

REES France

Réseau d'Évaluation en Économie de la Santé

PROTOCOLE EN REANIMATION D'ÉVALUATION MEDICO- ECONOMIQUE D'UNE INNOVATION DANS LE SEPSIS SEVERE : PREMISS

*Evaluation médico-économique du traitement des états septiques graves par la drotrécogine
alfa*

Réduction des biais de recrutement par Propensity Score

25 avril 2005

PRESENTATION DE LA SOCIETE

Le Réseau d'Évaluation en Économie de la Santé, **REES**, est un bureau d'études créé par des économistes, docteurs en pharmacie, statisticiens et informaticiens. Notre objectif est de **R**assembler les **I**nformations **C**liniques, **H**umaines, **É**conomiques et **S**ociales en **S**anté dans le cadre du modèle de création de "**RICHESS**" dont nous sommes les promoteurs. Il s'agit d'intensifier les liaisons entre la recherche universitaire et les cliniciens en mettant à la disposition des professionnels de santé, des techniques innovantes qui leur permettent de mieux maîtriser leur environnement.

Son domaine d'activité privilégié : le choix du bon design d'étude, la construction de questionnaires informatiques Intranet (Étude PREMISS, Étude HER.ME.S), la création de CD-ROM d'impact budgétaire, l'analyse de bases de données hospitalières ou ambulatoires de grande dimension sous SAS, le développement de simulations mathématiques stochastiques ou déterministes et les études de qualité de vie.

- Enquêtes observationnelles et études quasi-expérimentales
- CD Rom d'impact budgétaire du médicament dans les services cliniques
- Construction de questionnaires informatiques et de dossiers médicaux partagés
- Création et validation d'échelle de qualité de vie spécifique d'une pathologie
- Analyse statistique sous SAS des bases de données
- Analyse décisionnelle, modèle de MARKOV, étude de sensibilité probabiliste et bootstrap
- Analyse d'impact des réseaux de soins coordonnés

REES a publié plus d'une centaine d'articles dans des revues à Comité de Lecture et réalisé plus de trente études évaluatives en appliquant ces techniques.

SON EQUIPE

- ☞ **Robert LAUNOIS**, Professeur des Facultés de Sciences Economiques (Université de Paris 13), diplômé de l'Institut Politique de Paris, Harkness Fellow, Harvard University.
- ☞ **Katell LE LAY**, Chef de Projets, DESS Traitement de l'Information Médicale et Hospitalière, Maîtrise Biochimie (Université de Rennes I)
- ☞ **Lionel RIOU FRANÇA**, Chef de Projets, DEA de biostatistique, ENSA Montpellier 1, MST de Santé Publique (Université de Bordeaux 2)
- ☞ **Stéphanie PAYET**, Chargée de Projets, Ingénieur ENSAI, filière économie de la santé
- ☞ **Gérard PRESENTE**, Ingénieur informaticien
- ☞ **Marylène SARTOR**, Assistante de direction.

RECHERCHES & TRAVAUX EN COURS

Evaluation du Réseau Asthme RESALIS (1999-2002) CPAM. Evreux - Alliance Médica ; Evaluation des Réseaux Groupama (1999-2001) ; Evaluation des comportements de prescriptions des génériques en France (2000-2001) Direction de la Sécurité Sociale (DSS) ; Programme de soutien aux Innovations diagnostiques et thérapeutiques coûteuses – Direction des Hôpitaux et de l'organisation des Soins (DHOS) ; Programme HER.ME.S (AP-HP, CRLCC, 12 services d'oncologie) ; Programme PREMISS (SFAR, SRLF, 118 services de réanimation) ; Construction et validation d'une échelle spécifique de qualité de vie dans le lymphœdème des membres supérieurs après cancer du sein.

NOUVELLES PUBLICATIONS REES France

- Launois R., Giroud M., Mégnigbêto AC., Durand I., Gaudin A.F., M.H. Mahagne "Estimating the cost-effectiveness of stroke units in France, compared with conventional care". Stroke 2004. Stroke 2004; 35:770-775.
- Launois R. "A quality of life tool kit in venous disorders". Medicographia 2004; 26(2): 1-7.
- P. Tilleul, G. Perrocheau, K. Le Lay, A. Mégnigbêto, A. Bellanger, A.C. Joly, C. Dernaroutunian, J.F. Latour, G. Guidicelli, C. Ollivier, J.M. Canonge, I. Mogenet, E. de Saint-Fuscien, X. Simoens, R. Launois "Le microcosting est-il utile dans les études pharmacoéconomiques, Application au traitement du cancer colorectal métastatique " J Pharm Clin. 2003; 22, 4: 175-80.
- R. Launois, S. Payet "Evaluation pharmacoéconomique des épisodes dépressifs majeurs et des récidives " Neuropsych News 2003. 2 (5): 209-221.
- Launois R. " L'évaluation médico-économique en psychiatrie ". Encycl Méd Chir (Editions scientifiques et médicales Elsevier SAS, Paris), Psychiatrie, 37-860-A-15, 2003, 12 p.
- Launois R. " Notions et mesure des coûts en fonction de la perspective choisie ". Bulletin du cancer. 2003 ; 90 (11) : 946-954
- Launois R. " L'évaluation socio-économique : une discipline entre recherche clinique et études observationnelles ". Bulletin du cancer 2003; 90 (1) : 97-104.
- Launois R., Mégnigbêto AC., Perez V., Roy M., Camus A., Gailhardou M., Lançon F., Quéniart M. " Expérimentation de la Prise en Charge de l'Asthme en Réseau de Soins Coordinés® : Bilan de l'Expérience RESALIS dans l'Eure, 18 mois Avant et 12 mois Après Intervention " - Journal d' Economie Médicale 2002; 20 (5) : 333-352.
- Launois R., Mégnigbêto AC., Pocquet K., Alliot F. "A specific Quality of Life Scale in Upper Limb Lymphoedema : the ULL- 27 questionnaire". In C. Campisi, MH Witte, CL Witte (Ed) Progress in Lymphology XVIII International Congress of Lymphology. Sept 2001, Genoa. Lymphology 35 (suppl) : 1-760,2002 : 181-187.
- Le Lay K., Riou-Franca L., Launois R. " Analyse Coût-Efficacité des chimiothérapies par voie orale en médecine de ville : le cas de la vinorelbine " - Journal d' Economie Médicale 2002; 20 (7-8) : 379 - 400.
- Launois R. " Les Réseaux : Quelles Conditions de Pérennité ? ". Les Echos, VIII^{ème} Forum International de la gestion de la santé, Paris 20 et 21 Novembre 2002.
- Launois R. " Groupama Partenaires Santé : Une démonstration jamais faite jusqu' alors ". Le Concours Médical. NS Nov. 2002.
- Launois R. " L'Évaluation par un Economiste du Service Rendu au Malade par les Médicaments et les Dispositifs de Santé ". In. J.P. Alix, L. Degos, D. Jolly, Eds " Doit-on soigner une maladie ou un malade ? ". Médecine-Sciences Flammarion, Paris 2002.
- Launois R., Lhoste F. " Quelle place pour l'évaluation économique en matière de politique de santé ? ". Journal d'Economie Médicale 2002, Vol.20, n°2.
- Launois R. " Grandeur et misère de l'enseignement économique en Faculté de Médecine ". Le Courrier de L'Evaluation en Santé - N°21-22, Mars 2002 : 25-27.

SOMMAIRE

1. INTRODUCTION	6
1.1. OBJECTIF DE L'ETUDE PREMISS	6
1.2. DESIGN DE L'ETUDE	6
1.3. STRUCTURE DES DONNEES	7
1.4. ANALYSE STATISTIQUE	8
2. DESCRIPTION DES DONNEES	11
2.1. ECHANTILLON FINAL	11
2.2. VALEURS MANQUANTES PARMIS LES CARACTERISTIQUES INITIALES	11
2.3. MESURE DES BIAIS DE SELECTION	16
3. IMPUTATION DES DONNEES MANQUANTES	28
3.1. LOGICIEL CHOISI	28
3.2. NOMBRE D'IMPUTATIONS	28
4. CALCUL DU PROPENSITY SCORE	29
4.1. METHODOLOGIE	29
4.2. MODELE « COMPLET »	30
4.3. MODELE « PARCIMONIEUX »	35
4.4. MODELE AVEC « INTERACTIONS »	40
4.5. CHOIX DU MODELE	44
5. ANNEXES	46
5.1. REFERENCES	46
5.2. INDEX DES TABLEAUX	47
5.3. INDEX DES ILLUSTRATIONS	47

1. INTRODUCTION

1.1. Objectif de l'étude PREMISS

L'objectif de ce projet est d'évaluer les coûts et l'efficacité observationnelle de l'introduction d'un nouveau traitement adjuvant, la PCArh (drotrécogine alfa ou Xigris[®]), dans le traitement des syndromes septiques sévères et choc septique dans les unités de réanimation françaises.

L'*objectif principal* sera de chiffrer le coût des malades traités avec la drotrécogine alfa par rapport à des malades comparables non traités antérieurement.

Les objectifs secondaires seront les suivants :

- Estimer l'importance des moyens qu'il convient de dégager pour la prise en charge des patients atteints de sepsis sévère et résidant sur le territoire français ;
- Calculer le rapport coût-efficacité de la prise en charge des patients en état septique traités par Xigris[®].

1.2. Design de l'étude

L'efficacité de la molécule est déjà démontrée.¹ Afin de répondre aux objectifs de l'étude PREMISS, il convient de privilégier la validité externe de l'étude à sa validité interne. La *validité externe* d'une étude se réfère à sa capacité à produire des résultats généralisables à d'autres populations ou à d'autres situations que celle qui est étudiée. La *validité interne* d'une étude se réfère à sa capacité à produire des résultats pouvant être attribuables à l'intervention étudiée plutôt qu'à des biais ou à d'autres phénomènes. Les essais cliniques randomisés sont considérés comme produisant la meilleure validité interne.² En effet, la randomisation des patients selon les stratégies à évaluer permet de minimiser le risque de biais dus à la non-comparabilité des patients ou à l'existence de facteurs de confusion connus ou inconnus. Cependant, du fait de la rigidité du protocole de l'essai, et puisque l'évaluation se fait sur une population ultra-sélectionnée, les résultats de l'étude sont difficilement extrapolables à ce qu'il se passe en pratique courante. L'étude PREMISS se fonde sur un design quasi-expérimental de type Avant/Après (pre-post study), la méthode étant plus appropriée pour recueillir des coûts sans exposer les patients non traités à une perte de chance.

➤ Phase « Avant »

Les patients ont été recrutés pour la phase « Avant » dans la période précédant la mise à disposition dans le cadre réglementaire de la molécule. Les inclusions ont débuté au 9 septembre 2002, leur fin était prévue au 9 décembre 2002 à minuit. Une période de un mois a été accordée aux réanimateurs ayant eu des difficultés de recrutement, la phase « Avant » a ainsi été clôturée le 09 janvier 2003.

Le nombre de patients inclus dans la phase « Avant » était de 517, pour un objectif initial de 786 inclusions.

➤ Phase « Après »

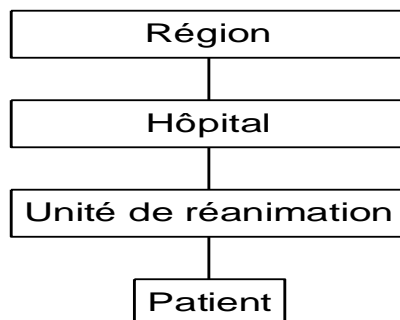
Les inclusions en phase « Après » ont débuté au 9 janvier 2003. Le gel de la base a eu lieu le 15 février 2005. A cette date, les inclusions en phase « Après » étaient au nombre de 586.

1.3. Structure des données

Les patients inclus dans la base PREMISS sont issus d'un processus d'échantillonnage stratifié : ce sont des centres de réanimation qui ont été contactés, ceux-ci incluant sur la base du volontariat des patients dans chaque phase de l'étude.

Les données ont donc une structure hiérarchique :

Figure 1 Structure hiérarchique des observations de l'étude PREMISS



Une région peut comporter plusieurs hôpitaux participant à l'étude, un hôpital peut comporter plusieurs services investigateurs, un service investigateur peut inclure plusieurs patients.

Chaque patient est relié à un service par un code se voulant unique. Chaque service comporte la mention de l'hôpital dont il est issu, chaque hôpital comporte une adresse.

Au niveau de l'hôpital, un certain nombre de variables étaient recueillies : le nombre de lits, la valeur du point ISA, le type d'établissement...

Au niveau du service étaient recueillies d'autres variables : type de service, nombre de sorties, durée moyenne de séjour...

Enfin, la majorité des informations recueillies se font à l'échelle du patient : gravité initiale, type d'infection, actes médicaux réalisés, survie...

1.4. Analyse statistique

1.4.1. Mode d'échantillonnage

La structure hiérarchisée de l'étude PREMISS demande de faire appel à l'analyse multiniveau. Cependant, certaines contraintes rendent sa mise en œuvre malaisée :

- Tout d'abord, les variables recueillies au niveau de l'hôpital et du service de réanimation ont été moins bien renseignées que celles relevant du patient :
 - Au niveau de l'hôpital, l'information recueillie est très lacunaire. Ainsi, le nombre de lits est manquant pour 40 % des hôpitaux. Pour les 60 % restants, de surcroît, l'information est de qualité variable, l'information saisie n'étant pas toujours cohérente avec d'autres sources officielles, lorsqu'elles existent. La situation empire lorsque l'on s'intéresse à d'autres variables. Ainsi, le prix de journée des spécialités onéreuses est manquant pour plus de la moitié des hôpitaux.
 - Au niveau du service, la situation est meilleure, mais la proportion de valeurs manquantes reste non négligeable : le nombre de personnel n'est pas renseigné dans plus d'un service sur cinq, le nombre de sorties dans plus d'un service sur quatre, le pourcentage de séjours avec épuration extra-rénale n'est pas renseigné dans plus d'un service sur trois...
- De plus, il s'est avéré que certains centres investigateurs ont inclus des patients provenant de plusieurs services de réanimation à la fois :
 - Le centre investigateur identifié par le n°22 (16 inclusions) a ainsi inclus des patients issus d'un service de réanimation médicale polyvalente et des patients issus d'un service de réanimation chirurgicale.
 - Le centre investigateur identifié par le n°73 (16 inclusions également) a lui inclus des patients de trois services de réanimation, médico-chirurgicale, médicale et chirurgicale.
 - Le centre investigateur identifié par le n°89 (19 inclusions) a inclus des patients issus de deux services de réanimation, toutes deux chirurgicales.

L'impossibilité de relier 51 patients à leur service de réanimation et la mauvaise qualité des données recueillies au niveau de l'hôpital et, dans une moindre mesure, de l'unité de réanimation, rend difficile la prise en compte explicite de la structure hiérarchique des données.

On emploiera – en faisant abstraction des incertitudes relatives au 51 patients précédents – si nécessaire des modèles à effets aléatoires pour prendre en compte l'effet « clustering » au niveau des services de réanimation, mais on se bornera à introduire un facteur aléatoire au niveau de l'intercept, l'information relative aux hôpitaux et services de réanimation n'étant pas exploitable.

1.4.2. Biais de recrutement

Contrairement à un essai clinique randomisé, rien ne garantit dans l'étude observationnelle PREMISS la comparabilité des patients inclus dans la phase « Avant » et celle des patients inclus dans la phase « Après ». Afin de contrôler les biais, le statisticien dispose de plusieurs outils. Une revue de littérature faisait récemment le point sur ces méthodes³ :

Tableau 1 : Méthodes employées dans les études observationnelles revues par Klungel et al.³

Strengths and limitations of methods to assess treatment effects in nonrandomized, observational studies			
Method	Used	Strengths	Limitations
Design approaches			
Historical controls	Infrequently	• Easy to identify comparison group	• Treatment effect often biased
Candidates for treatment	Infrequently	• Useful for preliminary selection	• Difficult to identify not treated candidates
Treatments for the same indication	Infrequently, when possible	• Similarity of prognostic factors	• Only useful for diseases treated with several drugs
			• Only effectiveness of one drug compared to another
Case–crossover and case–time–control designs	Infrequently	• Reduced variability by intersubject comparison	• Only useful to assess time-limited effects
			• Possible crossover effects
Data-analytical approaches			
Stratification and (weighted) matching	Frequently	• Clear interpretation / no assumptions	• Only a few covariates or rough categories can be used
		• Clarity of incomparability on used covariates	
Asymmetric stratification	Not used	• More covariates than with normal stratification	• Still limited number of covariates
Common statistical techniques: regression, logistic regression, survival analysis	Standard, very often	• More covariates than matching or stratification	• Focus is not on balancing groups
		• Easy to perform	• Adequate overlap between groups difficult to assess
Propensity scores	More often	• Many covariates possible	• Performs better with only a few number of events per confounder
Multivariate confounder score	Scarcely	• Less insensitive to misspecification	• Exaggerates significance
Ecologic studies	Scarcely	• Immune to confounding by indication	• Loss of power by reduced number of units
			• Loss of information at the individual level
Instrumental variables (IV), two-stage least squares; grouped-treatment effects	Infrequently	• Large differences per area are needed	• Difficult to identify instrumental variable(s)
			• Strong assumption that IV is unrelated with factors directly affecting outcome

Les méthodes les plus classiques en épidémiologie sont les suivantes :

- L'*appariement* : à chaque sujet de la phase « Avant », on fait correspondre un sujet de même âge, sexe, etc... de la phase « Après ». L'appariement, réalisé au début de l'enquête, n'a pas été prévu dans ce protocole en raison des difficultés pour chaque centre de recruter sur une durée d'étude limitée deux patients comparables.
- La *stratification* : l'échantillon est divisé en sous-échantillons homogènes quant à certains facteurs de confusion (sexe, âge, etc...). La stratification peut se pratiquer a posteriori, mais implique des effectifs suffisants dans chaque strate et des comparaisons par strate.
- La *standardisation* : on rend l'échantillon de l'une des phases artificiellement comparable à l'autre sur plusieurs facteurs de confusion. Cette technique s'appliquerait mal dans ce protocole.
- L'*ajustement* : c'est la méthode la plus fréquemment employée de nos jours, elle consiste à introduire les facteurs de confusion potentiels dans un modèle multivarié. Notons cependant une faiblesse majeure de cette technique : l'ajustement est pratiqué à l'aide de modèles multivariés. Sa qualité est donc fortement dépendante de celle du modèle : s'il s'avère trop éloigné de la réalité, le biais sur une covariable donnée sur laquelle on ajuste peut s'en trouver augmenté.

L'étude PREMISS fera usage de la méthode du Propensity Score.⁴ Le propensity score (PS) désigne la probabilité d'être exposé (à un traitement, par exemple) conditionnellement à un vecteur de covariables donné. On peut également l'interpréter comme la probabilité, pour une

personne de caractéristiques données, d'être exposée à un traitement. Le PS réduit ainsi un nombre potentiellement élevé de covariables en une seule variable composite qui résume correctement l'ensemble des caractéristiques mesurées.

