
18. Blackstone EH. Neurologic injury from cardiac surgery-an important but enormously complex phenomenon. *J Thorac Cardiovasc Surg.* Oct 2000;120(4):629-631.
19. Grunkemeier GL, Payne N. Bayesian analysis: a new statistical paradigm for new technology. *Ann Thorac Surg.* Dec 2002;74(6):1901-1908.
20. Grunkemeier GL, Wu Y. Bootstrap resampling methods: something for nothing? *Ann Thorac Surg.* Apr 2004;77(4):1142-1144.
21. Grunkemeier GL, Furnary AP. Mandatory database participation: risky business? *Ann Thorac Surg.* Sep 2005;80(3):799-801.
22. Grunkemeier GL, Jin R, Wu Y. Cumulative sum curves and their prediction limits. *Ann Thorac Surg.* Feb 2009;87(2):361-364.
23. Grunkemeier GL, Wu YX, Furnary AP. Cumulative sum techniques for assessing surgical results. *Ann Thorac Surg.* Sep 2003;76(3):663-667.

24. Grunkemeier GL, Jin R. Power and sample size: how many patients do I need? *Ann Thorac Surg.* Jun 2007;83(6):1934-1939.
25. Grunkemeier GL, Jin R, Starr A. Prosthetic heart valves: Objective Performance Criteria versus randomized clinical trial. *Ann Thorac Surg.* Sep 2006;82(3):776-780.
26. Grunkemeier GL, Payne N, Jin R, Handy JR, Jr. Propensity score analysis of stroke after off-pump coronary artery bypass grafting. *Ann Thorac Surg.* Aug 2002;74(2):301-305.
27. Grunkemeier GL, Jin R. Receiver operating characteristic curve analysis of clinical risk models. *Ann Thorac Surg.* Aug 2001;72(2):323-326.
28. Grunkemeier GL, Wu Y. What are the odds? *Ann Thorac Surg.* Apr 2007;83(4):1240-1244.
29. Grunkemeier GL, Wu Y, Furnary AP. What is the value of a p value? *Ann Thorac Surg.* May 2009;87(5):1337-1343.
30. Jones CM, Athanasiou T. Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. *Ann Thorac Surg.* Jan 2005;79(1):16-20.
31. Mackenzie TA, Brown JR, Likosky DS, Wu Y, Grunkemeier GL. Review of case-mix corrected survival curves. *Ann Thorac Surg.* May 2012;93(5):1416-1425.
32. Takkenberg JJ, Puvimanasinghe JP, Grunkemeier GL. Simulation models to predict outcome after aortic valve replacement. *Ann Thorac Surg.* May 2003;75(5):1372-1376.

33. Wu Y, Takkenberg JJ, Grunkemeier GL. Measuring follow-up completeness. *Ann Thorac Surg.* Apr 2008;85(4):1155-1157.
34. Wu Y, Jin R, Gao G, Grunkemeier GL, Starr A. Cost-effectiveness of aortic valve replacement in the elderly: an introductory study. *J Thorac Cardiovasc Surg.* Mar 2007;133(3):608-613.
35. Rice TW, Blackstone EH, Goldblum JR, DeCamp MM, Murthy SC, Falk GW, Ormsby AH, Rybicki LA, Richter JE, Adelstein DJ. Superficial adenocarcinoma of the esophagus. *J Thorac Cardiovasc Surg.* Dec 2001;122(6):1077-1090.
36. Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med.* May 30 2008;27(12):2037-2049.
37. Austin PC. Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: a systematic review and suggestions for improvement. *J Thorac Cardiovasc Surg.* Nov 2007;134(5):1128-1135.
38. Austin PC. Primer on statistical interpretation or methods report card on propensity-score matching in the cardiology literature from 2004 to 2006: a systematic review. *Circulation. Cardiovascular quality and outcomes.* Sep 2008;1(1):62-67.
39. Austin RL, Rune A, Bouzakri K, Zierath JR, Krook A. siRNA-mediated reduction of inhibitor of nuclear factor-kappaB kinase prevents tumor necrosis factor-alpha-induced insulin resistance in human skeletal muscle. *Diabetes.* Aug 2008;57(8):2066-2073.

40. Stuart EA. Developing practical recommendations for the use of propensity scores: discussion of 'A critical appraisal of propensity score matching in the medical literature between 1996 and 2003' by Peter Austin, *Statistics in Medicine*. *Stat Med*. May 30 2008;27(12):2062-2065; discussion 2066-2069.
41. Hill J. Discussion of research using propensity-score matching: comments on 'A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003' by Peter Austin, *Statistics in Medicine*. *Stat Med*. May 30 2008;27(12):2055-2061; discussion 2066-2059.
42. Walter CJ, Dumville JC, Hewitt CE, Moore KC, Torgerson DJ, Drew PJ, Monson JR. The quality of trials in operative surgery. *Ann Surg*. Dec 2007;246(6):1104-1109.
43. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med*. Aug 13 1987;317(7):426-432.
44. Olsen CH. Review of the use of statistics in infection and immunity. *Infect Immun*. Dec 2003;71(12):6689-6692.
45. Gore SM, Jones IG, Rytter EC. Misuse of statistical methods: critical assessment of articles in BMJ from January to March 1976. *Br Med J*. Jan 8 1977;1(6053):85-87.
46. Pocock SJ, Collier TJ, Dandreo KJ, de Stavola BL, Goldman MB, Kalish LA, Kasten LE, McCormack VA. Issues in the reporting of epidemiological studies: a survey of recent practice. *BMJ*. Oct 16 2004;329(7471):883.

47. Mullner M, Matthews H, Altman DG. Reporting on statistical methods to adjust for confounding: a cross-sectional survey. *Ann Intern Med.* Jan 15 2002;136(2):122-126.
48. Groenwold RH, Van Deursen AM, Hoes AW, Hak E. Poor quality of reporting confounding bias in observational intervention studies: a systematic review. *Ann Epidemiol.* Oct 2008;18(10):746-751.
49. Poorolajal J, Cheraghi Z, Irani AD, Rezaeian S. Quality of Cohort Studies Reporting Post the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement. *Epidemiol Health.* 2011;33:e2011005.
50. Nagele P. Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals. *Br J Anaesth.* Apr 2003;90(4):514-516.
51. Goodman NW, Powell CG. Could do better: statistics in anaesthesia research. *Br J Anaesth.* Jun 1998;80(6):712-714.
52. Huang W, LaBerge JM, Lu Y, Glidden DV. Research publications in vascular and interventional radiology: research topics, study designs, and statistical methods. *J Vasc Interv Radiol.* Mar 2002;13(3):247-255.
53. Altman DG. The scandal of poor medical research. *BMJ.* Jan 29 1994;308(6924):283-284.
54. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet.* Apr 14 2001;357(9263):1191-1194.

55. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *Jama*. Aug 28 1996;276(8):637-639.
56. Blackstone EH. CONSORT and beyond. *J Thorac Cardiovasc Surg*. Aug 2006;132(2):229-232.
57. von Elm E, Altman DG, Egger M, Pocock SJ, Gotsche PC, Vandenbroucke JP. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ*. Oct 20 2007;335(7624):806-808.
58. Allison PD, SAS Institute. *Survival analysis using the SAS system : a practical guide*. Cary, NC: SAS Institute; 1995.
59. Collett D. *Modeling survival data in medical research*. 2nd ed. Boca Raton, Florida: Chapman & Hall / CRC; 2003.
60. Efron B, Tibshirani RJ. *An introduction to the bootstrap (Monographs on statistics and applied probability)*. 1st ed. New York: Chapman & Hall; 1994.
61. Fitzmaurice GM, Laird NM, Ware JH. *Applied longitudinal analysis*. Hoboken, New Jersey: Wiley; 2004.
62. Hosmer DW, Lemeshow S. *Applied logistic regression*. 2nd ed. New York ; Toronto: Wiley; 2000.
63. Hosmer DW, Lemeshow S, May S. *Applied survival analysis : regression modeling of time-to-event data*. 2nd ed. Hoboken, N.J.: Wiley-Interscience; 2008.

64. Kleinbaum DG. *Applied regression analysis and other multivariable methods*. 3rd ed. Pacific Grove ; Toronto: Duxbury Press; 1998.
65. Kleinbaum DG, Klein M. *Survival analysis : a self-learning text*. 2nd ed. New York, NY: Springer; 2005.
66. Rosner B. *Fundamentals of biostatistics*. 6th ed. Belmont, CA: Thomson-Brooks/Cole; 2006.
67. Hesterberg TC, Moore DS, Monaghan S, Clipson A, Epstein R. Chapter 18: Bootstrap Methods and Permutation Tests. In: McCabe GP, Moore DS, eds. *Introduction to the Practice of Statistics*. 5th ed. New York: W.H. Freeman & Company; 2005:820.
68. Stevens LM, Madsen JC, Isselbacher EM, Khairy P, MacGillivray TE, Hilgenberg AD, Agnihotri AK. Surgical management and long-term outcomes for acute ascending aortic dissection. *J Thorac Cardiovasc Surg*. Dec 2009;138(6):1349-1357 e1341.
69. Stevens LM, Khairy P, Agnihotri AK. Coronary artery bypass grafting after recent or remote percutaneous coronary intervention in the Commonwealth of Massachusetts. *Circ Cardiovasc Interv*. Oct 2010;3(5):460-467.
70. Stevens LM, Agnihotri AK, Khairy P, Torchiana DF. Financial consequences of implementing a partner-in-care in cardiac surgery. *Ann Thorac Surg*. Sep 2010;90(3):805-812.
71. Chiappini B, Schepens M, Tan E, Dell' Amore A, Morshuis W, Dossche K, Bergonzini M, Camurri N, Reggiani LB, Marinelli G, Di Bartolomeo R. Early

- and late outcomes of acute type A aortic dissection: analysis of risk factors in 487 consecutive patients. *Eur Heart J*. Jan 2005;26(2):180-186.
72. Halstead JC, Meier M, Etz C, Spielvogel D, Bodian C, Wurm M, Shahani R, Griep RB. The fate of the distal aorta after repair of acute type A aortic dissection. *J Thorac Cardiovasc Surg*. Jan 2007;133(1):127-135.
73. Kazui T, Washiyama N, Bashar AH, Terada H, Suzuki T, Ohkura K, Yamashita K. Surgical outcome of acute type A aortic dissection: analysis of risk factors. *Ann Thorac Surg*. Jul 2002;74(1):75-81; discussion 81-72.
74. Sabik JF, Lytle BW, Blackstone EH, McCarthy PM, Loop FD, Cosgrove DM. Long-term effectiveness of operations for ascending aortic dissections. *J Thorac Cardiovasc Surg*. May 2000;119(5):946-962.
75. Bachet J, Goudot B, Dreyfus GD, Brodaty D, Dubois C, Delentdecker P, Guilmet D. Surgery for acute type A aortic dissection: the Hopital Foch experience (1977-1998). *Ann Thorac Surg*. Jun 1999;67(6):2006-2009; discussion 2014-2009.
76. David TE, Armstrong S, Ivanov J, Barnard S. Surgery for acute type A aortic dissection. *Ann Thorac Surg*. Jun 1999;67(6):1999-2001; discussion 2014-1999.
77. Arciero TJ, Jacobsen SJ, Reeder GS, Frye RL, Weston SA, Killian JM, Roger Vr VL. Temporal trends in the incidence of coronary disease. *Am J Med*. Aug 15 2004;117(4):228-233.

78. Shahian DM, Normand SL, Torchiana DF, Lewis SM, Pastore JO, Kuntz RE, Dreyer PI. Cardiac surgery report cards: comprehensive review and statistical critique. *Ann Thorac Surg.* Dec 2001;72(6):2155-2168.
79. Shahian DM, Torchiana DF, Normand SL. Implementation of a cardiac surgery report card: lessons from the Massachusetts experience. *Ann Thorac Surg.* Sep 2005;80(3):1146-1150.
80. Frye RL, August P, Brooks MM, Hardison RM, Kelsey SF, MacGregor JM, Orchard TJ, Chaitman BR, Genuth SM, Goldberg SH, Hlatky MA, Jones TL, Molitch ME, Nesto RW, Sako EY, Sobel BE. A randomized trial of therapies for type 2 diabetes and coronary artery disease. *N Engl J Med.* Jun 11 2009;360(24):2503-2515.
81. Briguori C, Condorelli G, Airoidi F, Focaccio A, D'Andrea D, Cannavale M, Abarghouei AA, Giordano S, De Vivo F, Ricciardelli B, Colombo A. Comparison of coronary drug-eluting stents versus coronary artery bypass grafting in patients with diabetes mellitus. *Am J Cardiol.* Mar 15 2007;99(6):779-784.
82. Gioia G, Matthai W, Gillin K, Dralle J, Benassi A, Gioia MF, White J. Revascularization in severe left ventricular dysfunction: Outcome comparison of drug-eluting stent implantation versus coronary artery by-pass grafting. *Catheter Cardiovasc Interv.* Jun 21 2007;70(1):26-33.
83. Seung KB, Park DW, Kim YH, Lee SW, Lee CW, Hong MK, Park SW, Yun SC, Gwon HC, Jeong MH, Jang Y, Kim HS, Kim PJ, Seong IW, Park HS, Ahn T, Chae IH, Tahk SJ, Chung WS, Park SJ. Stents versus coronary-artery