Le PS permet de juger directement de la comparabilité entre les populations exposées et non exposées au traitement. Si les courbes de distribution du PS pour les deux populations sont confondues, on pourra les considérer comme comparables.

Le PS peut ensuite s'employer de trois principales manières pour contrôler les biais de sélection :

- soit pour *stratifier* les patients en sous-populations aux caractéristiques comparables ;
- soit pour *appairer* a posteriori les patients exposés et non exposés ;
- soit pour *ajuster* les modèles multivariés.

L'objectif premier du PS est de simuler un essai clinique randomisé, en s'assurant que les patients traités et non traités ne diffèrent pas quant à leurs caractéristiques de base. Cependant, alors qu'un essai randomisé assure en principe une totale comparabilité des patients (à l'exception du traitement reçu) grâce à la randomisation, le PS ne permet que d'assurer la comparabilité des patients sur les variables *mesurées*. Tout facteur de confusion non relevé dans le protocole peut donc biaiser les résultats.

En cas d'*appariement*, la difficulté technique réside dans le choix de la méthode choisie pour faire correspondre un patient non exposé à un patient exposé : le PS étant continu et compris entre 0 et 1, la probabilité de trouver deux patients dont le PS soit identique est nulle. Il faudra par ailleurs trouver un compromis entre un appariement plutôt strict, garantissant une bonne comparabilité des patients mais réduisant de beaucoup la taille d'échantillon (on parle alors d'appariement incomplet, puisque des patients seront exclus de l'analyse faute d'avoir trouvé leur paire), et un appariement moins strict, permettant d'obtenir un échantillon plus important mais moins comparable (on parle alors d'appariement inexact, puisque certaines paires s'avèreront dissemblables). L'appariement inexact est considéré comme préférable lorsque l'on s'intéresse à inférer des résultats sur la population. Il faudra alors avoir recours à des analyses supplémentaires (en pratique, l'ajustement) pour gérer la moindre comparabilité des sujets.

La *stratification* peut permettre de contourner les problèmes propres à l'appariement. Il est admis dans ce cas que stratifier selon les 5 quintiles du PS permet un contrôle satisfaisant des biais,^{5,6} les estimations résultantes ont tendance à être plus fiables que celles obtenues par simple régression multiple. Tout en conservant les avantages de l'appariement, la stratification permet de travailler sur des échantillons complets. Enfin, la construction même des strates (par quantiles) permet l'obtention d'un estimateur final synthétisant les estimateurs des différentes strates, obtenu par simple moyenne, puisque les effectifs de chaque strate sont en principe les mêmes. Cette technique s'apparente alors à une méthode de *standardisation*.

On peut également procéder à une *pondération* des observations : le poids d'un sujet exposé est alors égal à l'inverse de son PS, celui d'un sujet non exposé est l'inverse du complémentaire du PS (i.e. $w_i = 1/[1-PS_i]$).

Enfin, on peut employer le PS directement dans les analyses multivariées en tant que variable d'*ajustement*. Les modèles construits incluront donc le PS mais l'analyse sera faite de manière classique. Cette dernière option est plus rarement employée : en effet, elle partage les failles de tout ajustement multivarié (dépendance au modèle spécifié).

L'estimation d'un propensity score ne peut se faire que si l'ensemble des variables intervenant dans son calcul sont renseignées : en procédant par régression logistique, tout cas manquant est exclu de l'échantillon. Avant de procéder au calcul du PS, il faudra imputer les valeurs manquantes.

2. DESCRIPTION DES DONNEES

2.1. Echantillon final

On va travailler sur la base gelée et corrigée, en date du 04 mars 2005.

Certains centres ont inclus des patients sans les renseigner, ils ont été supprimés de l'analyse.

On supprime également tout patient comportant plus de 100 valeurs manquantes sur les données principales (tables IdPat, QM, DO, CCAM, Omega et Sortie).

On travaille au final sur 1096 patients, 509 en phase « Avant » et 587 en phase « Après ».

2.2. Valeurs manquantes parmi les caractéristiques initiales

On s'intéresse aux caractéristiques des patients à l'inclusion :

❖ Caractéristiques des patients

1. Date de naissance
 2. Sexe
 3. Poids à l'inclusion
 4. Date d'inclusion
 5. Date d'entrée à l'hôpital
 6. Date d'entrée en réa
 7. Type de Séjour
 8. Mode d'entrée en réanimation
 9. Date de début du sepsis
 10. Présence de choc septique
 11. IGS II à l'entrée en réa
 12. Phase de l'étude
- #### ❖ Questionnaire médical
13. Insuffisance rénale chronique
 14. Insuffisance hépatique chronique
 15. Insuffisance cardiaque chronique
 16. Insuffisance respiratoire chronique
 17. Diabète insulino-dépendant
 18. Traitement immunodépresseur
 19. Chimiothérapie
 20. Cancer métastatique
 21. Maladie hématologique maligne
 22. VIH +
 23. Corticothérapie (> 3 semaines)
 24. Indice de McCabe
 25. Site d'infection inconnu
 26. Site d'infection : endo-cardiovasculaire
 27. Site d'infection : infection cathéter central
 28. Site d'infection : intra-abdominale
 29. Site d'infection : neurologique
 30. Site d'infection : os articulaires
 31. Site d'infection : plaies opératoires
 32. Site d'infection : pulmonaire
 33. Site d'infection : septicémie primaire
 34. Site d'infection : tissus mous

35. Site d'infection : tractus urinaire
36. Nombre total de sites d'infection
 - ❖ Défaillances d'organe
37. Température
38. Fréquence cardiaque
39. Pression artérielle systolique
40. Pression partielle en O₂
41. Bilirubine
42. Prothrombine
43. Score de Glasgow
44. Globules blancs
45. Plaquettes
46. Kaliémie
47. Natrémie
48. HCO₃⁻
49. Urée
50. Créatinine
51. Diurèse

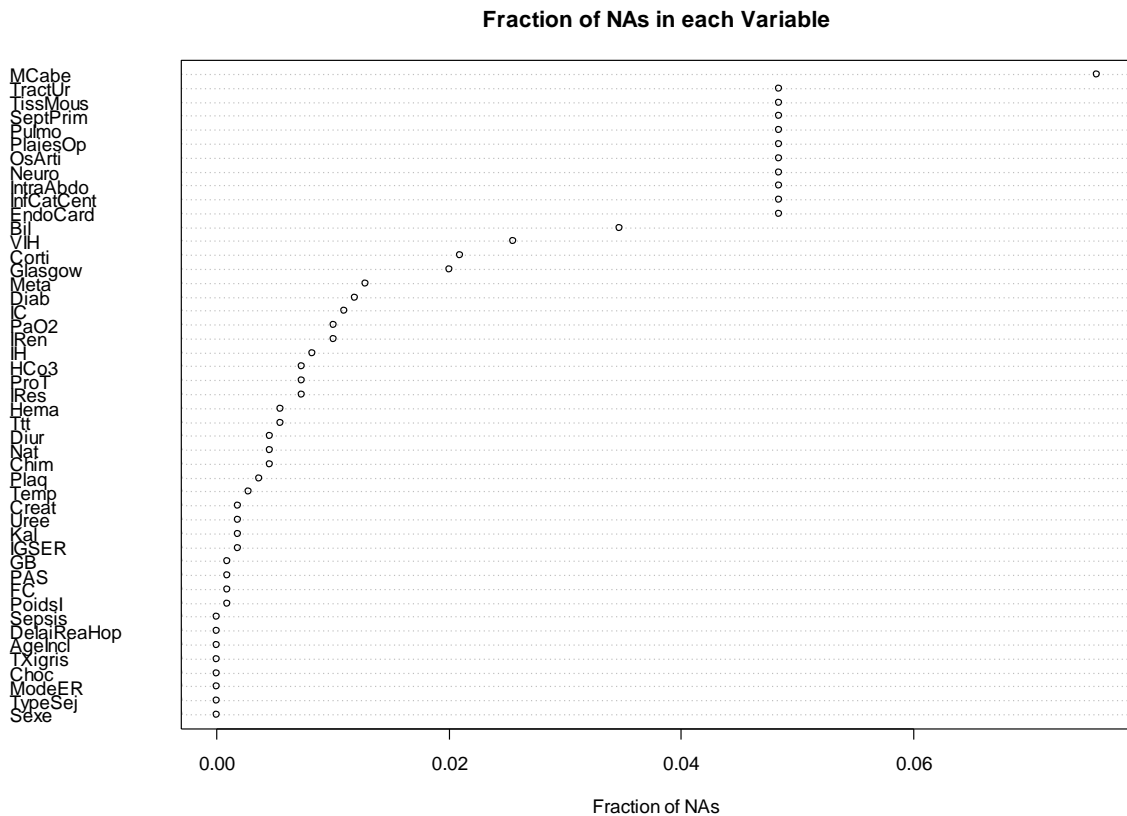
Parmi ces 51 variables, toutes ne sont pas pertinentes à conserver.

- Les dates ne peuvent être employées comme telles. On conservera donc :
 - o A partir de la date de naissance et de la date d'inclusion, l'âge des patients ;
 - o A partir de la date d'entrée à l'hôpital et de la date d'entrée en réanimation, le délai entre l'arrivée à l'hôpital et l'arrivée en réa (notons que cette variable sera corrélée au mode d'entrée, puisqu'un délai nul correspond forcément à une entrée directe : SAMU primaire ou urgences)
 - o A partir de la date d'entrée en réa et la date de sepsis, le délai entre l'entrée en réa et le début du sepsis. Puisqu'il était demandé aux réanimateurs, en cas de date d'apparition du Sepsis Sévère antérieure à l'entrée en réanimation, de coder la date d'entrée en réanimation, la valeur numérique du délai calculé n'a pas de sens. On va donc regrouper le délai entre l'entrée en réa et l'apparition du sepsis en 4 classes : « 0 jours ou moins », « 1 jour », « 2 à 5 jours », « plus de 5 jours ». Les classes ainsi formées n'ont d'autre justification que de fournir des effectifs à peu près équilibrés (exception faite de la classe « 0 jours ou moins » qui regroupe la majorité des patients)
- Pour les sites d'infection :
 - o On ne raisonnera pas sur le nombre total de sites, entièrement déterminé par les sites en question ;
 - o On ne conservera pas la variable « site d'infection inconnu », considérant celle-ci comme une indicatrice de valeurs manquantes sur tous les sites.

On travaille donc en définitive, sans compter l'identifiant du service ni l'identifiant du patient, sur 47 variables (46 en si l'on élimine la phase de recrutement, qui est toujours connue).

Parmi celles-ci, les variables manquantes se répartissent comme suit :

Figure 2 Proportion de valeurs manquantes par variable



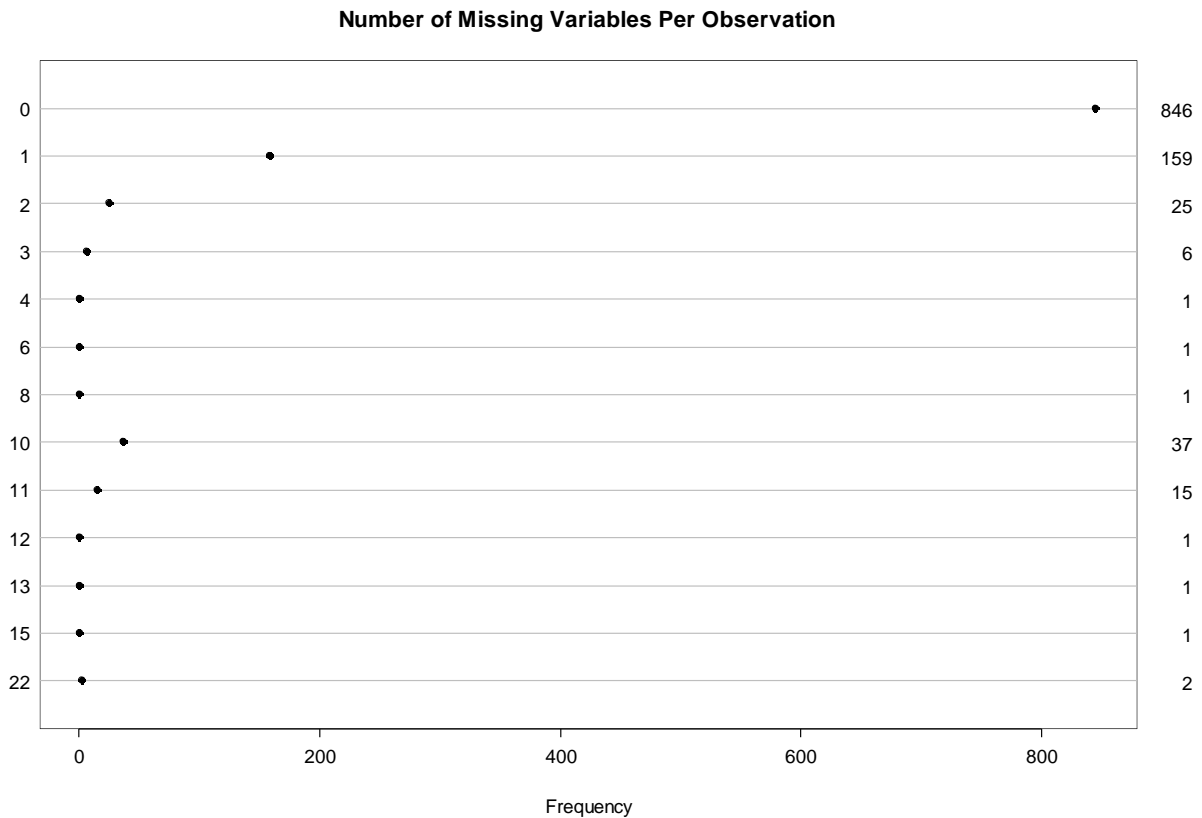
Le score de McCabe est la variable la plus souvent manquante, puisqu'elle n'est pas renseignée dans 7.6 % des cas.

Suivent les informations relatives aux sites d'infection, le site d'infection étant inconnu pour 51 patients et non renseigné (e.g. case « site d'infection inconnu » non cochée mais aucun site d'infection quand même) pour 2 patients (ceux-ci n'ayant pas rempli la page Internet correspondante).

Par ordre décroissant de valeurs manquantes, suivent la bilirubine, le statut VIH, les corticothérapies, le score de Glasgow, la présence d'un cancer métastatique...

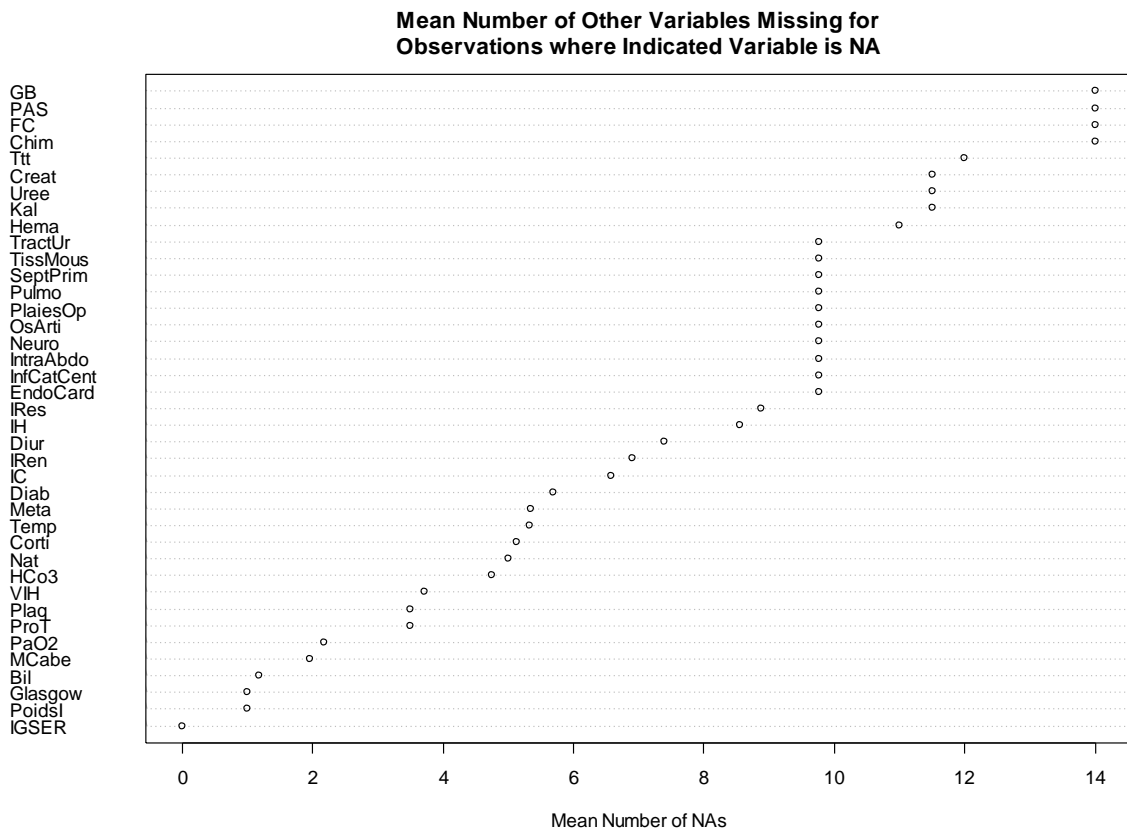
Les variables Sexe, Type de séjour, Mode d'entrée en réanimation, présence de choc septique, Phase de traitement, âge à l'inclusion, délai entre l'entrée en réa et l'entrée à l'hôpital et délai entre l'entrée en réa et l'apparition du sepsis sévère ne comportent aucune valeur manquante.

Figure 3 Répartition du nombre de valeurs manquantes par patient



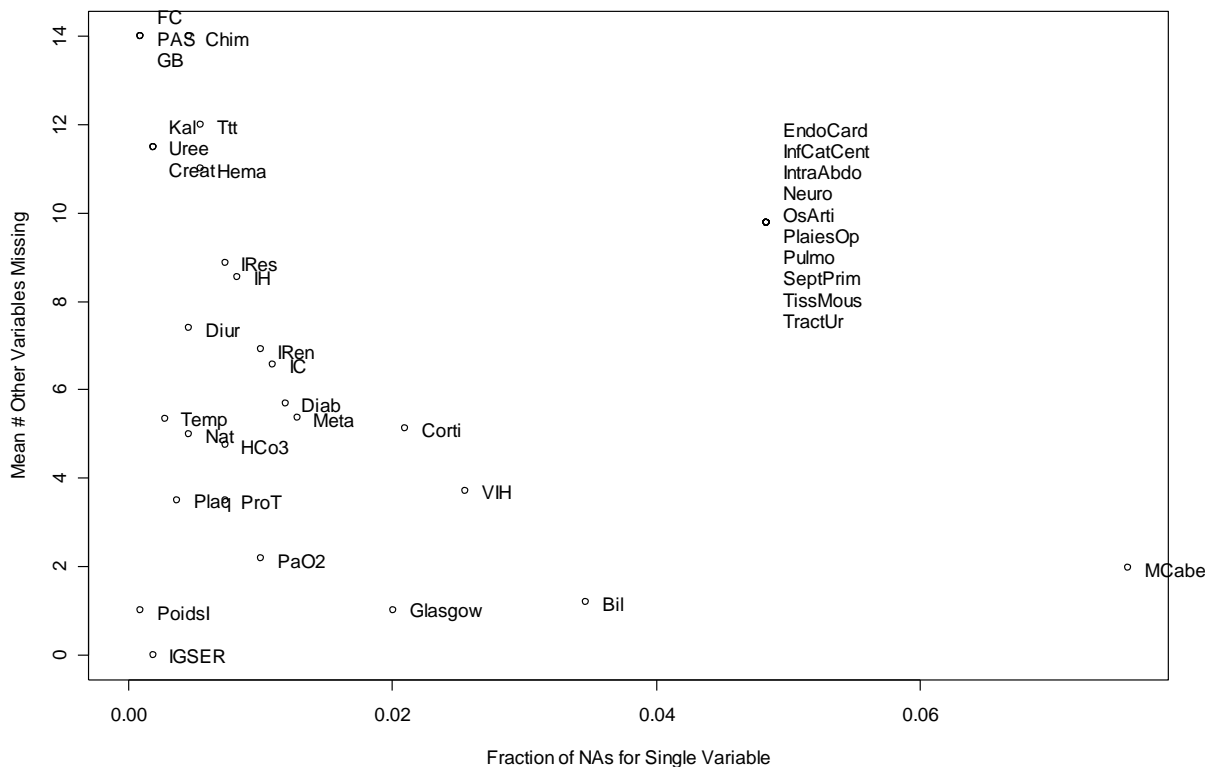
Pour 846 patients, les données sont complètes sur l'ensemble des variables. Pour les autres, le cas le plus fréquent est de n'observer qu'une valeur manquante. Signalons tout de même deux patients avec 22 valeurs manquantes (l'ensemble du questionnaire médical) et 37 avec 10 valeurs manquantes (pour 36 d'entre eux, il s'agit simplement des 10 sites d'infection).

Figure 4 Nombre moyen de valeurs manquantes si la variable est manquante



Les variables qui risquent de poser le plus de problèmes lors de l'imputation sont le taux de leucocytes, la pression artérielle systolique, la fréquence cardiaque et la présence d'une chimiothérapie : en moyenne, si l'une de ces valeurs est manquante, on observe 14 autres valeurs manquantes. A l'inverse, si l'IGS à l'entrée en réa est manquant, ce sera la seule variable à manquer. On retrouve là aussi le bloc des sites d'infection : si le site est inconnu, c'est forcément l'ensemble des sites qui seront manquants.

Figure 5 Moyenne conditionnelle de valeurs manquantes et pourcentage de valeurs manquantes



Cette figure reprend les informations précédentes en mettant en valeur les variables qui ont souvent tendance à être manquantes sans que cela entraîne plus d'autres variables manquantes (comme pour le McCabe) et celles qui sont rarement manquantes mais, lorsqu'elles le sont, sont liées au fait que d'autres variables soient manquantes (FC, PAS, GB, Chimio). Là encore, le bloc des sites d'infection se détache.

2.3. Mesure des biais de sélection

2.3.1. Les différences standardisées

La différence standardisée est un indicateur du bon équilibre d'une caractéristique entre les deux bras.

Pour une variable quantitative, elle se calcule comme suit :

$$d = \frac{\overline{X}_{Xigris} - \overline{X}_{Avant}}{\sqrt{\frac{S_{Xigris}^2 + S_{Avant}^2}{2}}}$$

Il s'agit donc de la différence entre les moyennes pondérée par l'écart-type commun. La généralisation au cas d'une variable binaire est immédiate :

$$d = \frac{p_{Xigris} - p_{Avant}}{\sqrt{\frac{p_{Xigris}(1-p_{Xigris}) + p_{Avant}(1-p_{Avant})}{2}}}$$

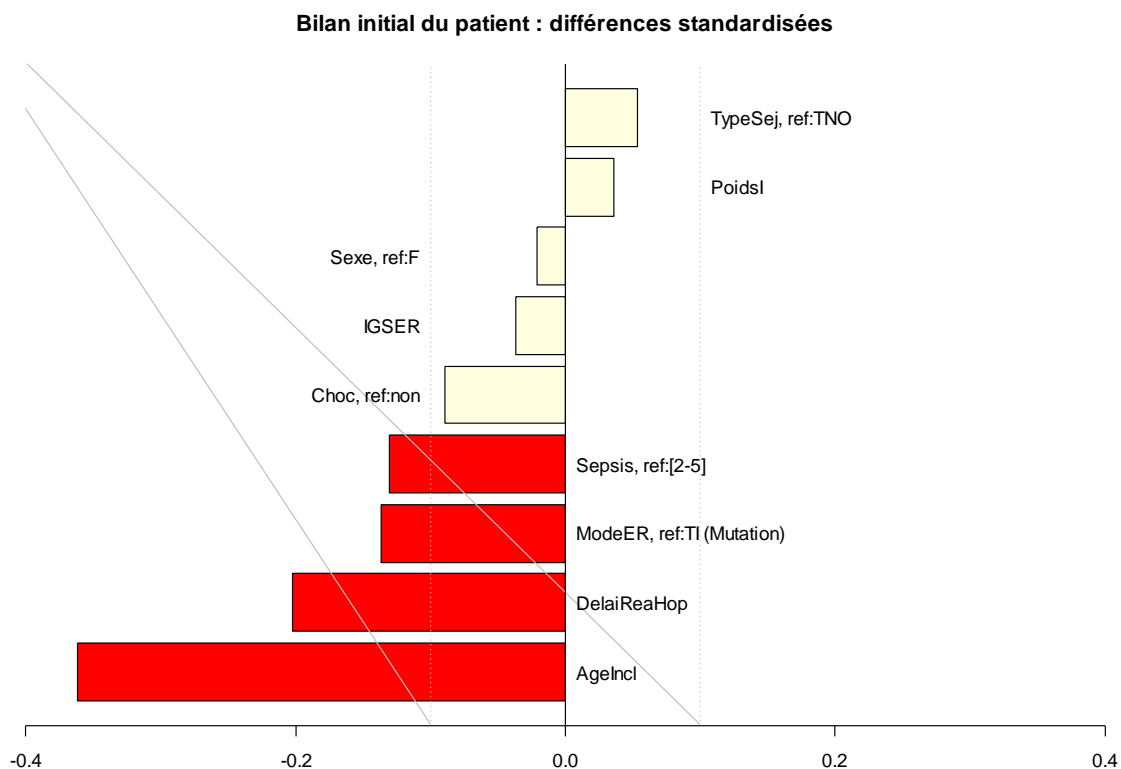
Pour les variables qualitatives à plus de deux modalités, la valeur de la différence standardisée dépend de la valeur de la modalité prise comme modalité de référence.

On choisira dans ce cas la modalité maximisant la valeur absolue de la différence standardisée.

2.3.2. Bilan initial

On recense 9 variables : l'âge du patient, son sexe, son poids, le délai écoulé entre l'entrée à l'hôpital et l'entrée en réa, le type de séjour, le mode d'entrée, le délai écoulé entre l'entrée en réa et le début du sepsis, la présence d'un choc septique, l'IGS II à l'entrée en réanimation.

Figure 6 Différences standardisées calculées sur les caractéristiques du patient à son bilan initial



On peut soupçonner la présence de biais de recrutement sur 4 des 9 variables : l'âge à l'inclusion, le délai entre l'entrée à l'hôpital et en réanimation, le mode d'entrée en réanimation et le délai entre l'entrée en réanimation et le début du sepsis.

Les patients en phase « Après » sont plus jeunes que ceux en phase « Avant » (âge moyen de 58.1 et 63.9 ans respectivement, $p = 2.10^{-8}$, test non paramétrique de Mann-Whitney).

Figure 7 Distribution de l'âge des patients en fonction de la phase de recrutement

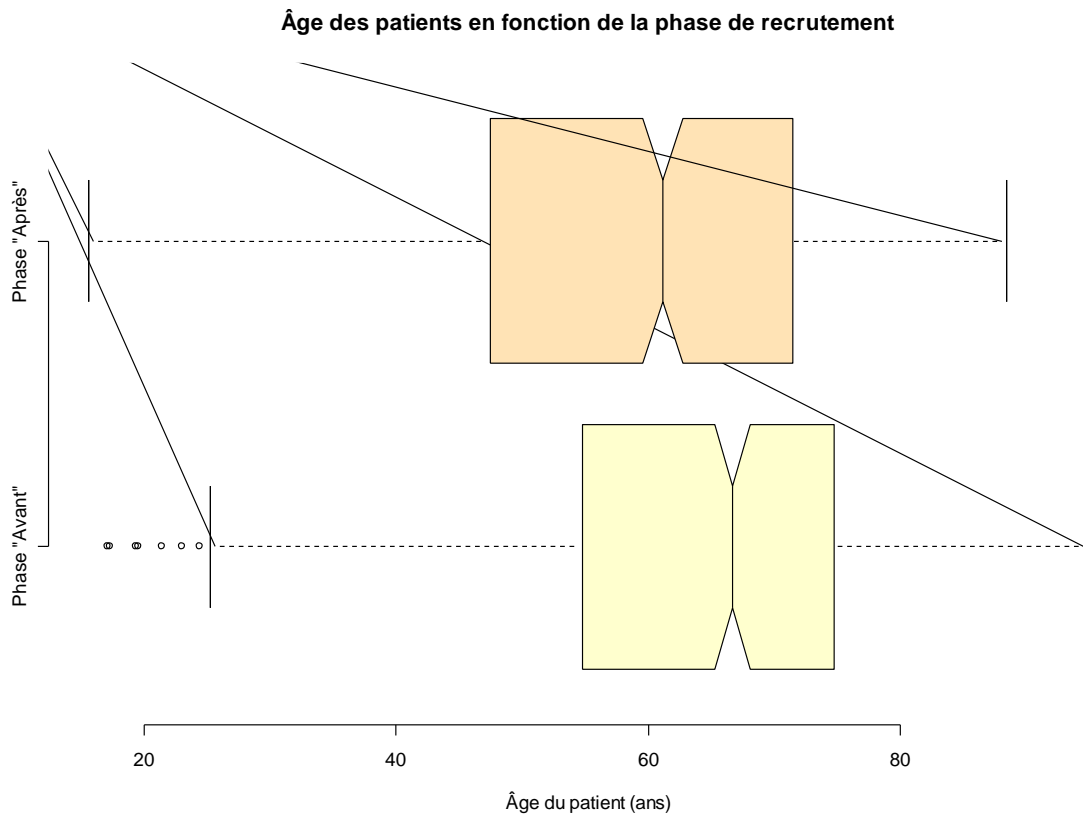
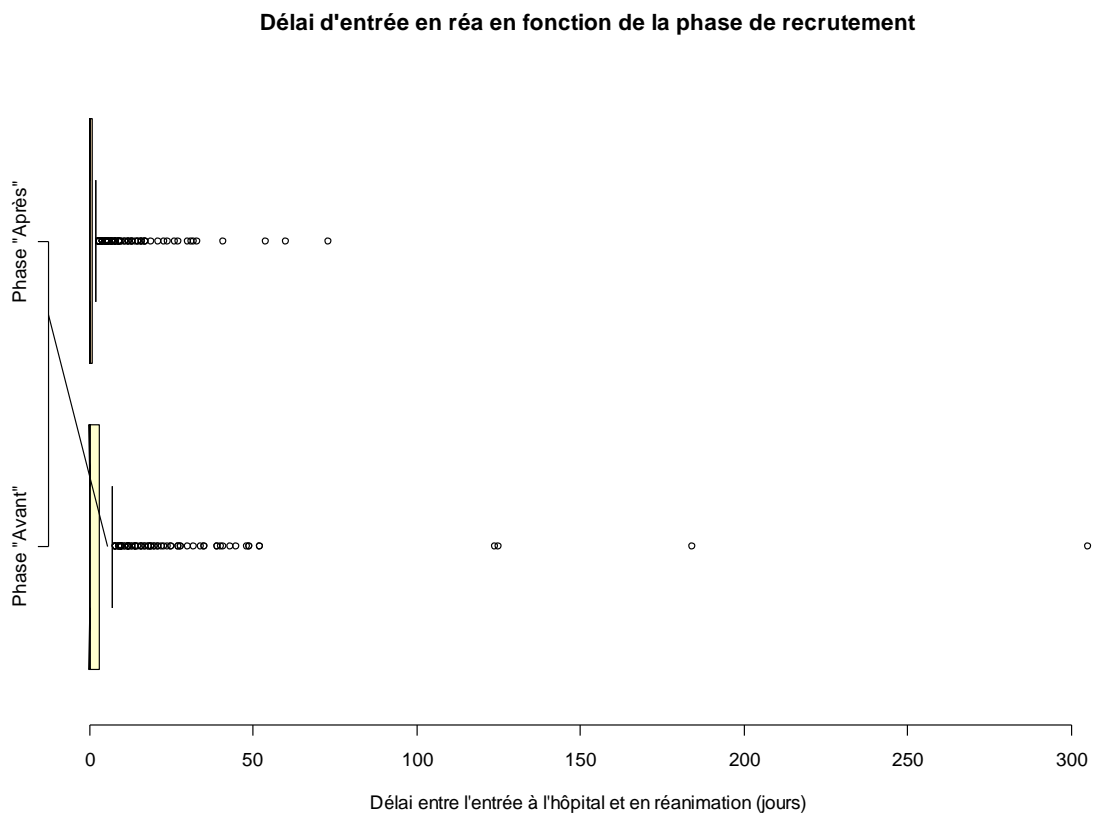


Figure 8 Délai entre l'entrée à l'hôpital et l'entrée en réanimation et phase de recrutement



L'examen de la Figure 8 montre que le délai entre l'entrée à l'hôpital et l'entrée en réa est extrêmement dispersé. Si une majeure partie des patients entre directement en réanimation (54.4 % des patients en phase « Avant » et 60.8 % en phase « Après », $p = 0.037$, test exact de Fisher), certains parviennent en réanimation suite à un séjour hospitalier de durée importante. Le délai moyen entre l'entrée à l'hôpital et l'entrée en réanimation est de 5.1 jours en phase « Avant » et de 2.2 jours en phase « Après » ($p = 0.004$, test non-paramétrique de Mann-Whitney).

Le délai d'entrée en réanimation est corrélé au mode d'entrée en réanimation :

Tableau 2 : Mode d'entrée en réanimation en fonction de la phase de recrutement

Mode d'entrée	Urgences	SAMU primaire	SAMU secondaire	Mutation
Phase « Avant »	27.1 %	9.0 %	19.8 %	44.0 %
Phase « Après »	29.5 %	8.2 %	25.0 %	37.3 %
Total	28.4 %	8.6 %	22.6 %	40.4 %

Les patients en phase « Après » n'ont pas les mêmes modes d'entrée en réanimation ($p = 0.070$, test du Chi²). En particulier, ils ont plus tendance que ceux de phase « Avant » à l'entrée par transfert externe que par transfert interne.

Tableau 3 : Délai entre l'apparition du sepsis et l'entrée en réa en fonction de la phase de recrutement

Délai (jours)	0 ou moins	1	2 à 5	Plus de 5
Phase « Avant »	74.9 %	9.6 %	9.2 %	6.3 %
Phase « Après »	75.6 %	10.6 %	5.8 %	8.0 %
Total	75.3 %	10.1 %	7.4 %	7.2 %

Malgré une différence standardisée supérieure à 10 %, on ne peut rejeter l'hypothèse d'indépendance entre le délai d'apparition du sepsis et la phase de recrutement ($p = 0.125$, test du Chi²). On ne peut conclure à la présence d'un biais de recrutement sur cette variable.

2.3.3. Bilan Médical

On s'intéresse cette fois à 22 variables : Les insuffisances chroniques rénale, hépatique, cardiaque, et respiratoire, le diabète insulino-dépendant, la présence d'un traitement immunodépresseur, d'une chimiothérapie, d'un cancer métastatique, d'une maladie hématologique maligne, du VIH, d'une corticothérapie de durée supérieure à trois semaines, le score de McCabe, et les sites d'infection (au nombre de 10).

7 caractéristiques pourraient participer à un biais de recrutement : 4 sites d'infection (Endocardiovasculaire et Tractus Urinaire moins fréquents dans la phase « Après », Tissus mous et neurologique moins fréquents dans la phase « Avant »), la présence d'un cancer métastaté (moins fréquent en phase « Après »), d'une insuffisance rénale (moins fréquente en phase « Après »), ou un score de McCabe de 3 (maladie mortelle à court terme, moins fréquent en phase « Après »).