- bypass grafting for left main coronary artery disease. *N Engl J Med*. Apr 24 2008;358(17):1781-1792.
- 84.** Khattab AA, Hamm CW, Senges J, Toelg R, Geist V, Bonzel T, Kelm M, Levenson B, Neumann FJ, Nienaber CA, Pfannebecker T, Sabin G, Schneider S, Tebbe U, Richardt G. Sirolimus-Eluting Stent Treatment for Unprotected versus Protected Left Main Coronary Artery Disease in the Widespread Clinical Routine: 6-Month and 3-Year Clinical Follow-up Results From The Prospective Multi-Centre German Cypher Registry. *Heart*. May 8 2007;93(10):1251-1255.
- 85.** Serruys PW, Morice MC, Kappetein AP, Colombo A, Holmes DR, Mack MJ, Stahle E, Feldman TE, van den Brand M, Bass EJ, Van Dyck N, Leadley K, Dawkins KD, Mohr FW. Percutaneous coronary intervention versus coronary-artery bypass grafting for severe coronary artery disease. *N Engl J Med*. Mar 5 2009;360(10):961-972.
- 86.** Hannan EL, Wu C, Walford G, Culliford AT, Gold JP, Smith CR, Higgins RS, Carlson RE, Jones RH. Drug-eluting stents vs. coronary-artery bypass grafting in multivessel coronary disease. *N Engl J Med*. Jan 24 2008;358(4):331-341.
- 87.** Daemen J, Boersma E, Flather M, Booth J, Stables R, Rodriguez A, Rodriguez-Granillo G, Hueb WA, Lemos PA, Serruys PW. Long-term safety and efficacy of percutaneous coronary intervention with stenting and coronary artery bypass surgery for multivessel coronary artery disease: a

- meta-analysis with 5-year patient-level data from the ARTS, ERACI-II, MASS-II, and SoS trials. *Circulation*. Sep 9 2008;118(11):1146-1154.
- 88.** Malenka DJ, Leavitt BJ, Hearne MJ, Robb JF, Baribeau YR, Ryan TJ, Helm RE, Kellett MA, Dauerman HL, Dacey LJ, Silver MT, VerLee PN, Weldner PW, Hettleman BD, Olmstead EM, Piper WD, O'Connor GT. Comparing long-term survival of patients with multivessel coronary disease after CABG or PCI: analysis of BARI-like patients in northern New England. *Circulation*. Aug 30 2005;112(9 Suppl):I371-376.
- 89.** Mack MJ, Prince SL, Herbert M, Brown PP, Katz M, Palmer G, Edgerton JR, Eichhorn E, Magee MJ, Dewey TM. Current clinical outcomes of percutaneous coronary intervention and coronary artery bypass grafting. *Ann Thorac Surg*. Aug 2008;86(2):496-503; discussion 503.
- 90.** Yang JH, Gwon HC, Cho SJ, Hahn JY, Choi JH, Choi SH, Lee YT, Lee SH, Hong KP, Park JE. Comparison of coronary artery bypass grafting with drug-eluting stent implantation for the treatment of multivessel coronary artery disease. *Ann Thorac Surg*. Jan 2008;85(1):65-70.
- 91.** Serruys PW, Morice MC, Kappetein AP, Colombo A, Holmes DR, Mack MJ, Stahle E, Feldman TE, van den Brand M, Bass EJ, Van Dyck N, Leadley K, Dawkins KD, Mohr FW. Percutaneous coronary intervention versus coronary-artery bypass grafting for severe coronary artery disease. *The New England journal of medicine*. Mar 5 2009;360(10):961-972.
- 92.** Mohr F, Fadajet J, de Bruyne B, Holmes D, Kappetein A, Morice M, Mack M, Serruys P, Stahle E, Feldman T, Leadley K, Dawkins K, Colombo A.

Three-year follow-up of the syntax trial: Optimal revascularization strategy in patients with three-vessel disease. *J. Am. Coll. Cardiol.* 2010;56(13):Suppl. B8.

93. Barner HB, Barnett MG. Fifteen- to twenty-one-year angiographic assessment of internal thoracic artery as a bypass conduit. *Ann Thorac Surg.* Jun 1994;57(6):1526-1528.
94. Loop FD, Lytle BW, Cosgrove DM, Stewart RW, Goormastic M, Williams GW, Golding LA, Gill CC, Taylor PC, Sheldon WC, et al. Influence of the internal-mammary-artery graft on 10-year survival and other cardiac events. *N Engl J Med.* Jan 2 1986;314(1):1-6.
95. Voutilainen SM, Jarvinen AA, Verkkala KA, Keto PE, Heikkinen LO, Voutilainen PE, Harjola PT. Angiographic 20-year follow-up of 61 consecutive patients with internal thoracic artery grafts. *Ann Surg.* Jan 1999;229(1):154-158.
96. Hayward PA, Buxton BF. Contemporary coronary graft patency: 5-year observational data from a randomized trial of conduits. *Ann Thorac Surg.* Sep 2007;84(3):795-799.
97. Bourassa MG, Fisher LD, Campeau L, Gillespie MJ, McConney M, Lesperance J. Long-term fate of bypass grafts: the Coronary Artery Surgery Study (CASS) and Montreal Heart Institute experiences. *Circulation.* Dec 1985;72(6 Pt 2):V71-78.

98. CUURA103SAM, CUUSA103SAM, Medical care, Boston-Brockton-Nashua, MA-NH-ME-CT. 2007; <http://www.bls.gov/data/home.htm> Accessed July 19, 2007.
99. Wilson CT, Fisher ES, Welch HG, Siewers AE, Lucas FL. U.S. trends in CABG hospital volume: the effect of adding cardiac surgery programs. *Health Aff (Millwood)*. Jan-Feb 2007;26(1):162-168.
100. Gerber Y, Rihal CS, Sundt TM, 3rd, Killian JM, Weston SA, Thorneau TM, Roger VL. Coronary revascularization in the community. A population-based study, 1990 to 2004. *J Am Coll Cardiol*. Sep 25 2007;50(13):1223-1229.
101. Smith PK, Mayer JE, Jr., Kanter KR, DiSesa VJ, Levett JM, Wright CD, Nichols FC, 3rd, Naunheim KS. Physician payment for 2007: a description of the process by which major changes in valuation of cardiothoracic surgical procedures occurred. *Ann Thorac Surg*. Jan 2007;83(1):12-20.
102. Ferguson TB, Jr., Hammill BG, Peterson ED, DeLong ER, Grover FL. A decade of change--risk profiles and outcomes for isolated coronary artery bypass grafting procedures, 1990-1999: a report from the STS National Database Committee and the Duke Clinical Research Institute. Society of Thoracic Surgeons. *Ann Thorac Surg*. Feb 2002;73(2):480-489; discussion 489-490.
103. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41-55.

104. D'Agostino RB, Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med.* Oct 15 1998;17(19):2265-2281.
105. Cochran WG. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics.* Jun 1968;24(2):295-313.
106. Stevens LM, Carrier M, Perrault LP, Hebert Y, Cartier R, Bouchard D, Fortier A, El-Hamamsy I, Pellerin M. Single versus bilateral internal thoracic artery grafts with concomitant saphenous vein grafts for multivessel coronary artery bypass grafting: effects on mortality and event-free survival. *J Thorac Cardiovasc Surg.* May 2004;127(5):1408-1415.
107. Stevens LM, Carrier M, Perrault LP, Hebert Y, Cartier R, Bouchard D, Fortier A, Pellerin M. Influence of diabetes and bilateral internal thoracic artery grafts on long-term outcome for multivessel coronary artery bypass grafting. *Eur J Cardiothorac Surg.* Feb 2005;27(2):281-288.
108. Hirano K, Imbens GW. Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services & Outcomes Research Methodology.* 2001;2 259-278.
109. Sato T, Matsuyama Y. Marginal structural models as a tool for standardization. *Epidemiology.* Nov 2003;14(6):680-686.
110. Shah BR, Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. *J Clin Epidemiol.* Jun 2005;58(6):550-559.

111. Weitzen S, Lapane KL, Toledano AY, Hume AL, Mor V. Principles for modeling propensity scores in medical research: a systematic literature review. *Pharmacoepidemiology and drug safety*. Dec 2004;13(12):841-853.
112. Sturmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J Clin Epidemiol*. May 2006;59(5):437-447.
113. Kurth T, Walker AM, Glynn RJ, Chan KA, Gaziano JM, Berger K, Robins JM. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *American journal of epidemiology*. Feb 1 2006;163(3):262-270.
114. Stevens LM, Basmadjian AJ, Bouchard D, El-Hamamsy I, Demers P, Carrier M, Perrault LP, Cartier R, Pellerin M. Late echocardiographic and clinical outcomes after mitral valve repair for degenerative disease. *J Card Surg*. Jan-Feb 2010;25(1):9-15.
115. Turnbull BW. The empirical distribution function with arbitrarily grouped, censored and truncated data. *J R Statist Soc*. 1976;B38:290-295.
116. Weibull W. A statistical distribution of wide applicability. *J. Appl. Mech. - Trans. ASME*. 1951;18(3):293-297.
117. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assn*. 1958;53:457-481.

118. Greenwood M. The natural duration of cancer. Reports on Public Health and Medical Subjects. . *London: Her Majesty's Stationery Office.* 1926;33:1–26.
119. Blackstone EH. Overview of Procedures: Parametric Analysis of Time-Related Events : Procedures (available at <http://my.clevelandclinic.org/professionals/software/hazard/documentation.aspx>). 2001:20 p. Accessed January 28th, 2011.
120. Blackstone EH. The HAZARD Procedure (available at <http://my.clevelandclinic.org/professionals/software/hazard/documentation.aspx>). 2001:21 p. Accessed January 28th, 2011.
121. Blackstone EH. The HAZPRED Procedure (available at <http://my.clevelandclinic.org/professionals/software/hazard/documentation.aspx>). 2001:10 p. Accessed January 28th, 2011.
122. Blackstone EH, Naftel DC, Turner Jr M. The decomposition of time-varying hazard into phases, each incorporating a separate stream of concomitant information. *J Am Stat Assn.* 1986; 81(September):615-624.
123. El-Hamamsy I, Eryigit Z, Stevens LM, Sarang Z, George R, Clark L, Melina G, Takkenberg JJ, Yacoub MH. Long-term outcomes after autograft versus homograft aortic root replacement in adults with aortic valve disease: a randomised controlled trial. *Lancet.* Aug 14 2010;376(9740):524-531.
124. El-Hamamsy I, Clark L, Stevens LM, Sarang Z, Melina G, Takkenberg JJ, Yacoub MH. Late outcomes following freestyle versus homograft aortic root replacement: results from a prospective randomized trial. *J Am Coll Cardiol.* Jan 26 2010;55(4):368-376.

125. Cox DR. Regression Models and Life Tables. *Journal of the Royal Statistical Society*. 1972;Series B 34(2):187–220.
126. Cox DR, Oakes D. *Analysis of survival data*. London ; New York: Chapman and Hall; 1984.
127. Harrell F, Lee K. Verifying assumptions of the Cox proportional hazards model. *Proceedings of the Eleventh Annual SAS User's Group International*. 1986:823-828.
128. Schoenfeld D. Partial residuals for the proportional hazards model. *Biometrika*. 1982;69:51-55.
129. Blackstone EH. Born of necessity: the dynamic synergism between advancement of analytic methods and generation of new knowledge. *J Heart Valve Dis*. Jul 1995;4(4):326-336.
130. Blackstone EH, Kirklin JW. Death and other time-related events after valve replacement. *Circulation*. Oct 1985;72(4):753-767.
131. Sabik JF, 3rd, Blackstone EH, Gillinov AM, Smedira NG, Lytle BW. Occurrence and risk factors for reintervention after coronary artery bypass grafting. *Circulation*. Jul 4 2006;114(1 Suppl):I454-460.
132. Kirklin JW, Kouchoukos NT. *Kirklin/Barratt-Boyes cardiac surgery : morphology, diagnostic criteria, natural history, techniques, results, and indications*. 3rd ed. Philadelphia, Pa.: Churchill Livingstone; 2003.
133. Efron B. Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*. 1981;68:589-599.

134. Efron B, Tibshirani RJ. Improvements on Cross-Validation: The .632+ Bootstrap Method. *Journal of the American Statistical Association*. 1997;92(June):548-560.
135. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J Clin Epidemiol*. May 2003;56(5):441-447.
136. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. Aug 2001;54(8):774-781.
137. Raspe R. *The Singular Adventures of Baron Munchausen*. 1786.
138. Shapira OM, Stevens LM, Korach A, Benjamin EJ, Lovett A, Silverstien T, Agnihotri AK, Torchiana DF, Shemin RJ, Normand SLT. Abstract 3999: The Impact of Race and Insurance Status on Utilization Rates and Outcomes after Coronary Artery Bypass Grafting Surgery in the Commonwealth of Massachusetts. Presented at the American Heart Association Scientific Sessions - Chicago (IL) *Circulation*. 2006;114:II_860.
139. Shapira OM, Stevens LM, Korach A, Benjamin EJ, Lovett A, Silverstien T, Agnihotri AK, Torchiana DF, Shemin RJ, A J, Normand SLT. Abstract 1012-127: The Impact of Race and Insurance Status on Utilization Rates and Outcomes after Percutaneous Coronary Interventions in the Commonwealth of Massachusetts. Presented at the American College of Cardiology 57th Annual Scientific Session - Chicago (IL) *JACC*. 2008;51(10-Suppl 1):A242.

140. El-Hamamsy I, Zaki M, Stevens LM, Clark LA, Rubens M, Melina G, Yacoub MH. Rate of progression and functional significance of aortic root calcification after homograft versus freestyle aortic root replacement. *Circulation*. Sep 15 2009;120(11 Suppl):S269-275.
141. Nguyen BK, Maltais S, Perrault LP, Tanguay JF, Tardif JC, Stevens LM, Borie M, Harel F, Mansour S, Noiseux N. Improved function and myocardial repair of infarcted heart by intracoronary injection of mesenchymal stem cell-derived growth factors. *J Cardiovasc Transl Res*. Oct 2010;3(5):547-558.
142. Waisbren EC, Stevens LM, Avery EG, Picard MH, Vlahakes GJ, Agnihotri AK. Changes in mitral regurgitation after replacement of the stenotic aortic valve. *Ann Thorac Surg*. Jul 2008;86(1):56-62.
143. El-Hamamsy I, Balachandran K, Yacoub MH, Stevens LM, Sarathchandra P, Taylor PM, Yoganathan AP, Chester AH. Endothelium-dependent regulation of the mechanical properties of aortic valve cusps. *J Am Coll Cardiol*. Apr 21 2009;53(16):1448-1455.
144. Center for Disease Control and Prevention / National Center for Health Statistics / National Death Index. <http://www.cdc.gov/nchs/ndi.htm>, 2009.
145. National Center for Health Statistics. *U.S. decennial life tables for 1989–91, vol II, State life tables no. 22, Massachusetts*. Hyattsville, Maryland 1998.
146. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41-55.
147. Aitchison J, Brown JAC. *The lognormal distribution, with special reference to its uses in economics*. Cambridge [Eng.]: University Press; 1957.