Figure 9 Différences standardisées calculées sur les caractéristiques du patient à son bilan médical

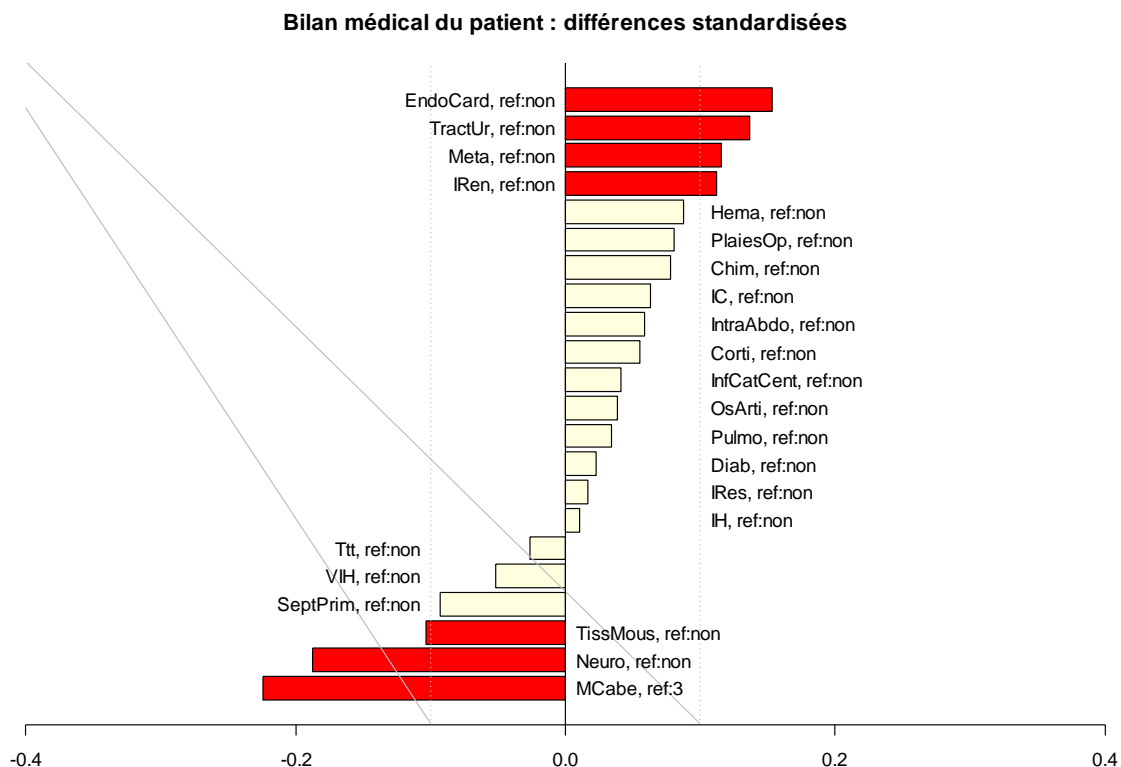


Tableau 4 : Caractéristiques du bilan médical avec une différence standardisée supérieure à 10 %

Caractéristique	Phase « Avant »	Phase « Après »	Total	p-value*
Site endo-cardiovasculaire	3.6 %	1.2 %	2.3 %	0.0130
Site tractus urinaire	12.0 %	7.9 %	9.8 %	0.0283
Cancer métastatique	6.3 %	3.8 %	5.0 %	0.0682
Insuffisance rénale chron.	7.7 %	5.0 %	6.3 %	0.0783
Site Tissus Mous	5.1 %	7.6 %	6.4 %	0.1013
Site Neurologique	2.7 %	6.7 %	4.9 %	0.0036
McCabe = 3	9.0 %	3.6 %	6.1 %	0.0004

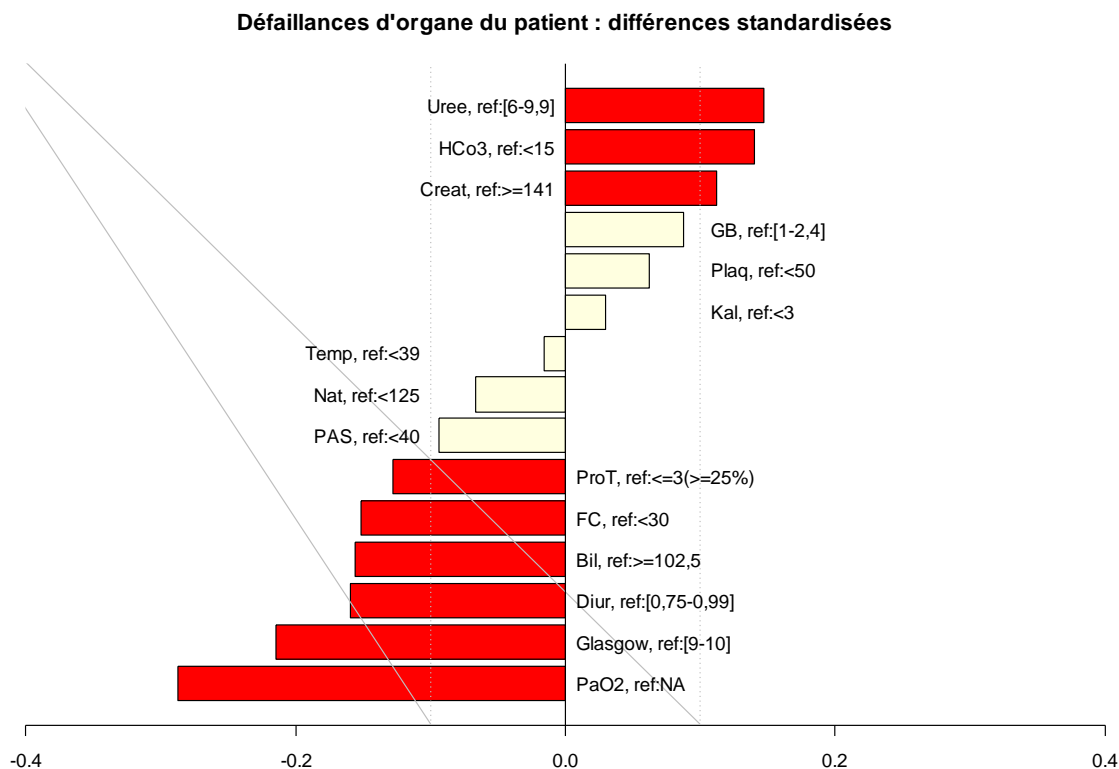
* Test exact de Fisher (bilatéral)

On retrouve des différences significatives au niveau des sites d'infection endo-cardiovasculaire, du tractus urinaire (plus fréquents en phase « Avant ») et neurologique (plus fréquent en phase « Après »). Les différences ne sont pas significatives au seuil de 5 % pour le cancer métastatique, l'insuffisance rénale chronique et le site d'infection des tissus mous. Le score de McCabe (dans le Tableau 4, seule est testée la présence d'une maladie mortelle à court terme, le test du score McCabe dans sa globalité par un Chi² donne une p-value de 2.10⁻⁶) est la caractéristique initiale pour laquelle les biais de recrutement sont les plus importants.

2.3.4. Défaillances d'organe à l'entrée dans l'étude

On s'intéresse ici aux variables biologiques intervenant dans le calcul du LODS et de l'IGS II, au nombre de 15 : température, fréquence cardiaque, pression artérielle systolique, pression partielle en O₂, bilirubine, prothrombine, score de Glasgow, globules blancs, plaquettes, kaliémie, natrémie, bicarbonatémie, urée, créatinine et diurèse (l'âge, les comorbidités et le type de séjour interviennent également dans le calcul de l'IGS mais ont déjà été décrits dans les chapitres précédents).

Figure 10 Différences standardisées calculées sur les caractéristiques biologiques du patient



Les figures suivantes décrivent les 9 variables dont les différences standardisées sont supérieures à 10 %.

Pour la pression partielle en O₂, on retrouve plus de patients non ventilés à l'entrée dans l'étude en phase « Avant », plus de patients avec une pression inférieure à 100 mmHg en phase « Après ». Les différences sont significatives. Les patients de phase « Après » seraient plus sévères sur le plan respiratoire.

A l'inverse, les patients de la phase « Avant » seraient plus sévères sur le plan neurologique.

La diurèse est plus souvent inférieure à 0,75 L/24h en phase « Après », plus souvent supérieure à ce seuil en phase « Avant ».

La fréquence cardiaque est plus fréquemment inférieure à 30 en phase « Avant ». Les patients de cette phase sont aussi plus nombreux à avoir une fréquence cardiaque normale (70-119).

Le taux de prothrombine est plus souvent inférieur à 25 % en phase « Avant ».

La bicarbonatémie est plus fréquemment inférieure à 15 mEq/L en phase « Avant ».

Les différences observées au niveau de la bilirubine, de l'urée et de la créatinine ne sont pas significatives.

Figure 11 PaO2, Score de Glasgow et Diurèse en fonction de la phase de recrutement

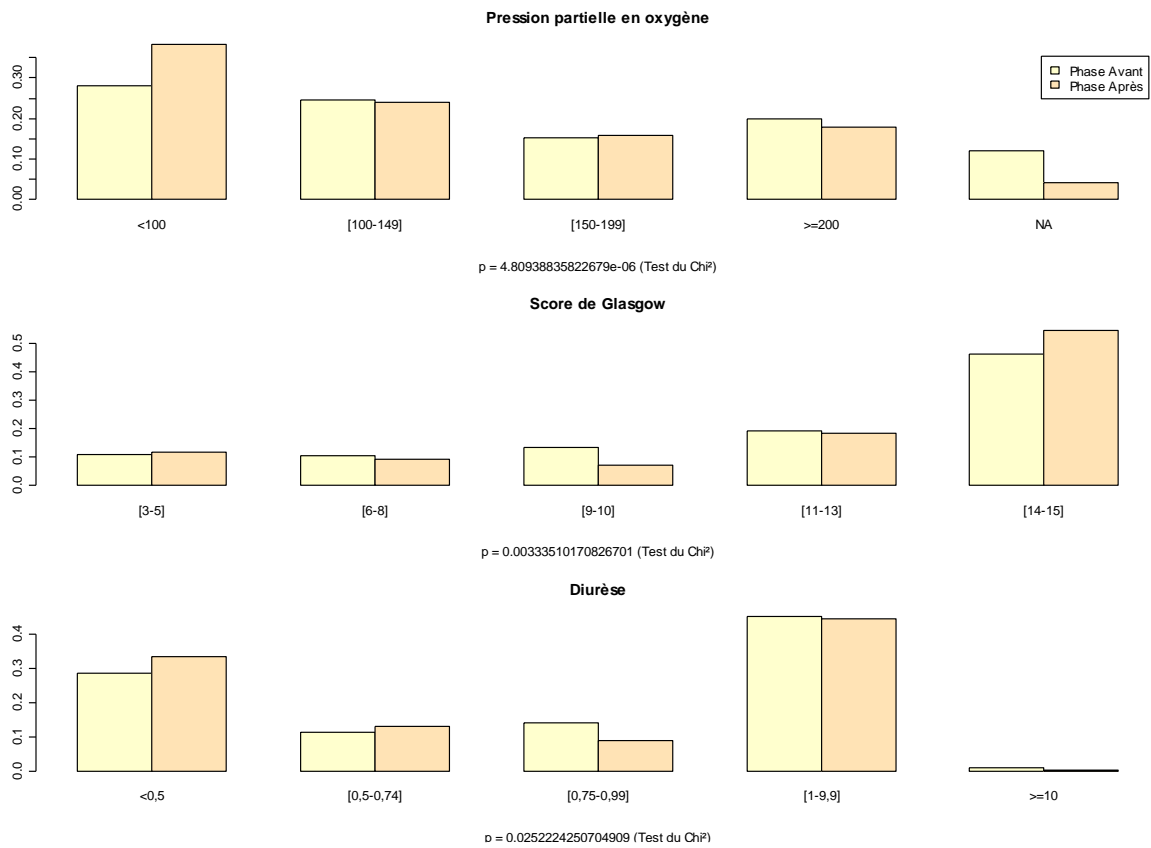


Figure 12 Bilirubine, Fréquence cardiaque et Prothrombine en fonction de la phase de recrutement

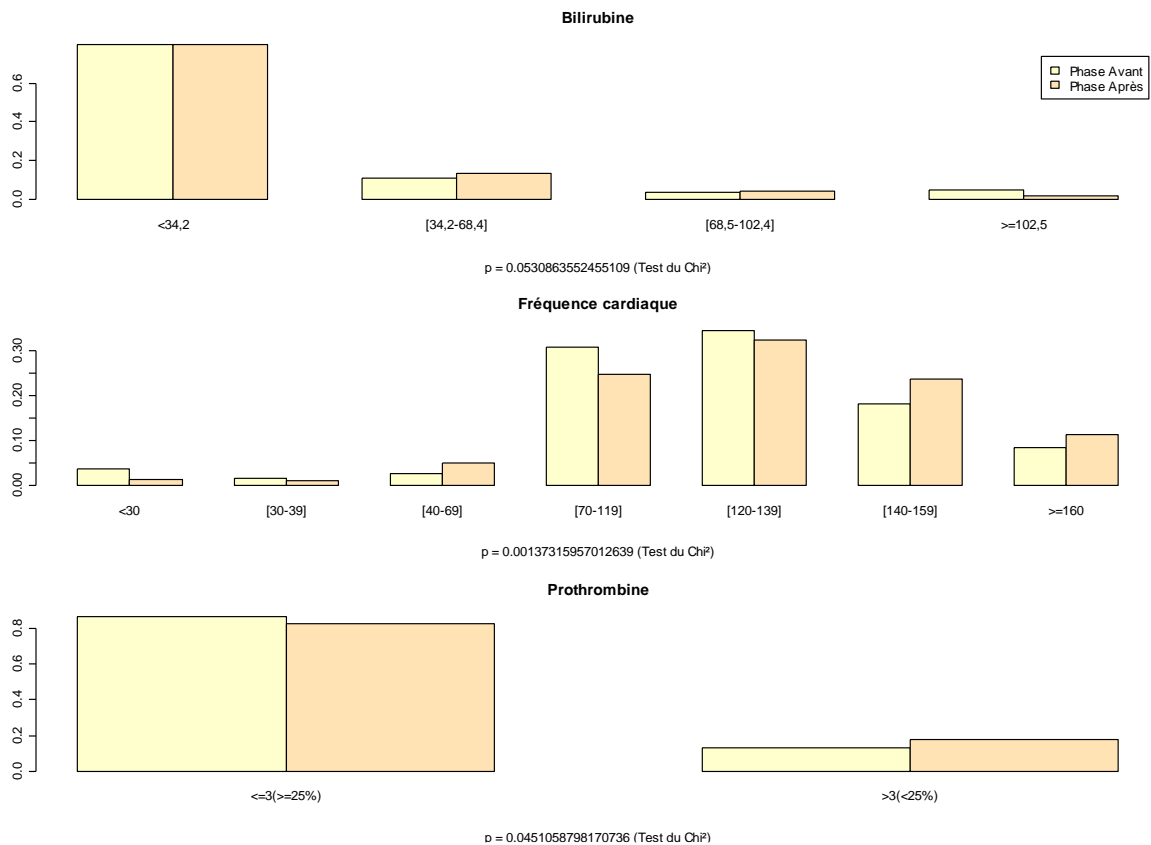
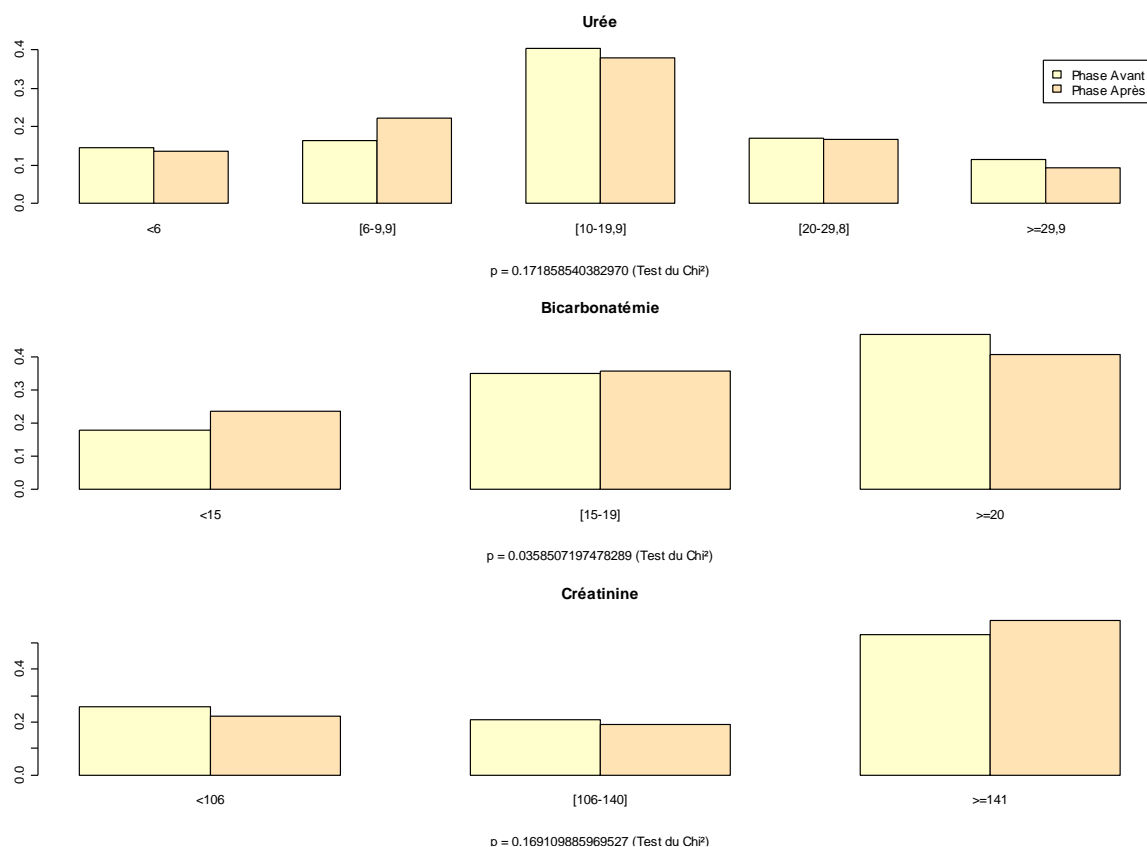


Figure 13 Urée, Bicarbonatémie et Créatinine en fonction de la phase de recrutement



2.3.5. Scores

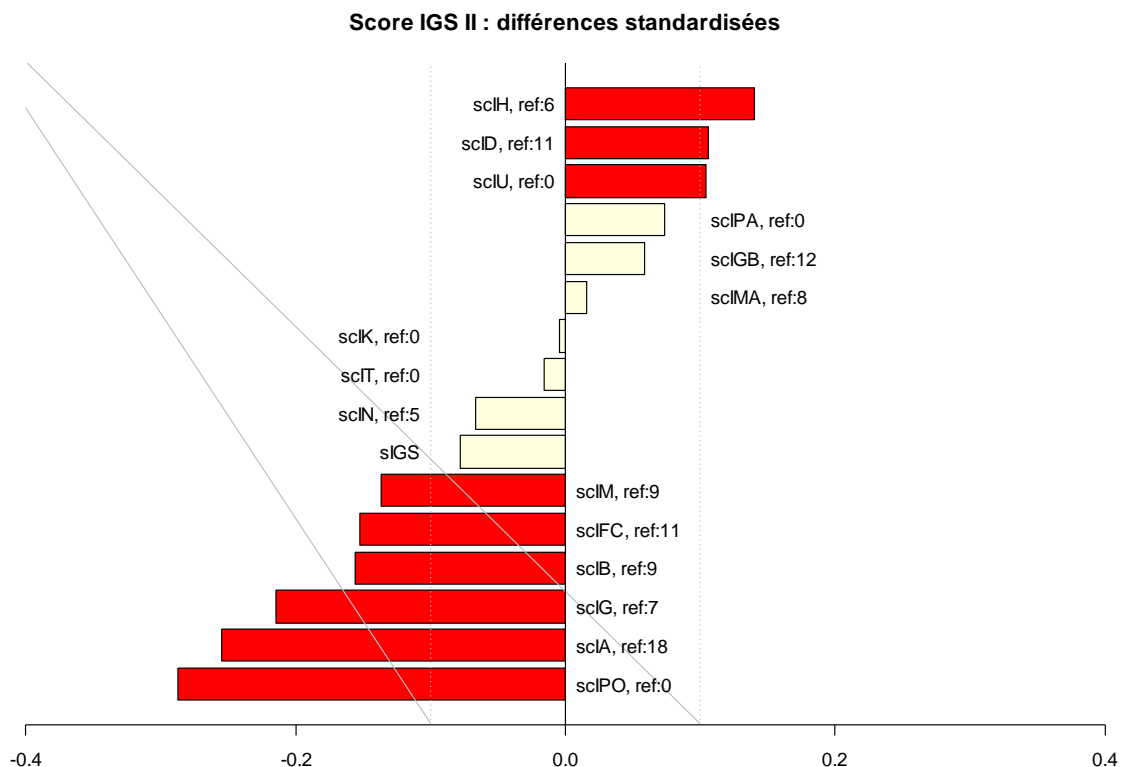
➤ Score IGS II

15 paramètres interviennent dans le calcul de l'IGS II. A chaque paramètre est associé un score de sévérité, leur somme donnant la valeur de l'indice de gravité.

L'IGS II global semble bien équilibré dans les deux phases de l'étude (avec une légère tendance à un IGS II plus important en phase « Avant » – p = 0.322, test non paramétrique de Mann-Whitney). Ce n'est pas le cas des variables biologiques qui en sont à l'origine (voir la Figure 10). La Figure 14 nous renseigne sur les mécanismes expliquant le phénomène :

- Les patients de la phase « Avant » sont plus âgés, plus sévères sur la bilirubine, l'urée, les comorbidités, la fréquence cardiaque, la bilirubine et le score de Glasgow ;
- Les patients de la phase « Après » sont plus sévères sur la PaO₂, la bicarbonatémie et la diurèse.

Figure 14 Différences standardisées calculées sur les composantes de l'IGS II



L'assimilation des composantes du score IGS II à des variables quantitatives peut masquer certaines tendances, comme en témoigne le Tableau 5 :

Tableau 5 : Caractéristiques des composantes du score IGS II

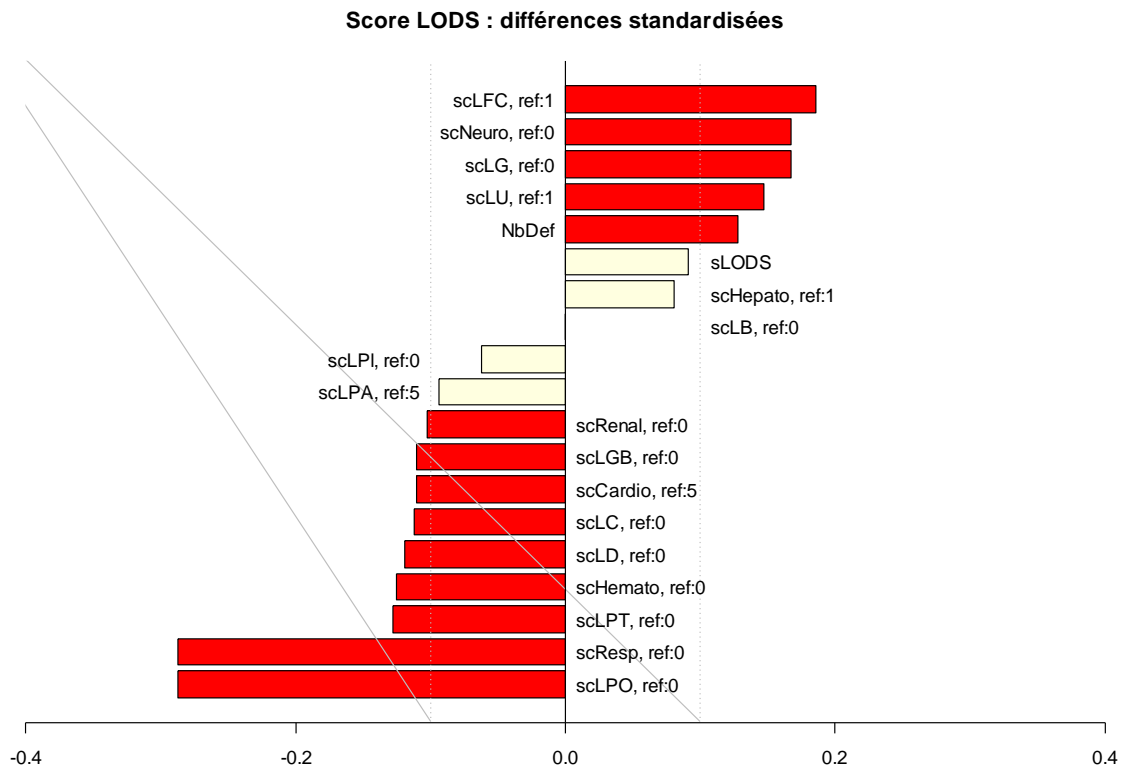
Caractéristique	Phase « Avant »	Phase « Après »	p-value*
Caractéristiques pour lesquelles les patients de la phase « Avant » sont plus graves			
Score Âge	11.3 (5.2)	9.6 (5.5)	4.10 ⁻⁶
Score Bilirubine	0.6 (2.1)	0.4 (1.5)	0.0371
Score Urée	4.6 (3.3)	4.2 (3.4)	0.1800
Score Comorbidités	1.2 (3.5)	0.9 (3.3)	0.0365
IGS II	59.0 (17.6)	57.7 (17.0)	0.3217
Score Natrémie	0.4 (1.1)	0.3 (1.0)	0.4887
Score de Glasgow	6.1 (8.1)	5.6 (8.4)	0.0033
Score Pression Artérielle	7.5 (4.3)	7.2 (4.3)	NA†
Caractéristiques pour lesquelles les patients de la phase « Après » sont plus graves			
Score Kaliémie	1.0 (1.4)	1.0 (1.4)	0.9976
Score Température	1.5 (1.5)	1.5 (1.5)	0.8402
Score Mode d'admission	6.2 (1.6)	6.2 (1.6)	0.9513
Score Fréquence cardiaque	3.3 (2.8)	3.4 (2.5)	0.0018
Score Leucocytes	1.3 (2.4)	1.3 (2.6)	0.5819
Score Diurèse	4.2 (4.6)	4.6 (4.8)	0.1664
Score Bicarbonatémie	2.1 (2.3)	2.5 (2.3)	0.0359
Score PaO ₂	7.9 (3.4)	8.8 (2.6)	1.10 ⁻⁶

* : Test de Mann-Whitney si le nombre de valeurs prises par la variable est supérieur ou égal à 10, test du Chi² sinon (le nombre limité de modalités des variables rend discutable les méthodes spécifiques aux variables quantitatives). † : Les conditions de validité du Chi² ne sont pas réunies (faibles effectifs dans l'une des modalités)

➤ Score LODS

Le score LODS synthétise 11 variables biologiques pour définir 6 types de défaillances : Cardiovasculaire, Hématologique, Hépatique, Neurologique, Rénale et Respiratoire. La somme des scores de ces 6 défaillances donne le score final.

Figure 15 Différences standardisées calculées sur les composantes du LODS



Alors que l'IGS II tendait à être plus important en phase « Avant », le LODS est plus important en phase « Après ». Les patients de la phase « Après » sont plus graves sur le plan respiratoire, ont tendance à avoir plus de défaillances telles que définies par le LODS (non significatif), à être plus graves sur le plan hématologique, hépatique et rénal (tendances non significatives). Les patients de la phase « Avant » sont plus graves sur le plan cardiovasculaire (non significatif) et Neurologique (statistiquement significatif).

Tableau 6 : Caractéristiques des composantes du score LODS

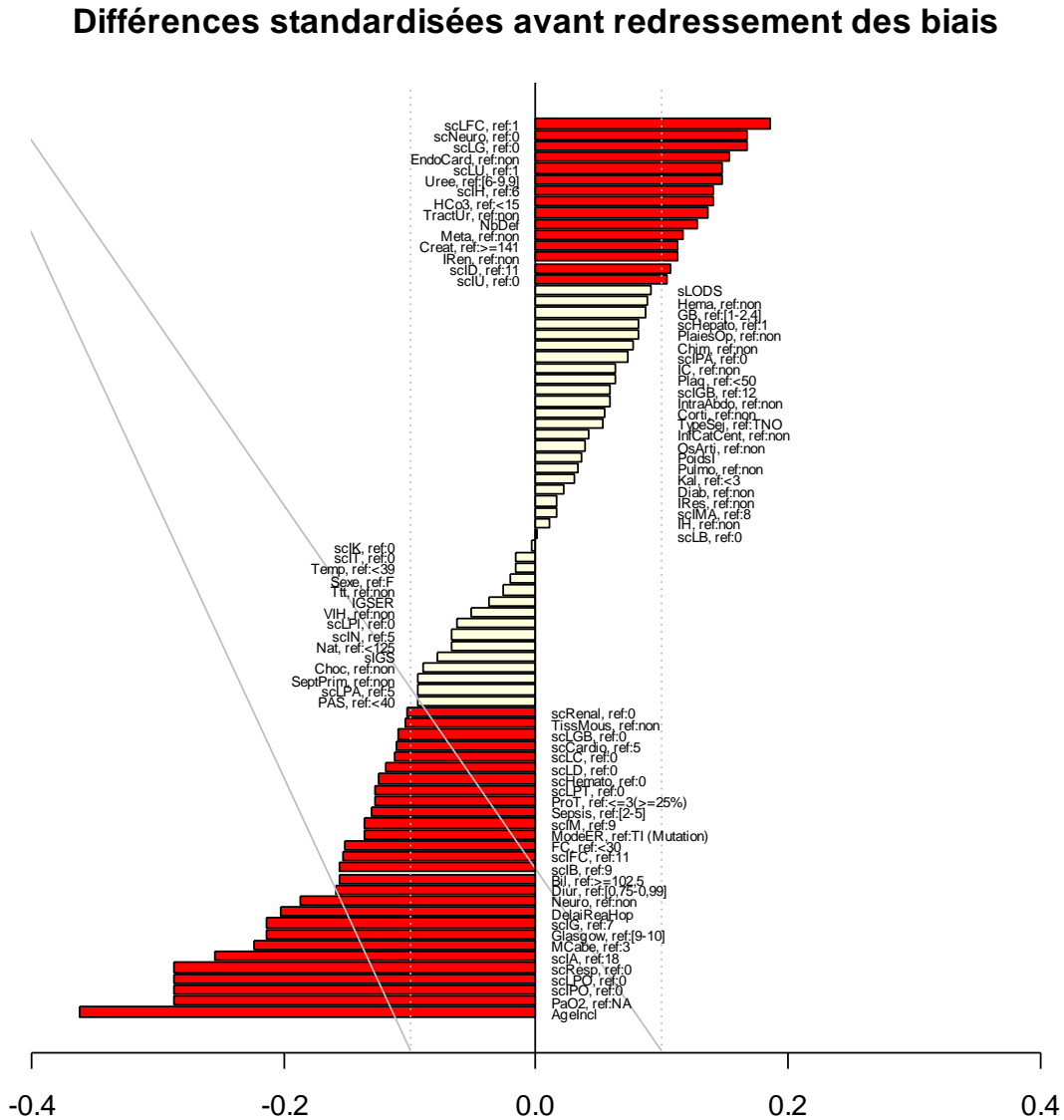
Caractéristique	Phase « Avant »	Phase « Après »	p-value*
Caractéristiques pour lesquelles les patients de la phase « Avant » sont plus graves			
Score Urée	2.8 (1.8)	2.7 (1.7)	0.1182
Score Pression artérielle	1.6 (1.4)	1.5 (1.3)	0.4456
Score de Glasgow	1.2 (1.6)	1.1 (1.7)	0.0207
Score Neurologique (Glasgow)	1.2 (1.6)	1.1 (1.7)	0.0207
Score Cardiovasculaire (FC, PAS)	1.7 (1.3)	1.6 (1.2)	0.2603
Score Fréquence cardiaque	0.5 (1.0)	0.4 (0.7)	0.0009
Score Bilirubine	0.2 (0.4)	0.2 (0.4)	0.9393
Caractéristiques pour lesquelles les patients de la phase « Après » sont plus graves			
Score Rénal (Urée, Créat., Diur.)	3.3 (1.8)	3.4 (1.7)	0.3405
Score Plaquettes	0.1 (0.3)	0.1 (0.3)	0.3603
Score Hépatique (Bil., Prothr.)	0.3 (0.5)	0.3 (0.5)	0.2052
LODS	8.6 (3.6)	8.9 (3.3)	0.0527
Score Leucocytes	0.2 (0.6)	0.3 (0.7)	0.1935
Score Créatinine	1.6 (1.5)	1.8 (1.5)	0.0742
Score Hématologique (Leuc., Plaq.)	0.3 (0.6)	0.3 (0.7)	0.1190
Score Diurèse	1.8 (2.2)	2.1 (2.3)	0.1354
Score Prothrombine	0.1 (0.3)	0.2 (0.4)	0.0451
Nombre de défaillances	3.6 (1.0)	3.8 (1.0)	0.3306
Score PaO ₂	1.9 (1.2)	2.2 (1.0)	3.10 ⁻⁶
Score Respiratoire (PaO ₂)	1.9 (1.2)	2.2 (1.0)	3.10 ⁻⁶

* : Test de Mann-Whitney si le nombre de valeurs prises par la variable est supérieur ou égal à 10, test du Chi² sinon.

2.3.6. Récapitulatif

Les variables pour lesquelles la différence standardisée est supérieure à 10 % sont les suivantes :

Figure 16 Différences standardisées calculées avant redressement des biais



3. IMPUTATION DES DONNEES MANQUANTES

Si l'on calculait un Propensity score à partir des données brutes, on serait amenés à exclure d'office 250 patients pour lesquels au moins l'une des caractéristiques initiales est manquante. On se prive ainsi de 23 % de l'échantillon.

Il conviendra donc d'imputer les valeurs manquantes pour calculer un PS sur l'ensemble des 1096 patients. On aura recours à des procédures d'imputation multiple.⁷

3.1. Logiciel choisi

On emploiera le logiciel IVEware (Imputation and Variance Estimation Software), un ensemble de routines C et FORTRAN qui peuvent être lancées à partir de SAS[®].

Le principe du logiciel est d'imputer pour chaque individu la valeur manquante conditionnellement aux autres variables observées pour l'individu. Pour chaque variable, on construit un modèle de régression multiple, puis on remplace la valeur manquante en tirant une valeur à partir de la distribution conditionnelle spécifiée par le modèle.⁸

Les variables peuvent appartenir à 5 catégories :

- Continues (ex : IGS à l'entrée en Réa)
- Binaires (ex : VIH+)
- Catégorielles (ex : PaO₂/FiO₂)
- Comptage (aucune variable de ce type dans les caractéristiques initiales)
- Mixtes : variable continue avec une densité de probabilité non nulle à 0 (ex : délai entre l'entrée à l'hôpital et l'entrée en réanimation : il est de 0 pour les entrées directes ou en urgence, ou continu sinon).

Le logiciel permet de limiter le nombre de prédicteurs employés dans chaque modèle de régression : soit en précisant directement leur nombre maximum, soit en ne rajoutant que les variables qui améliorent le R² d'une quantité donnée. Dans notre procédure d'imputation, une variable est incluse comme prédicteur si le R² augmente de 0,01.

Notons que le logiciel ne permet pas de prendre en compte le « clustering » des patients au niveau des unités de réanimation.

3.2. Nombre d'imputations

On impute les données manquantes 10 fois. On obtient ainsi 10 bases de données sans valeurs manquantes. On calcule pour chacune un propensity score, le propensity score global retenu en définitive étant tout simplement la moyenne des 10 autres.

4. CALCUL DU PROPENSITY SCORE

4.1. Méthodologie

Le propensity score sera estimé par régression logistique. La variable à prédire est la phase de recrutement, 46 variables peuvent être incluses en tant que variables qualitatives. Ces 46 variables ont été décrites au chapitre 2.2. On n'inclura pas dans les modèles les scores LODS et IGS II recalculés à partir de ces variables initiales, pour éviter les problèmes de multicollinéarité.

Certaines variables comprennent des modalités qui ont rarement été observées chez les patients. Lorsque cela est possible, on regroupe les modalités observées sur moins de 30 patients (phases « Avant » et « Après » confondues) :

- Pour le type de séjour, le traumatisme opéré (10 patients) sera regroupé avec la chirurgie non programmée, le traumatisme non opéré (4 patients) avec le séjour médical ;
- Pour la fréquence cardiaque, les fréquences inférieures à 30 (27 patients) et comprises entre 30 et 40 battements par minute (14 patients) ont été fusionnées en une seule catégorie ;
- Pour la pression artérielle systolique, les pressions comprises entre 200 et 239 mmHg (3 patients) ont été regroupées avec les pressions comprises entre 100 et 199 mmHg ;
- Pour la leucocytémie, les comptages supérieurs à 50 (21 patients) ont été regroupés avec ceux compris entre 20 et 50 milliards par litre ;
- Pour la diurèse, les valeurs de 10 L ou plus par jour (5 patients) ont été regroupées avec les valeurs comprises entre 1 et 10 litres.

On va comparer trois modèles :

- Modèle « Complet » (MC) : on inclut les 46 variables initiales, mais aucun terme d'interaction ;
- Modèle « Parcimonieux » (MP) : on sélectionne les variables par une procédure « stepwise » avec un seuil de sélection de 10 %, puis on rajoute les variables qui n'auraient pas été incluses mais dont la différence standardisée est supérieure à 10 % ;
- Modèle avec « Interactions » (MI) : on sélectionne les variables par procédure « stepwise », on sélectionne ensuite les termes d'interaction entre les variables précédentes, puis on ajoute les variables dont la différence standardisée est supérieure à 10 %.

Les modèles sont estimés sur les 10 bases issues de la procédure d'imputation multiple, le propensity score final est donné par la moyenne des 10 PS issus de chacun des modèles.