148. Briggs A, Nixon R, Dixon S, Thompson S. Parametric modelling of cost data: some simulation evidence. *Health Econ.* Apr 2005;14(4):421-428.
149. Nixon RM, Thompson SG. Parametric modelling of cost data in medical studies. *Stat Med.* Apr 30 2004;23(8):1311-1331.
150. O'Hagan A, Stevens JW. Assessing and comparing costs: how robust are the bootstrap and methods based on asymptotic normality? *Health Econ.* Jan 2003;12(1):33-49.
151. Briggs A, Gray A. The distribution of health care costs and their statistical analysis for economic evaluation. *J Health Serv Res Policy.* Oct 1998;3(4):233-245.
152. Gallo A, Davies RR, Coe MP, Elefteriades JA, Coady MA. Indications, timing, and prognosis of operative repair of aortic dissections. *Semin Thorac Cardiovasc Surg.* Fall 2005;17(3):224-235.
153. DeBakey ME, Henly WS, Cooley DA, Morris GC, Jr., Crawford ES, Beall AC, Jr. Surgical Management of Dissecting Aneurysms of the Aorta. *J Thorac Cardiovasc Surg.* Jan 1965;49:130-149.
154. Apaydin AZ, Buket S, Posacioglu H, Islamoglu F, Calkavur T, Yagdi T, Ozbaran M, Yuksel M. Perioperative risk factors for mortality in patients with acute type A aortic dissection. *Ann Thorac Surg.* Dec 2002;74(6):2034-2039; discussion 2039.
155. Ehrlich MP, Grabenwoger M, Kilo J, Kocher AA, Grubhofer G, Lassnig AM, Tschernko EM, Schlechta B, Hutschala D, Domanovits H, Sodeck G, Wolner

- E. Surgical treatment of acute type A dissection: is rupture a risk factor? *Ann Thorac Surg*. Jun 2002;73(6):1843-1848.
- 156.** Lai DT, Miller DC, Mitchell RS, Oyer PE, Moore KA, Robbins RC, Shumway NE, Reitz BA. Acute type A aortic dissection complicated by aortic regurgitation: composite valve graft versus separate valve graft versus conservative valve repair. *J Thorac Cardiovasc Surg*. Dec 2003;126(6):1978-1986.
- 157.** Mehta RH, Suzuki T, Hagan PG, Bossone E, Gilon D, Llovet A, Maroto LC, Cooper JV, Smith DE, Armstrong WF, Nienaber CA, Eagle KA. Predicting death in patients with acute type a aortic dissection. *Circulation*. Jan 15 2002;105(2):200-206.
- 158.** Rampoldi V, Trimarchi S, Eagle KA, Nienaber CA, Oh JK, Bossone E, Myrmel T, Sangiorgi GM, De Vincentiis C, Cooper JV, Fang J, Smith D, Tsai T, Raghupathy A, Fattori R, Sechtem U, Deeb MG, Sundt TM, 3rd, Isselbacher EM. Simple risk models to predict surgical mortality in acute type A aortic dissection: the International Registry of Acute Aortic Dissection score. *Ann Thorac Surg*. Jan 2007;83(1):55-61.
- 159.** Trimarchi S, Nienaber CA, Rampoldi V, Myrmel T, Suzuki T, Mehta RH, Bossone E, Cooper JV, Smith DE, Menicanti L, Frigiola A, Oh JK, Deeb MG, Isselbacher EM, Eagle KA. Contemporary results of surgery in acute type A aortic dissection: The International Registry of Acute Aortic Dissection experience. *J Thorac Cardiovasc Surg*. Jan 2005;129(1):112-122.

- 160.** Westaby S, Saito S, Katsumata T. Acute type A dissection: conservative methods provide consistently low mortality. *Ann Thorac Surg.* Mar 2002;73(3):707-713.
- 161.** Tsai TT, Evangelista A, Nienaber CA, Trimarchi S, Sechtem U, Fattori R, Myrmet T, Pape L, Cooper JV, Smith DE, Fang J, Isselbacher E, Eagle KA. Long-term survival in patients presenting with type A acute aortic dissection: insights from the International Registry of Acute Aortic Dissection (IRAD). *Circulation.* Jul 4 2006;114(1 Suppl):I350-356.
- 162.** Kallenbach K, Oelze T, Salcher R, Hagl C, Karck M, Leyh RG, Haverich A. Evolving strategies for treatment of acute aortic dissection type A. *Circulation.* Sep 14 2004;110(11 Suppl 1):II243-249.
- 163.** Bavaria JE, Brinster DR, Gorman RC, Woo YJ, Gleason T, Pochettino A. Advances in the treatment of acute type A dissection: an integrated approach. *Ann Thorac Surg.* Nov 2002;74(5):S1848-1852; discussion S1857-1863.
- 164.** Hagan PG, Nienaber CA, Isselbacher EM, Bruckman D, Karavite DJ, Russman PL, Evangelista A, Fattori R, Suzuki T, Oh JK, Moore AG, Malouf JF, Pape LA, Gaca C, Sechtem U, Lenferink S, Deutsch HJ, Diedrichs H, Marcos y Robles J, Llovet A, Gilon D, Das SK, Armstrong WF, Deeb GM, Eagle KA. The International Registry of Acute Aortic Dissection (IRAD): new insights into an old disease. *Jama.* Feb 16 2000;283(7):897-903.
- 165.** Crawford ES, Kirklin JW, Naftel DC, Svensson LG, Coselli JS, Safi HJ. Surgery for acute dissection of ascending aorta. Should the arch be included? *J Thorac Cardiovasc Surg.* Jul 1992;104(1):46-59.

166. Fann JI, Smith JA, Miller DC, Mitchell RS, Moore KA, Grunkemeier G, Stinson EB, Oyer PE, Reitz BA, Shumway NE. Surgical management of aortic dissection during a 30-year period. *Circulation*. Nov 1 1995;92(9 Suppl):III113-121.
167. Hilgenberg AD, Akins CW, Logan DL, Vlahakes GJ, Buckley MJ, Madsen JC, Torchiana DF. Composite aortic root replacement with direct coronary artery implantation. *Ann Thorac Surg*. Oct 1996;62(4):1090-1095.
168. De Paepe A, Devereux RB, Dietz HC, Hennekam RC, Pyeritz RE. Revised diagnostic criteria for the Marfan syndrome. *Am J Med Genet*. Apr 24 1996;62(4):417-426.
169. Caus T, Frapier JM, Giorgi R, Aymard T, Riberi A, Albat B, Chaptal PA, Mesana T. Clinical outcome after repair of acute type A dissection in patients over 70 years-old. *Eur J Cardiothorac Surg*. Aug 2002;22(2):211-217.
170. Mehta RH, O'Gara PT, Bossone E, Nienaber CA, Myrmel T, Cooper JV, Smith DE, Armstrong WF, Isselbacher EM, Pape LA, Eagle KA, Gilon D. Acute type A aortic dissection in the elderly: clinical characteristics, management, and outcomes in the current era. *J Am Coll Cardiol*. Aug 21 2002;40(4):685-692.
171. Beckman JA, Mehta RH, Isselbacher EM, Bossone E, Cooper JV, Smith DE, Fang J, Sechtem U, Pape LA, Myrmel T, Nienaber CA, Eagle KA, O'Gara PT. Branch vessel complications are increased in aortic dissection patients with renal insufficiency. *Vasc Med*. Nov 2004;9(4):267-270.

172. Scholl FG, Coady MA, Davies R, Rizzo JA, Hammond GL, Kopf GS, Elefteriades JA. Interval or permanent nonoperative management of acute type A aortic dissection. *Arch Surg*. Apr 1999;134(4):402-405; discussion 405-406.
173. Slonim SM, Miller DC, Mitchell RS, Semba CP, Razavi MK, Dake MD. Percutaneous balloon fenestration and stenting for life-threatening ischemic complications in patients with acute aortic dissection. *J Thorac Cardiovasc Surg*. Jun 1999;117(6):1118-1126.
174. Moore AG, Eagle KA, Bruckman D, Moon BS, Malouf JF, Fattori R, Evangelista A, Isselbacher EM, Suzuki T, Nienaber CA, Gilon D, Oh JK. Choice of computed tomography, transesophageal echocardiography, magnetic resonance imaging, and aortography in acute aortic dissection: International Registry of Acute Aortic Dissection (IRAD). *Am J Cardiol*. May 15 2002;89(10):1235-1238.
175. Shiga T, Wajima Z, Apfel CC, Inoue T, Ohe Y. Diagnostic accuracy of transesophageal echocardiography, helical computed tomography, and magnetic resonance imaging for suspected thoracic aortic dissection: systematic review and meta-analysis. *Arch Intern Med*. Jul 10 2006;166(13):1350-1356.
176. Bartram U, Wahlers T, Aebert H, Muegge H. Coronary artery bypass grafting after failed percutaneous angioplasty compared to direct coronary bypass grafting in patients with unstable angina. *Thorac Cardiovasc Surg*. Feb 1996;44(1):31-34.

177. Ladowski JS, Dillon TA, Deschner WP, DeRiso AJ, 2nd, Peterson AC, Schatzlein MH. Durability of emergency coronary artery bypass for complications of failed angioplasty. *Cardiovasc Surg*. Feb 1996;4(1):23-27.
178. Berger PB, Stensrud PE, Daly RC, Grill D, Bell MR, Garratt KN, Holmes DR, Jr. Time to reperfusion and other procedural characteristics of emergency coronary artery bypass surgery after unsuccessful coronary angioplasty. *Am J Cardiol*. Sep 15 1995;76(8):565-569.
179. Nollert G, Amend J, Detter C, Reichart B. Coronary artery bypass grafting after failed coronary angioplasty: risk factors and long-term results. *Thorac Cardiovasc Surg*. Feb 1995;43(1):35-39.
180. Tuzcu M, Simpfendorfer C, Dorosti K, Franco I, Golding L, Hollman J, Whitlow P. Long-term outcome of unsuccessful percutaneous transluminal coronary angioplasty. *Am Heart J*. Apr 1990;119(4):791-796.
181. Chocron S, Baillot R, Rouleau JL, Warnica WJ, Block P, Johnstone D, Myers MG, Calciu CD, Nozza A, Martineau P, van Gilst WH. Impact of previous percutaneous transluminal coronary angioplasty and/or stenting revascularization on outcomes after surgical revascularization: insights from the imagine study. *Eur Heart J*. Mar 2008;29(5):673-679.
182. Barakate MS, Hemli JM, Hughes CF, Bannon PG, Horton MD. Coronary artery bypass grafting (CABG) after initially successful percutaneous transluminal coronary angioplasty (PTCA): a review of 17 years experience. *Eur J Cardiothorac Surg*. Feb 2003;23(2):179-186.

183. Kalaycioglu S, Sinci V, Oktar L. Coronary artery bypass grafting (CABG) after successful percutaneous transluminal coronary angioplasty (PTCA): is PTCA a risk for CABG? *Int Surg*. Jul-Sep 1998;83(3):190-193.
184. Society of Thoracic Surgeons Adult Cardiac Database - Specifications For Vendors And PGS Sites - Version 2.52.1. April 9, 2004; http://www.ctsnet.org/file/rptDataSpecifications252_1_ForVendorsPGS.pdf
Accessed July 2nd, 2007.
185. Parsons L. Performing a 1:N Case-Control Match on Propensity Score. Paper presented at: Proceedings of the Twenty-Nine Annual SAS Users Group International Conference 2004; Cary, NC.
186. Hassan A, Buth KJ, Baskett RJ, Ali IS, Maitland A, Sullivan JA, Ghali WA, Hirsch GM. The association between prior percutaneous coronary intervention and short-term outcomes after coronary artery bypass grafting. *Am Heart J*. Nov 2005;150(5):1026-1031.
187. Thielmann M, Neuhauser M, Knipp S, Kottenberg-Assemacher E, Marr A, Pizanis N, Hartmann M, Kamler M, Massoudy P, Jakob H. Prognostic impact of previous percutaneous coronary intervention in patients with diabetes mellitus and triple-vessel disease undergoing coronary artery bypass surgery. *J Thorac Cardiovasc Surg*. Aug 2007;134(2):470-476.
188. Thielmann M, Leyh R, Massoudy P, Neuhauser M, Aleksic I, Kamler M, Herold U, Piotrowski J, Jakob H. Prognostic significance of multiple previous percutaneous coronary interventions in patients undergoing elective coronary artery bypass surgery. *Circulation*. Jul 4 2006;114(1 Suppl):I441-447.

189. Boylan MJ, Lytle BW, Loop FD, Taylor PC, Borsh JA, Goormastic M, Cosgrove DM. Surgical treatment of isolated left anterior descending coronary stenosis. Comparison of left internal mammary artery and venous autograft at 18 to 20 years of follow-up. *J Thorac Cardiovasc Surg.* Mar 1994;107(3):657-662.
190. Alderman EL, Kip KE, Whitlow PL, Bashore T, Fortin D, Bourassa MG, Lesperance J, Schwartz L, Stadius M. Native coronary disease progression exceeds failed revascularization as cause of angina after five years in the Bypass Angioplasty Revascularization Investigation (BARI). *J Am Coll Cardiol.* Aug 18 2004;44(4):766-774.
191. Smith SC, Jr., Feldman TE, Hirshfeld JW, Jr., Jacobs AK, Kern MJ, King SB, 3rd, Morrison DA, O'Neil WW, Schaff HV, Whitlow PL, Williams DO, Antman EM, Adams CD, Anderson JL, Faxon DP, Fuster V, Halperin JL, Hiratzka LF, Hunt SA, Nishimura R, Ornato JP, Page RL, Riegel B. ACC/AHA/SCAI 2005 guideline update for percutaneous coronary intervention: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (ACC/AHA/SCAI Writing Committee to Update 2001 Guidelines for Percutaneous Coronary Intervention). *Circulation.* Feb 21 2006;113(7):e166-286.
192. Long KH, McMurtry EK, Lennon RJ, Chapman AC, Singh M, Rihal CS, Wood DL, Holmes DR, Jr., Ting HH. Elective percutaneous coronary intervention without on-site cardiac surgery: clinical and economic implications. *Med Care.* May 2006;44(5):406-413.