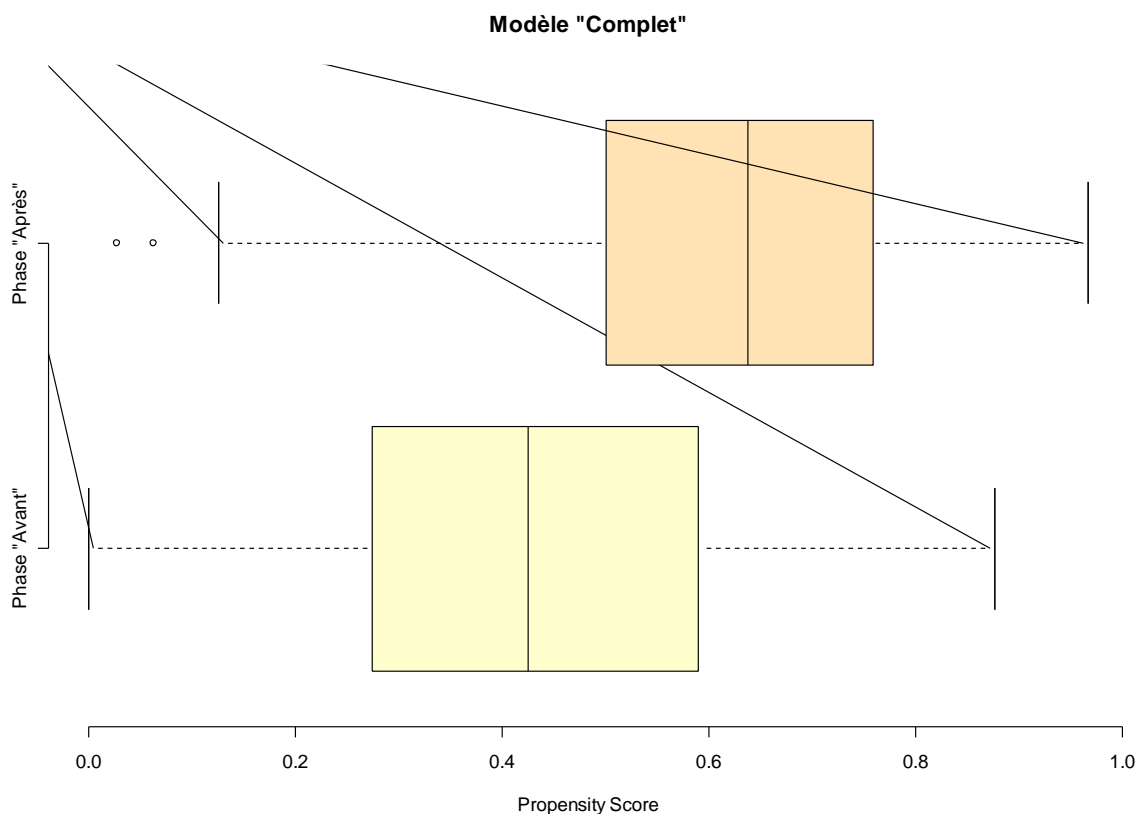
4.2. Modèle « Complet »

Ce modèle a un AIC compris entre 1443 et 1457 (selon les imputations) et classe correctement entre 74,2 % et 75,0 % des patients.

4.2.1. Propensity scores

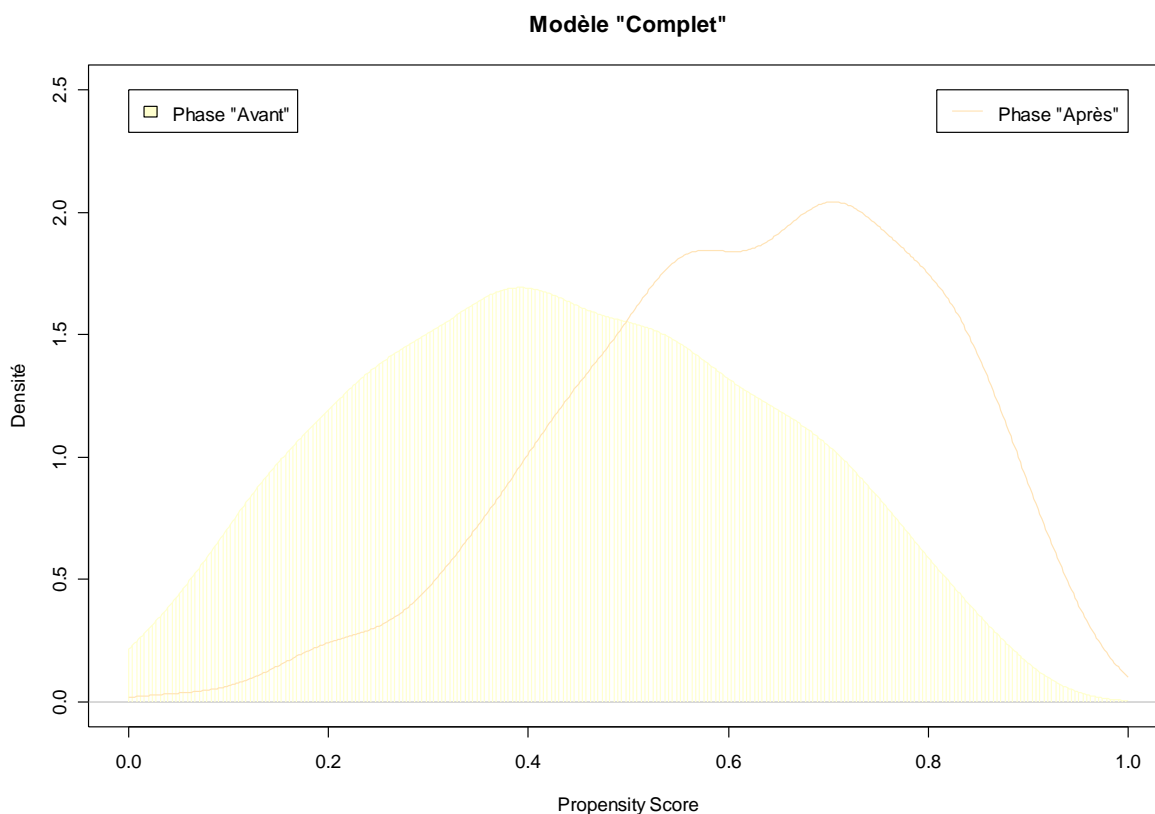
La répartition des propensity scores est donnée dans les deux figures suivantes :

Figure 17 Répartition des propensity scores en phase « Avant » et en phase « Après » - Modèle MC



Sans surprise, les patients en Phase « Après » ont tendance à avoir des PS supérieurs à ceux en phase « Avant ». Les PS se superposent cependant relativement bien pour les valeurs centrales (entre 40 % et 60 %).

Figure 18 Densité de probabilité des propensity scores en phase « Avant » et en phase « Après » - MC



Les résultats de ce modèle sont encourageants, les deux courbes se superposent relativement bien.

Si l'on s'intéresse à la répartition des patients par quintiles de PS :

Tableau 7 : Répartition des patients par quintiles de PS en fonction de la phase de recrutement - MC

Quintile	1 ^{er}	2 ^e	3 ^e	4 ^e	5 ^e
Phase « Avant »	177	121	96	74	41
Phase « Après »	42	98	124	145	178
Total	219	219	220	219	219

Les effectifs sont plus équilibrés sur les deuxième et troisième quintile. Néanmoins, même sur les quantiles extrêmes, on a au moins 40 patients de l'une des phases.

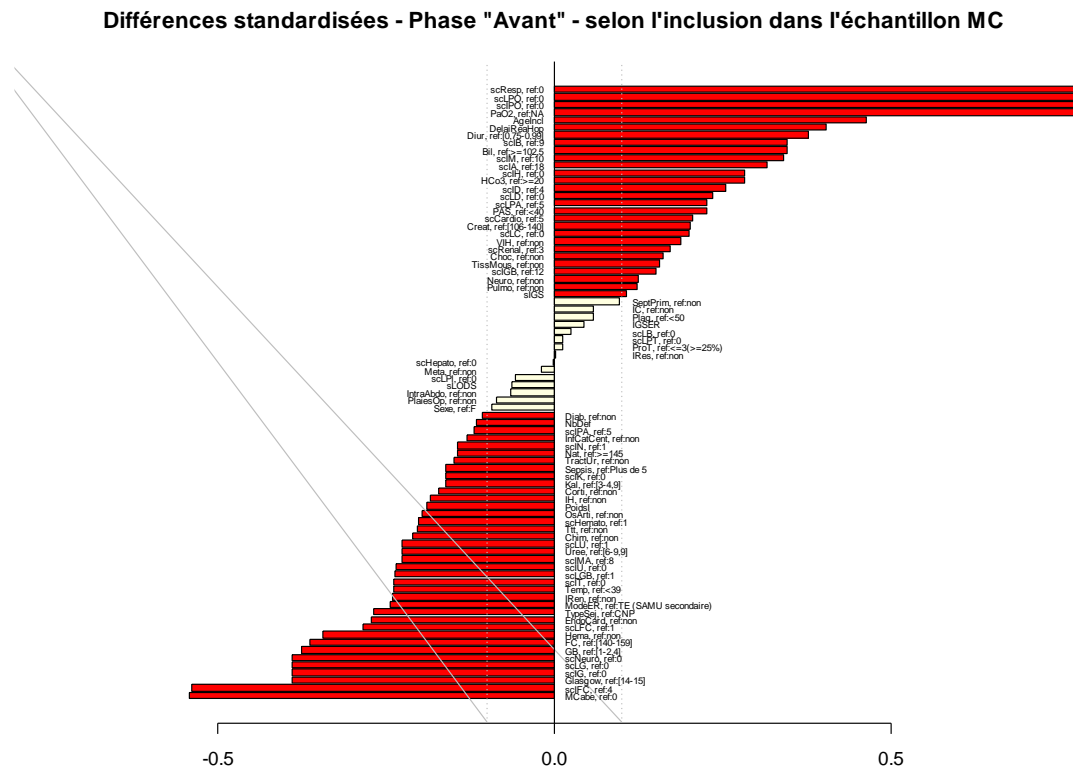
4.2.2. Comparabilité des patients appariés et non appariés

On procède à un appariement optimal à l'aide de la macro SAS[®] « match »¹. Cette macro cherche à appairer les patients des deux phases en fonction de leur propensity score de manière à minimiser la distance totale entre les PS des patients appariés (calculée comme la valeur absolue de la différence entre les deux PS du couple de patients appariés).

Le modèle « Complet » permet d'appairer 420 patients de la phase « Avant » à des patients de la phase « Après ». On obtient ainsi un échantillon de 840 patients, soit 76,6 % de l'échantillon initial.

On peut s'intéresser au processus de sélection des patients en calculant les différences standardisées, pour chaque phase de l'étude, entre les patients appariés et ceux exclus de l'échantillon.

Figure 19 Différences standardisées des patients de phase « Avant » en fonction de l'inclusion dans MC



Les différences standardisées négatives correspondent aux caractéristiques plus souvent observées sur les patients appariés.

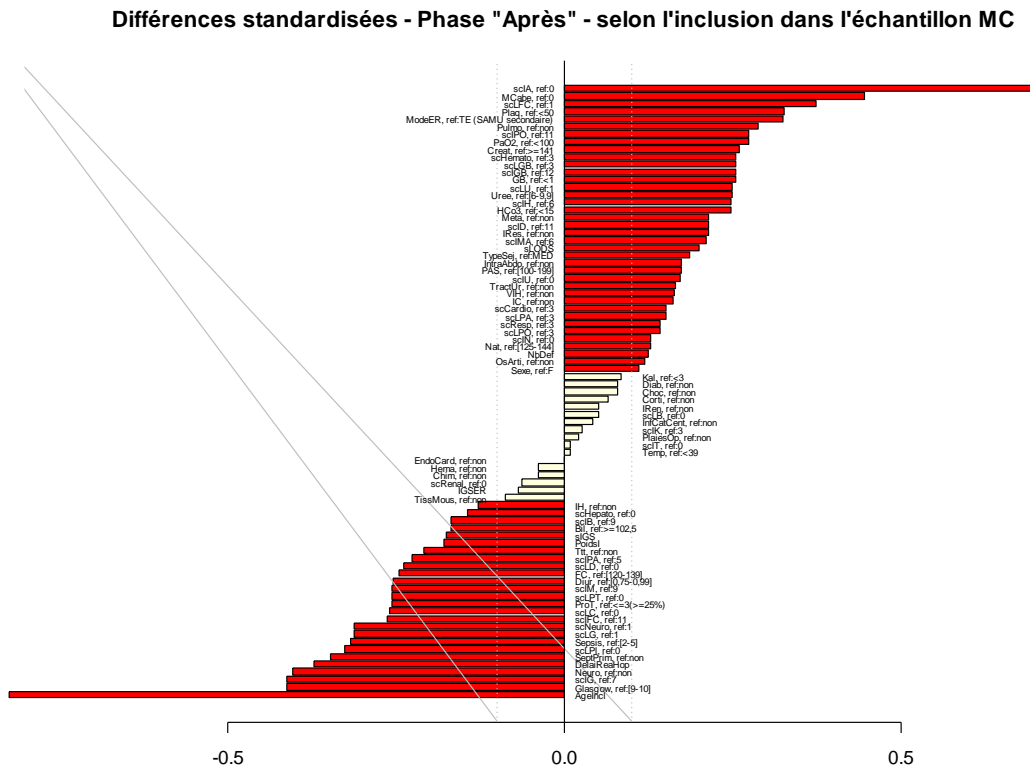
Ainsi, les patients de phase « Avant » retenus dans l'échantillon apparié ont plus souvent un indice de McCabe de 0 (absence de comorbidités), un score de Glasgow de 14 ou 15 (réactivité maximale). Ils ont une proportion moindre de défaillances neurologiques (définies par le LODS), de maladies hématologiques, de sites d'infection endo-cardiovasculaires...

¹ Mayo Clinic College of Medicine. <http://mayoresearch.mayo.edu/mayo/research/biostat/upload/match.sas>

Si les patients appariés de la phase « Avant » ont tendance à être plus jeunes et moins graves (au sens de l'IGS II), l'appariement sélectionne toutefois des patients avec plus de défaillances d'organe (définies par le LODS) et notamment ventilés.

Les appariements en phase « Après » sont symétriques à ceux en phase « Avant » : les patients appariés sont plus âgés, ont plus fréquemment un score de Glasgow entre 9 et 10, ont un IGS II et un score de McCabe supérieurs...

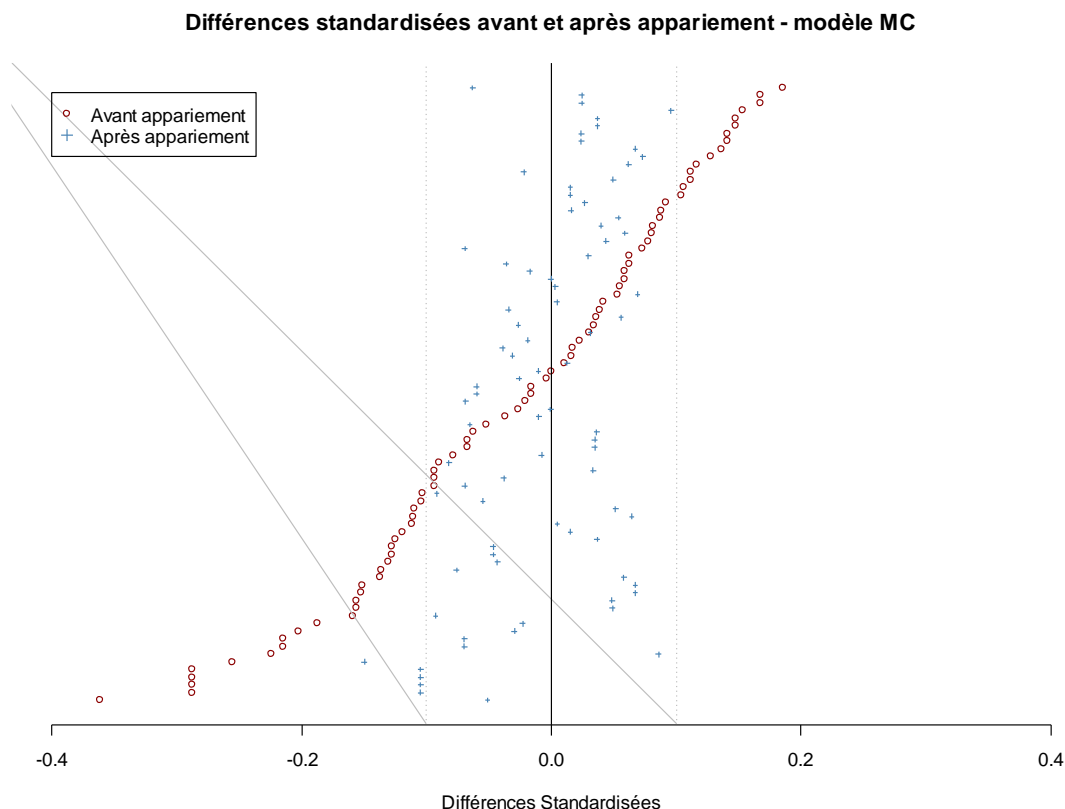
Figure 20 Différences standardisées des patients de phase « Après » en fonction de l'inclusion dans MC



4.2.3. Réduction des biais de recrutement

On a vu que les patients appariés l'étaient de manière à réduire les écarts de caractéristiques entre les deux groupes. Le calcul des différences standardisées sur l'échantillon apparié permet de juger de l'amplitude de la réduction de ces biais de recrutement :

Figure 21 Différences standardisées avant et après appariement sur le PS du modèle MC



La moyenne de la valeur absolue des différences standardisées passe de 11,21 % sur l'échantillon total à 4,69 % sur l'échantillon apparié.

Les différences standardisées sont presque toutes en dessous du seuil de 10 %. Seules 5 variables le dépassent :

- Le score IGS II lié à l'âge, les scores de 18 (au moins 80 ans) restent plus nombreux en phase « Avant » : 42 patients contre 25 (sur 420 patients chacun). Si les âges des patients des deux phases sont comparables ($p = 0.3312$, test non paramétrique de Mann-Whitney) ainsi que la répartition globale des scores IGS II liés à l'âge ($p = 0.2552$, test du χ^2), la proportion de patients de 80 ans ou plus reste significativement différente entre les deux phases ($p = 0.0409$, test exact de Fisher) ;
- La proportion de patients non ventilés ($\text{PaO}_2 = \text{NA}$), et par ricochet les 2 scores LODS et le score IGS correspondants est plus importante en phase « Avant » : 29 patients contre 19. Néanmoins, cette différence n'est pas statistiquement significative ($p = 0.1394$, test exact de Fisher).

4.3. Modèle « Parcimonieux »

4.3.1. Construction du modèle

On teste les variables à inclure dans le modèle une par une, la p-value maximale permettant l'entrée dans le modèle et la p-value minimale permettant d'éliminer une variable déjà incluse du modèle sont toutes deux de 10 %. Ce seuil, supérieur au seuil traditionnel de 5 %, a été retenu de manière à obtenir un modèle surparamétré.

Toute variable apparaissant dans au moins l'un des 10 modèles est retenue. Il s'agit des variables suivantes :

- Parmi les caractéristiques initiales : l'âge, le score de McCabe, le délai entre l'entrée à l'hôpital et l'entrée en réanimation, le délai entre l'entrée en réanimation et le début du sepsis.
- Parmi les variables biologiques : le score de Glasgow, la pression partielle en O₂, la prothrombine, la fréquence cardiaque, la diurèse apparaissent dans les 10 modèles. La créatinine, l'urée, la bilirubine apparaissent sur certains des modèles.
- Parmi les sites d'infection : les sites endo-cardiovasculaire et neurologique sont toujours sélectionnés, le site pulmonaire ne l'est qu'occasionnellement.