193. Singh M, Ting HH, Berger PB, Garratt KN, Holmes DR, Jr., Gersh BJ. Rationale for on-site cardiac surgery for primary angioplasty: a time for reappraisal. *J Am Coll Cardiol*. Jun 19 2002;39(12):1881-1889.
194. Teisberg EO, Wallace S. Creating a high-value delivery system for health care. *Semin Thorac Cardiovasc Surg*. Spring 2009;21(1):35-42.
195. Society of Thoracic Surgeons Adult Cardiac Database - Specifications For Vendors And PGS Sites - Version 2.52.1. April 9, 2004; http://www.ctsnet.org/file/rptDataSpecifications252_1_ForVendorsPGS.pdf Accessed March 14, 2008.
196. Duan N. Smearing estimate: a nonparametric retransformation method. *J Am Stat Assoc*. 1983;78(383):605-610.
197. Pennington DG. The impact of new technology on cardiothoracic surgical practice. *Ann Thorac Surg*. Jan 2006;81(1):10-18.
198. Bonchek LI, Harley DP, Wilbur RH, Rich JB, Bonchek MS, Erenhouse R, Mayer JE. The STS future planning conference for adult cardiac surgery. *Ann Thorac Surg*. Dec 2003;76(6):2156-2166.
199. Nallamothu BK, Rogers MA, Chernew ME, Krumholz HM, Eagle KA, Birkmeyer JD. Opening of specialty cardiac hospitals and use of coronary revascularization in medicare beneficiaries. *Jama*. Mar 7 2007;297(9):962-968.
200. Omoigui NA, Miller DP, Brown KJ, Annan K, Cosgrove D, 3rd, Lytle B, Loop F, Topol EJ. Outmigration for coronary bypass surgery in an era of

- public dissemination of clinical outcomes. *Circulation*. Jan 1 1996;93(1):27-33.
- 201.** Shahian DM, Normand SL. The volume-outcome relationship: from Luft to Leapfrog. *Ann Thorac Surg*. Mar 2003;75(3):1048-1058.
- 202.** Vaughan-Sarrazin MS, Hannan EL, Gormley CJ, Rosenthal GE. Mortality in Medicare beneficiaries following coronary artery bypass graft surgery in states with and without certificate of need regulation. *Jama*. Oct 16 2002;288(15):1859-1866.
- 203.** DiSesa VJ, O'Brien SM, Welke KF, Beland SM, Haan CK, Vaughan-Sarrazin MS, Peterson ED. Contemporary impact of state certificate-of-need regulations for cardiac surgery: an analysis using the Society of Thoracic Surgeons' National Cardiac Surgery Database. *Circulation*. Nov 14 2006;114(20):2122-2129.
- 204.** Marshall G, Grover FL, Henderson WG, Hammermeister KE. Assessment of predictive models for binary outcomes: an empirical approach using operative death from cardiac surgery. *Stat Med*. Aug 15 1994;13(15):1501-1511.
- 205.** Harrell FE, Jr., Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. *Stat Med*. Apr-Jun 1984;3(2):143-152.
- 206.** Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. Apr 1982;143(1):29-36.
- 207.** Anderson RN. Method for constructing complete annual U.S. life tables. *National Center for Health Statistics. Vital and health statistics. Data*

- evaluation and methods research*. 1999;2(129):DHHS publication no. (PHS) 99-1329.
- 208.** Miettinen OS. Individual matching with multiple controls in the case of all-or-none responses. *Biometrics*. Jun 1969;25(2):339-355.
- 209.** Miettinen OS. Matching and design efficiency in retrospective studies. *Am J Epidemiol*. Feb 1970;91(2):111-118.
- 210.** Sturmer T, Brenner H. Degree of matching and gain in power and efficiency in case-control studies. *Epidemiology*. Jan 2001;12(1):101-108.
- 211.** Raboud J, Breslow NE. Efficiency gains from the addition of controls to matched sets in cohort studies. *Stat Med*. Aug 1989;8(8):977-985.
- 212.** Walter SD. Matched case-control studies with a variable number of controls per case. *Applied Statistics*. 1980;29:172-179.
- 213.** Austin PC, Grootendorst P, Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Stat Med*. Feb 20 2007;26(4):734-753.
- 214.** Austin PC, Mamdani MM, Stukel TA, Anderson GM, Tu JV. The use of the propensity score for estimating treatment effects: administrative versus clinical data. *Stat Med*. May 30 2005;24(10):1563-1578.
- 215.** Imai K, King G, Stuart EA. Misunderstandings among Experimentalists and Observationalists about Causal Inference. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*. 2008;171(2):481-502.

216. Ho D, Imai K, King G, Stuart E. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*. 2007;15:199–236.
217. Flury B, Reidwyl H. Standard distance in univariate and multivariate analysis. *The American Statistician*. 1986;40:249–251.
218. Austin PC. Using the Standardized Difference to Compare the Prevalence of a Binary Variable Between Two Groups in Observational Research. *Communications in Statistics - Simulation and Computation*. 2009;38(6):1228-1234.
219. Agresti A, Min Y. Effects and non-effects of paired identical observations in comparing proportions with binary matched-pairs data. *Stat Med*. Jan 15 2004;23(1):65-75.
220. Klein JP, Moeschberger ML. *Survival analysis: techniques for censored and truncated data*. New York: Springer-Verlag; 1997.
221. Therneau TM, Gamsch PM. *Modeling survival data: extending the Cox model*. New York: Springer-Verlag; 2000.
222. Austin PC, Mamdani MM. A comparison of propensity score methods: a case-study estimating the effectiveness of post-AMI statin use. *Stat Med*. Jun 30 2006;25(12):2084-2106.
223. Limpert E, Stahel WA, Abbt M. Log-normal Distributions across the Sciences: Keys and Clues. *BioScience*. 2001;51(5):341-352.
224. Wang P, MyiLibrary Ltd. *Financial econometrics [ressource électronique]*. 2nd ed. London ; New York: Routledge; 2009.

- 225.** Pasta DJ, Cisternas MG. Estimating Standard Errors for CLASS Variables in Generalized Linear Models Using PROC IML. Paper presented at: Proceedings of the Twenty-Eight Annual SAS Users Group International Conference 2003; Cary, NC.
- 226.** Mills JL. Data torturing. *N Engl J Med.* Oct 14 1993;329(16):1196-1199.

APPENDICE A

FORMULAIRE DE DÉROGATION

Le 28 avril 2006

Monsieur Louis-Mathieu Stevens
5863, Desaulniers
Montréal QC H1N 3P9

Monsieur,

Suite à votre demande, la Faculté des études supérieures vous accorde une dérogation pour vous permettre d'être inscrit dans deux programmes en même temps.

Vous pourrez ainsi poursuivre vos travaux de recherche au doctorat dans le programme de Sciences biomédicales tout en effectuant un fellowship au Harvard School of Public Health et au Massachussetts General Hospital.

Nous vous souhaitons un franc succès dans vos études.

Veillez agréer, monsieur, mes sincères salutations.

Le vice-doyen,



André Ferron

APPENDICE B

AUTORISATION DE SOUMETTRE LA THÈSE PAR ARTICLES

AUTORISATION DE RÉDIGER LA THÈSE PAR ARTICLES

From: Fauteux France [REDACTED]
Sent: Thursday, February 03, 2011 2:17 PM
To: [REDACTED]
Cc: [REDACTED]

Subject: **Autorisation de rédiger par articles**

Bonjour,

Suite à votre demande, j'ai le plaisir de vous annoncer que vous êtes autorisé à présenter votre thèse sous forme d'articles. Il est entendu que vous devrez vous soumettre aux conditions minimales de dépôt décrites dans le « Guide de présentation des mémoires de maîtrise et des thèses de doctorat ». Ce document est disponible sur le site de la FES. Vous pouvez également vous le procurer à la Librairie de l'Université de Montréal.

Cependant, afin de respecter la loi canadienne sur les droits d'auteurs, vous devrez, au moment du dépôt de votre mémoire ou avant, remettre, avec les formulaires de reproduction et diffusion dûment complétés et signés, les déclarations écrites de tous les coauteurs des articles inclus dans votre thèse autorisant la reproduction et la diffusion de votre thèse de doctorat.

Merci et n'hésitez pas à communiquer avec moi pour toute autre question.

Meilleures salutations,

France Fauteux -
Gestion des dossiers étudiants
Programme de sciences biomédicales, Ph.D.

Faculté de médecine - Direction
Études supérieures
Pavillon Roger-Gaudry, bureau 707
2900 boul. Édouard-Montpetit
Montréal Qc H3T 1J4
Téléphone: 514-343-6111 poste 1-4134
Télécopieur : 514-343-5751

APPENDICE C

FORMULAIRES DE CONSENTEMENT DES CO-AUTEURS

***SECTION ÉLIMINÉE EN RAISON DES EXIGENCES DE
FORMAT DE THÈSE ÉLECTRONIQUE (PAS DE SIGNATURES
NI DE RENSEIGNEMENTS PERSONNELS)***

APPENDICE D

FORMULAIRES DE PERMISSION DES JOURNAUX

***SECTION ÉLIMINÉE EN RAISON DES EXIGENCES DE
FORMAT DE THÈSE ÉLECTRONIQUE (PAS DE SIGNATURES
NI DE RENSEIGNEMENTS PERSONNELS)***

APPENDICE E

CURRICULUM VITAE

CURRICULUM VITAE

LOUIS-MATHIEU STEVENS, MD, PhD, FRCSC

Lieu de naissance : Sherbrooke, Québec, Canada
 Lieu de pratique : Centre Hospitalier Universitaire de Montréal (CHUM)
 Pavillon Hôtel-Dieu de Montréal
 3840, rue St-Urbain
 Montréal (Québec) H2W 1T8
 Téléphones : Bureau : (514) 890-8131 ou (514) 890-8000 #14308
 Télécopieur : (514) 412-7231
 Courriel : [REDACTED]

CHEMINEMENT PROFESSIONNEL

2009- Chirurgien cardiaque, CHUM, Québec
 2009- Professeur adjoint de clinique, Université de Montréal, Québec
 2009- Chercheur au sein de l' Axe Risque à la Santé du Centre de Recherche du CHUM (CRCHUM), Québec
 2011- Chercheurs-boursier Junior 1 du Fonds de la Recherche en Santé du Québec (FRSQ)

FORMATION ET DIPLÔMES

1996-2000 **MD**, Université Laval, Québec
 2000-2006 **Résidence** en chirurgie cardiaque, Université de Montréal, Québec
 2001-2002 **MSc** (sciences biomédicales), Université de Montréal, Québec
 Étude effet tétrahydrobioptérine sur dysfonction endothéliale artères coronaires induite par circulation extracorporelle dans modèle porcin (directeur : Dr. L. P. Perrault, MD, PhD).
 2007-2008 **SM** (Master of Science – Epidemiology), Harvard School of Public Health, Harvard University, Boston, Massachusetts, U.S.A. (superviseur: Dr. M. Mittleman, PhD).
 2007 Program in Clinical Effectiveness, Harvard School of Public Health, Harvard University, Boston, Massachusetts, U.S.A.
 2007-2012 **PhD** (sciences biomédicales), Université de Montréal, Québec.
 Statistiques appliquées en chirurgie cardiaque adulte : Analyses de survie et applications du “*propensity score*” (directeur : Dr. P. Khairy, MD PhD).

FORMATION COMPLÉMENTAIRE OU POST-DOCTORALE

- 2006-2008 Recherche. Stage de formation complémentaire, à la *Harvard Medical School* et au *Massachusetts General Hospital, Harvard University*, Boston, Massachusetts, U.S.A.
 Stage de recherche clinique (*fellowship* académique) avec Dr. A.K. Agnihotri, MD, Dr. S.L.T. Normand, PhD, Dr. D.F. Torchiana, MD, Dr. P. Khairy, MD, PhD
- 2008-2009 Clinique. Stage de formation complémentaire, à la *Brody School of Medicine, East Carolina University*, Greenville, North Carolina, U.S.A.
 Stage de formation clinique et de recherche clinique en chirurgie cardiaque minimalement invasive et robotique avec Dr. W.R. Chitwood, MD.

BOURSES DE FORMATION OBTENUES

- 2006-2008 Bourse de formation post-diplôme professionnel du Fonds de la Recherche en Santé du Québec (**FRSQ**).
(décliné pour la bourse de recherche du ISRC/CIHR)
- 2006-2009 Bourse de formation post-diplôme professionnel « Initiative de Recherche Clinique » des Instituts de Recherche en Santé du Canada (**IRSC/CIHR**).
- 2006-2008 Bourse du Fonds *Rosetti, Massachusetts General Hospital*, Boston, Massachusetts, U.S.A.
- 2007-2008 Bourse de CRICO/RMF (compagnie d'assurance-responsabilité médicale), Boston, Massachusetts, U.S.A.
 Coordonnateur principal pour le projet de recherche clinique prospectif en simulation.
- 2007-2009 Bourse de *Partners High Performance Medicine*, Boston, Massachusetts, U.S.A.
 Expert biostatistique.

PUBLICATIONS

Articles : 35 articles avec comités de pairs

Parutions : 2 chapitres de livre

Abrégés : 45 publiés