On rajoute à ce modèle les variables dont les différences standardisées dépassent 10 % : le mode d'entrée en réanimation, l'insuffisance rénale, le cancer métastasé, la bicarbonatémie, les sites d'infection du tractus urinaire et des tissus mous.

Le modèle a un AIC compris entre 1405 et 1417 (selon les imputations) et classe correctement entre 72,8 % et 73,4 % des patients.

4.3.2. Propensity Scores

Si l'on s'intéresse à la répartition des patients par quintiles de PS :

Tableau 8 : Répartition des patients par quintiles de PS en fonction de la phase de recrutement - MP

Quintile	1 ^{er}	2 ^e	3 ^e	4 ^e	5 ^e
Phase « Avant »	172	128	87	75	47
Phase « Après »	47	91	133	144	172
Total	219	219	220	219	219

Les effectifs ne sont équilibrés sur aucun des quantiles, mais sur les deux quantiles extrêmes, on a 47 patients de la phase la moins représentée.

La répartition des propensity scores est illustrée dans les deux figures suivantes :

Figure 22 Répartition des propensity scores en phase « Avant » et en phase « Après » - MP

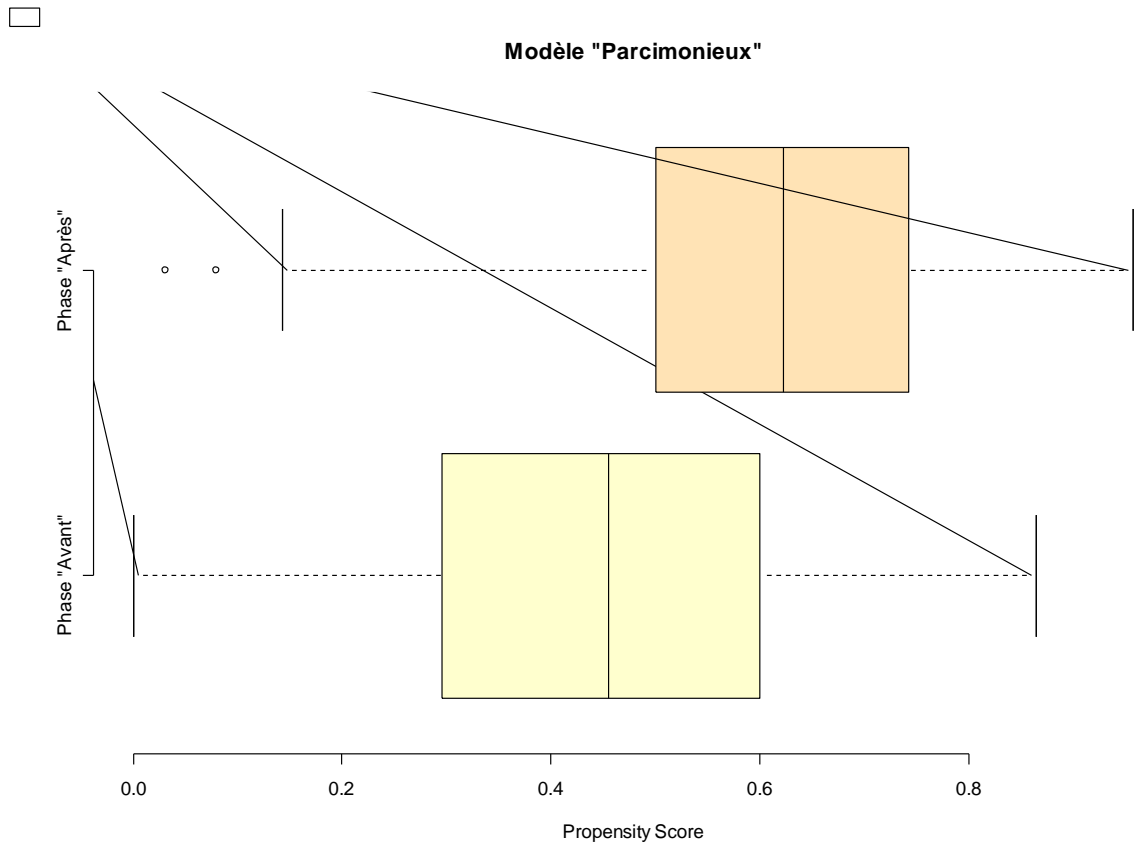
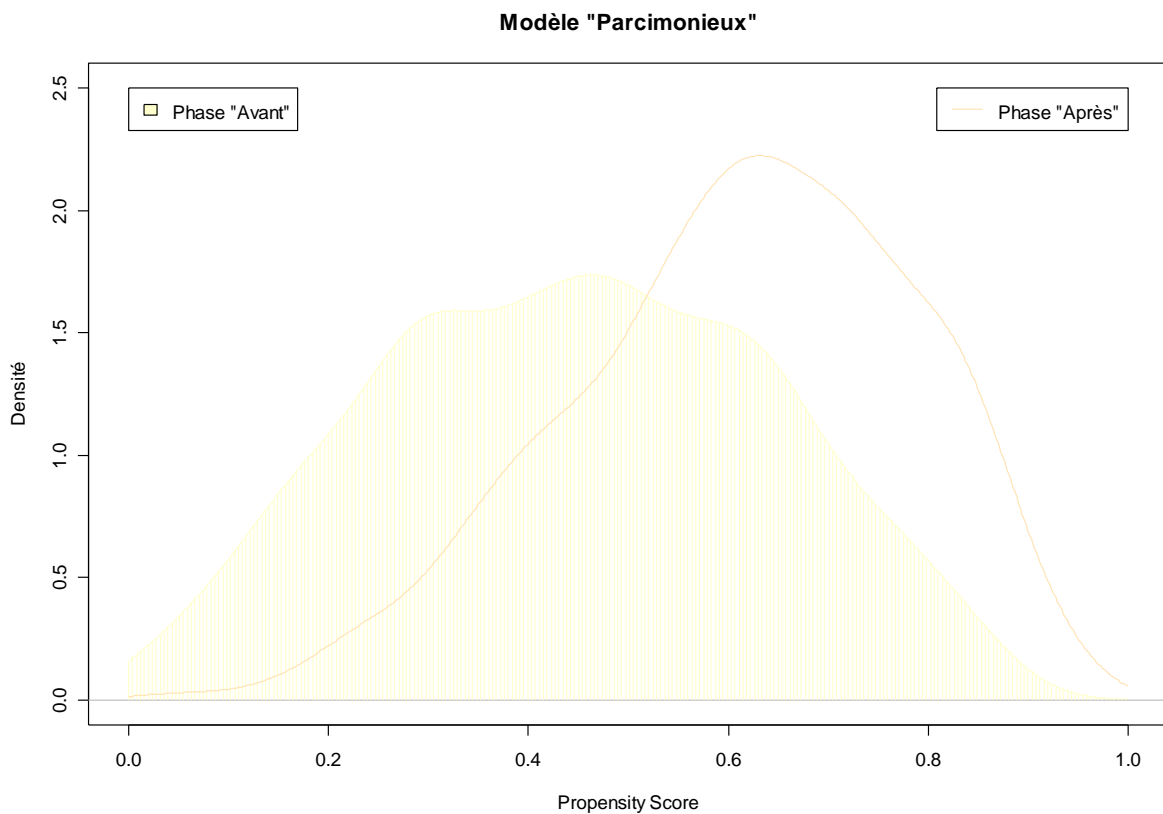


Figure 23 Densité de probabilité des propensity scores en phase « Avant » et en phase « Après » - MP

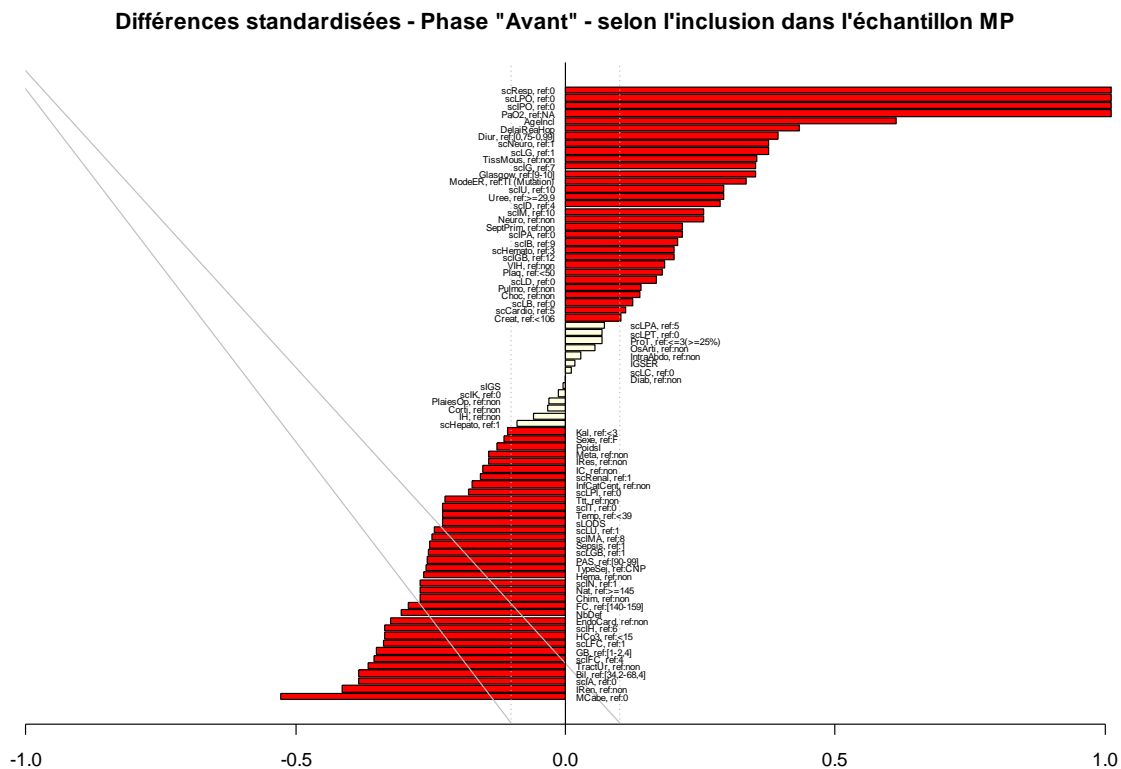


4.3.3. Comparabilité des patients appariés et non appariés

En suivant la même procédure que pour le modèle « Complet », Le modèle « Parcimonieux » permet d'apparier 435 patients de la phase « Avant » à des patients de la phase « Après ». On obtient ainsi un échantillon de 870 patients, 30 de plus qu'avec le modèle « Complet », soit 79,4 % de l'échantillon initial.

En phase « Avant », les patients appariés et non appariés diffèrent sur les caractéristiques suivantes :

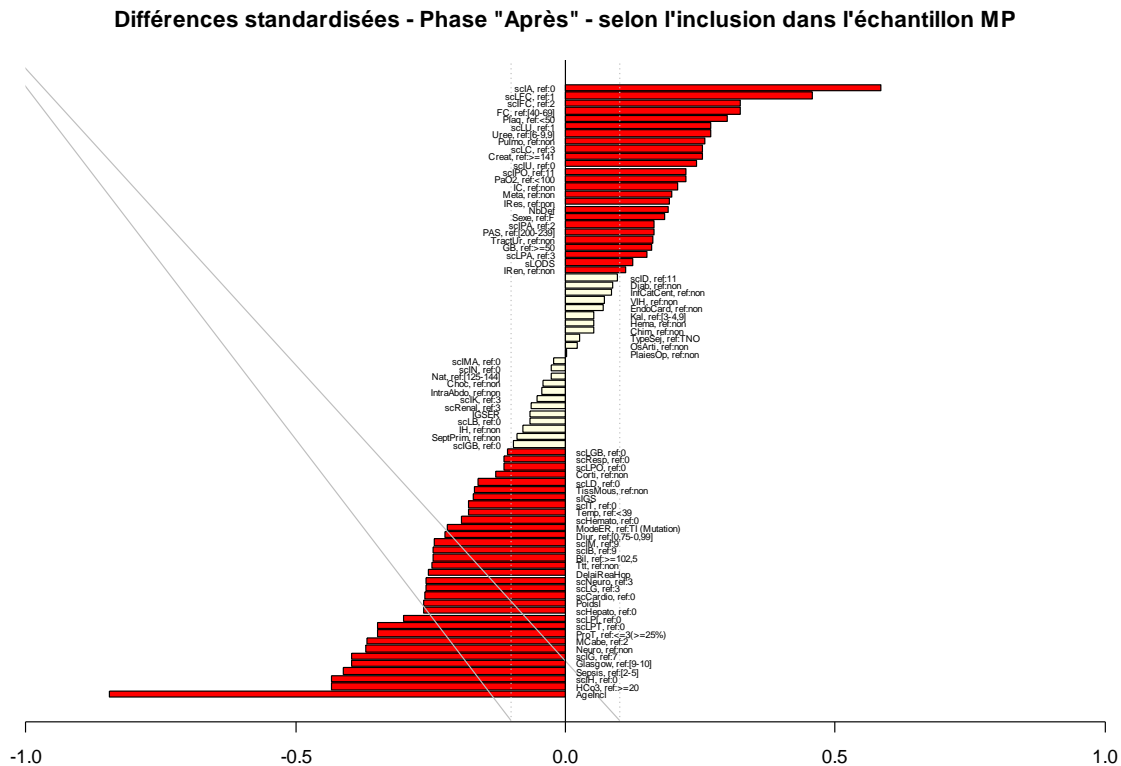
Figure 24 Différences standardisées des patients de phase « Avant » en fonction de l'inclusion dans MP



La différence standardisée la plus importante a trait à l'absence de ventilation mécanique : 6 % des patients appariés contre 46 % des non appariés ! Les différences sont comparables à celles relevées sur le modèle « Complet » : les patients appariés ont tendance à avoir plus de défaillances (définies par le LODS), moins de comorbidités (définies par le score de McCabe), à être plus jeunes... Dans ce modèle, néanmoins, l'IGS II des patients de phase « Avant » inclus et non inclus est comparable.

En phase « Après » (Figure 25), les patients appariés sont plus âgés que les non appariés, ont un IGS II à l'entrée dans l'étude plus important, un LODS et un nombre de défaillances moins important.

Figure 25 Différences standardisées des patients de phase « Après » en fonction de l'inclusion dans MP



4.3.4. Réduction des biais de recrutement

La moyenne de la valeur absolue des différences standardisées passe de 11,21 % sur l'échantillon total à 5,38 % sur l'échantillon apparié (4,69 % pour le modèle « Complet »).

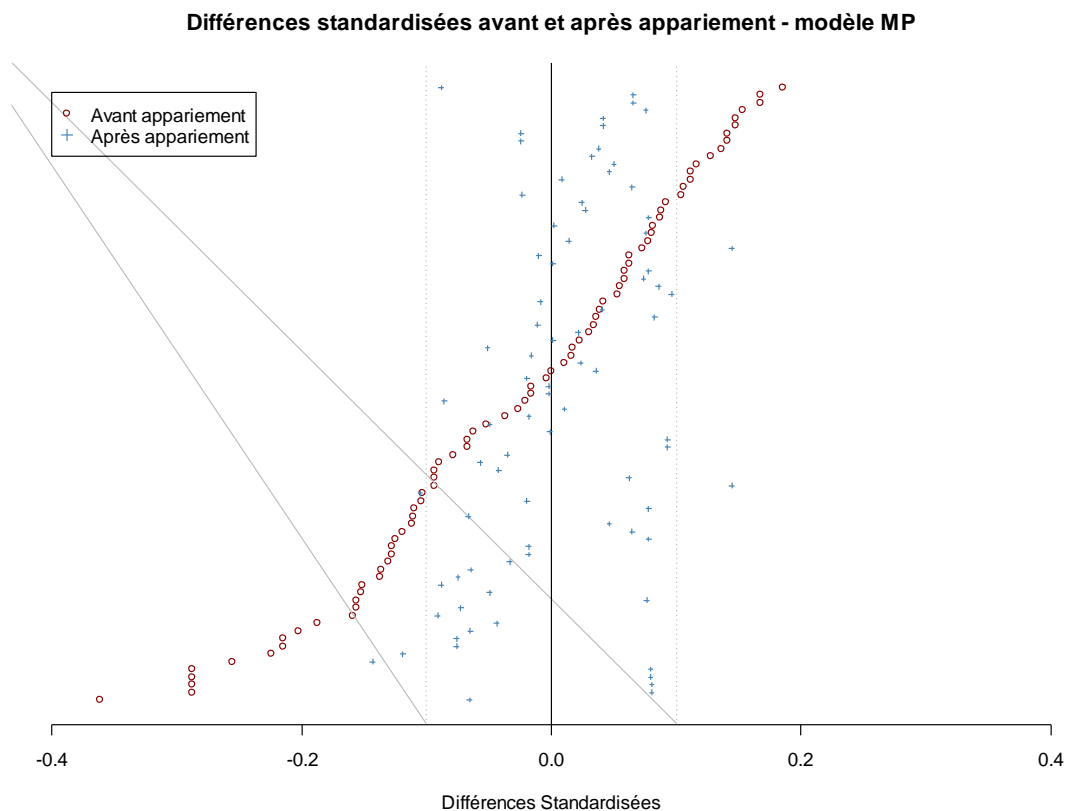
5 variables ont une différence standardisée supérieure à 10 % en valeur absolue :

- La pression artérielle et, par ricochet, le sous-score IGS II lié à cette variable. On observe plus de PAS comprises entre 100 et 199 en phase « Après », et par conséquent plus de sous-scores IGS II nuls pour cette variable (6,00 et 9,89 % en phase « Avant » et « Après respectivement). Il est intéressant de noter qu'avant appariement, la PAS était équilibrée entre les deux phases. Si la répartition globale des PAS n'est pas significativement différente entre les deux phases ($p = 0.2508$, test du χ^2), la répartition des PAS comprises entre 100 et 199 spécifiquement l'est ($p = 0.0440$, test exact de Fisher). La PAS rentre également dans le calcul du LODS, mais avec un système de seuils différent : une PAS comprise entre 90 et 99 est également considérée comme non pathologique. Les patients en phase « Avant » ayant plus souvent une PAS dans cet intervalle que les patients en phase « Après », leur score LODS s'en trouve équilibré.
- Le sous-score IGS II lié à l'âge est plus fréquemment égal à 18 en phase « Avant ». On retrouve ici le problème rencontré avec le modèle « Complet ». Si la répartition globale des scores reste comparable ($p = 0.4179$, test du χ^2), la proportion de 80 ans ou plus diffère entre les deux phases ($p = 0.0469$, test exact de Fisher).
- Le score McCabe de 3 (maladie mortelle à court terme) est plus fréquemment rencontré en phase « Avant » : 7,35 % contre 4,55 % en phase « Après ». Cependant, la répartition des scores n'est pas significativement différente dans les deux phases, que l'on teste le McCabe

global ($p = 0.1938$, test du χ^2) ou seulement la présence d'une maladie mortelle à court terme ($p = 0.1026$, test exact de Fisher).

- L'absence de défaillance rénale (selon le LODS, qui s'intéresse à l'urée, la créatinine et la diurèse) est plus fréquente en phase « Avant » : 12,21 % des patients, contre 8,99 % en phase « Après ». Cette différence n'est cependant pas statistiquement significative ($p = 0,1514$).

Figure 26 Différences standardisées avant et après appariement sur le PS du modèle MP



Le modèle « Parcimonieux » permet de conserver plus de patients dans l'échantillon apparié que le modèle « Complet », mais au coût d'un moindre équilibre entre les caractéristiques des deux phases : si les patients sont désormais comparables sur la présence d'une ventilation mécanique, la pression artérielle pose problème, ainsi que, dans une moindre mesure, le score de McCabe et les défaillances rénales. De plus, on observe toujours plus de patients âgés d'au moins 80 ans en phase « Avant ».

4.4. Modèle avec « Interactions »

4.4.1. Construction du modèle

Ce modèle complète le modèle « Parcimonieux » en rajoutant les termes d'interaction suivants (les interactions retenues étaient celles apparaissant dans au moins 3 des 10 modèles) :

- Interactions entre le délai de l'entrée à l'hôpital à l'entrée en réanimation et :
 - o Le score de McCabe ;
 - o La pression partielle en O₂ ;
 - o Le site d'infection endo-cardiovasculaire ;
 - o La bilirubine.
- Interactions entre le score de McCabe et :
 - o Le délai entre l'entrée en réanimation et l'apparition du sepsis ;
 - o La pression partielle en O₂ ;
 - o La bilirubine.
- Interaction entre la prothrombine et :
 - o La créatinine ;
 - o La bilirubine.
- Interaction entre l'âge et le site d'infection endo-cardiovasculaire.

Le modèle a un AIC compris entre 1393 et 1411 (selon les imputations) et classe correctement entre 77,6 % et 78,4 % des patients.

4.4.2. Propensity scores

Si l'on s'intéresse à la répartition des patients par quintiles de PS :

Tableau 9 : Répartition des patients par quintiles de PS en fonction de la phase de recrutement - MI

Quintile	1 ^{er}	2 ^e	3 ^e	4 ^e	5 ^e
Phase « Avant »	188	132	87	64	38
Phase « Après »	31	87	133	155	181
Total	219	219	220	219	219

Les patients de la phase « Avant » se concentrent majoritairement dans le premier quintile, ceux de la phase « Après » dans le dernier. De manière symétrique, on ne retrouve que 31 patients de la phase « Après » dans le premier quintile et 38 de la phase « Avant » dans le dernier.

L'intégration des interactions discrimine mieux entre les patients des deux phases. On observe plus de patients de phase « Avant » avec des PS inférieurs à 20 %, le recouplement des deux phases est moindre.

Figure 27 Répartition des propensity scores en phase « Avant » et en phase « Après » - MI

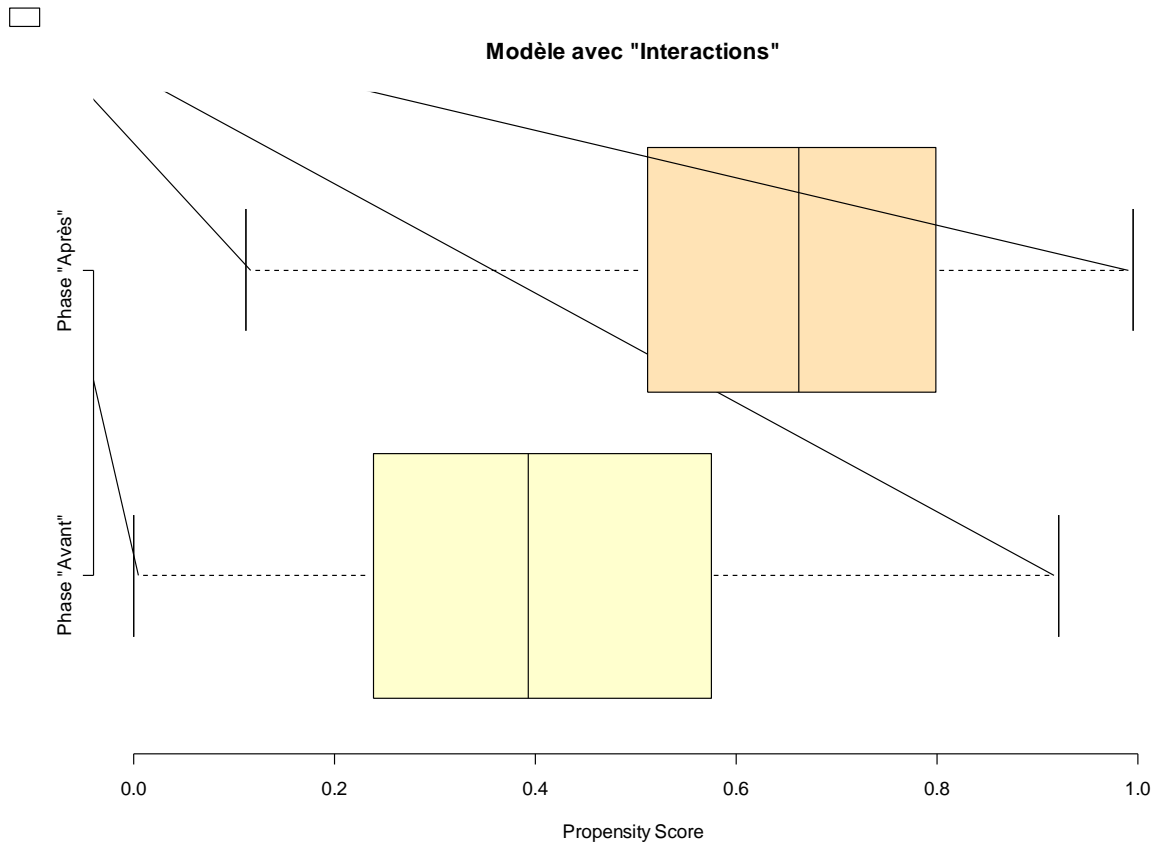
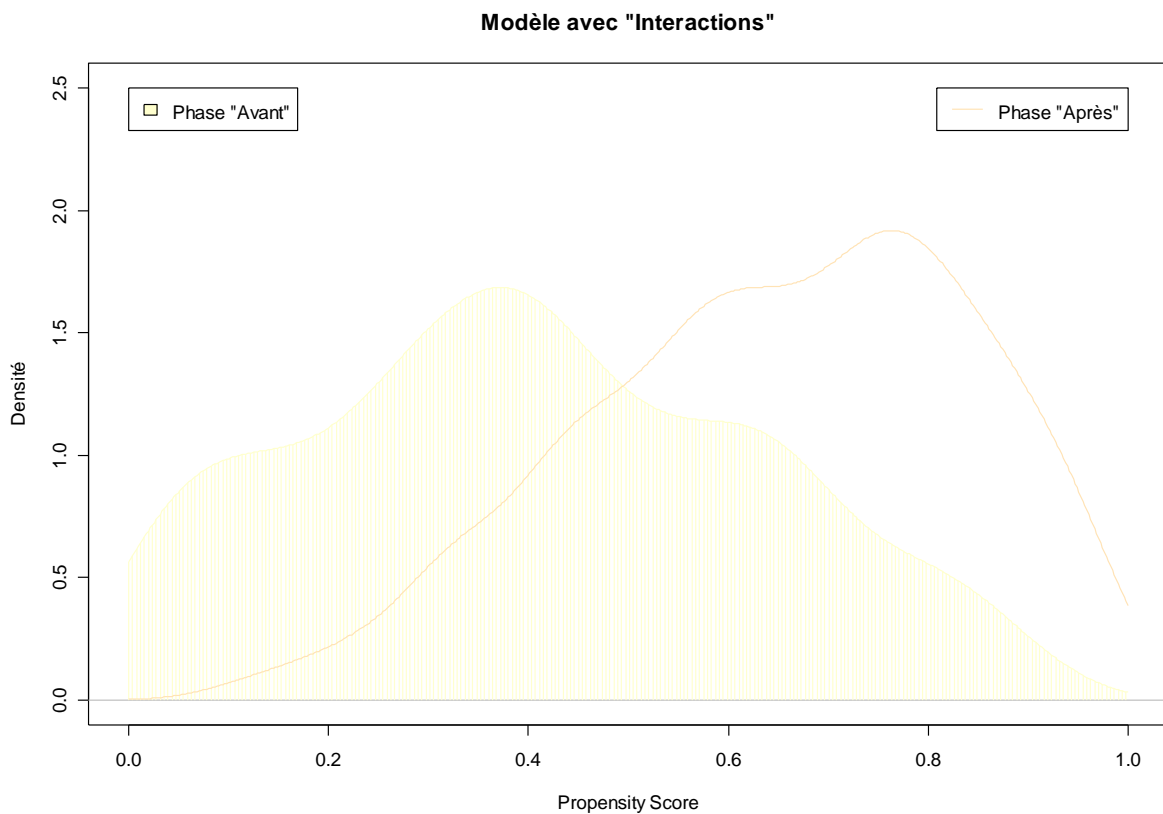


Figure 28 Densité de probabilité des propensity scores en phase « Avant » et en phase « Après » - MI



4.4.3. Comparabilité des patients appariés et non appariés

Le modèle avec « Interactions » permet d'apparier 374 patients, on obtient un échantillon final de 748 patients, soit 68.25 % de l'échantillon initial. Ce modèle est celui retenant le moins de patients dans l'échantillon apparié.

Le processus de sélection des patients pour l'appariement relève de la même logique que dans les deux autres modèles : les patients de la phase « Avant » (Figure 29) sont sélectionnés parmi ceux présentant le plus de défaillances et ayant un LODS plus sévère, ventilés, moins âgés. Les patients de la phase « Après » (Figure 30) sont au contraire sélectionnés parmi les plus âgés, avec moins de défaillances, avec plus de comorbidités.

Figure 29 Différences standardisées des patients de phase « Avant » en fonction de l'inclusion dans MP

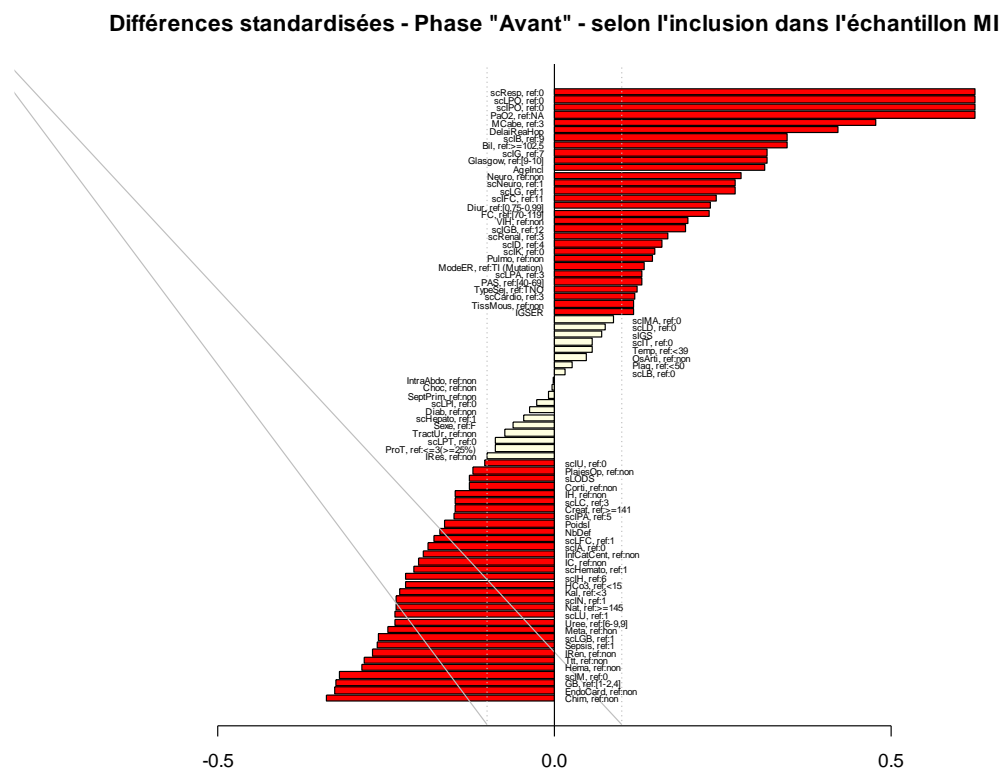
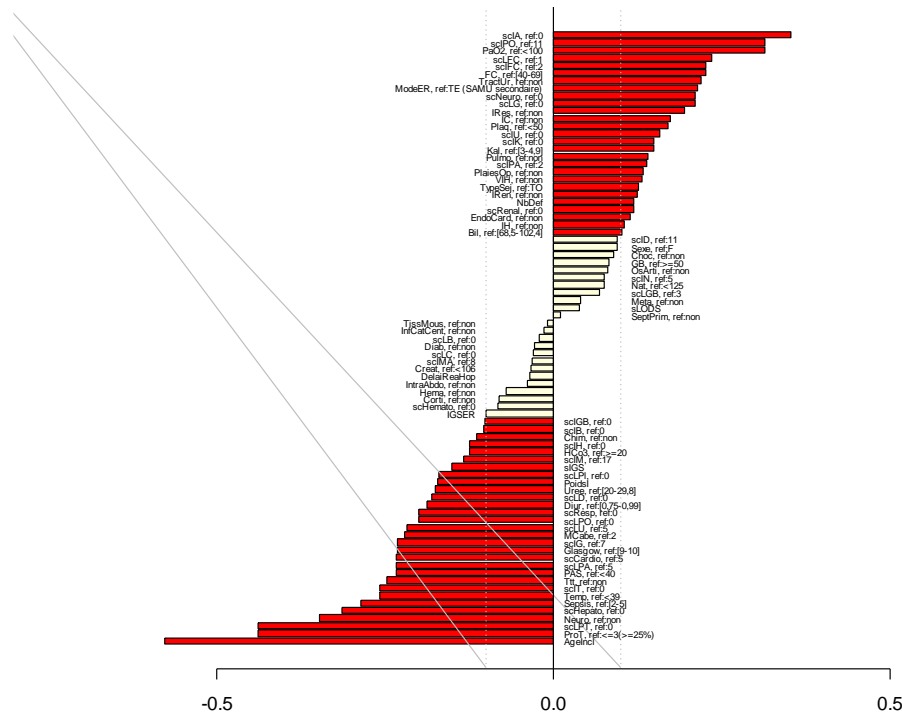


Figure 30 Différences standardisées des patients de phase « Avant » en fonction de l'inclusion dans MP

Différences standardisées - Phase "Après" - selon l'inclusion dans l'échantillon MI



4.4.4. Réduction des biais de recrutement

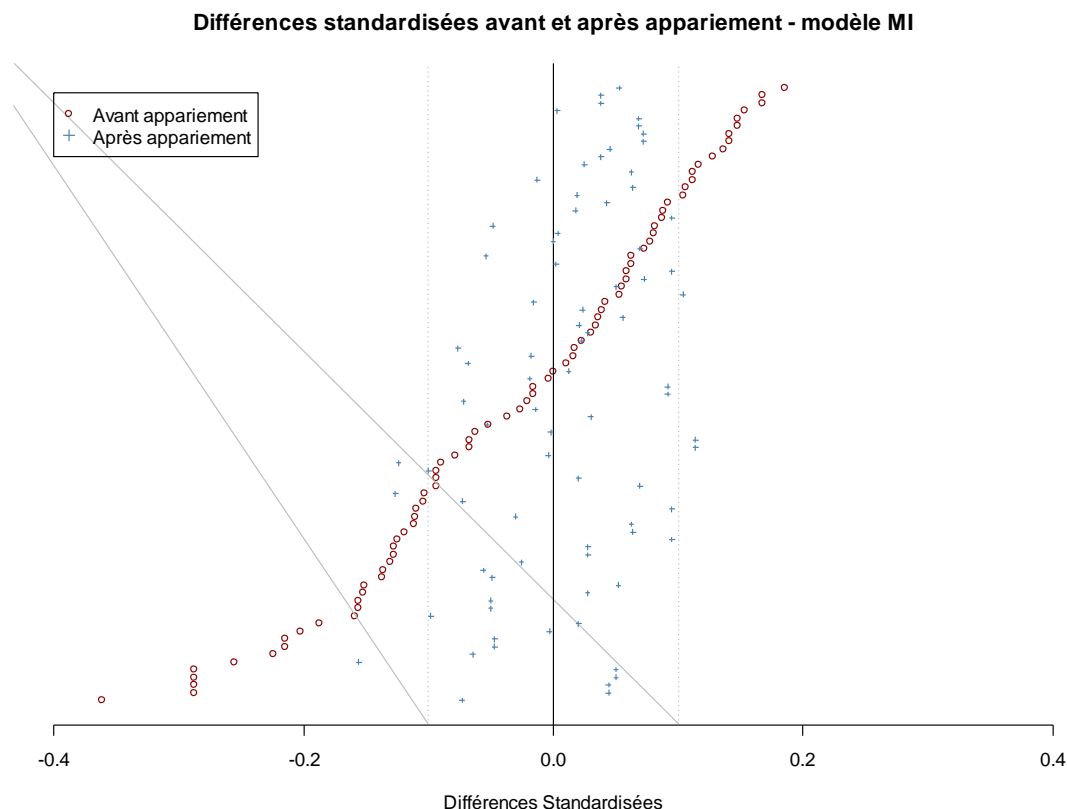
La moyenne de la valeur absolue des différences standardisées passe de 11,21 % sur l'échantillon total à 5,19 % sur l'échantillon apparié (4,69 % pour le modèle « Complet » et 5,38 % sur le modèle « Parcimonieux »). On obtient donc un modèle de même niveau de performance que le modèle « Parcimonieux », mais avec moins de patients appariés.

Ce sont cette fois 6 variables dont la différence standardisée est supérieure à 10 % (Figure 31) :

- Comme dans les modèles précédents, on retrouve une proportion supérieure de patients de 80 ans et plus en phase « Avant ». Si le sous-score IGS II lié à l'âge n'est pas significativement différent entre les deux phases pris dans sa globalité ($p = 0.2535$, test du χ^2), la différence devient significative si l'on s'intéresse uniquement à sa modalité extrême ($p = 0.0459$, test exact de Fisher).
- On retrouve plus de défaillances rénales (définies par le LODS) en phase « Après ». La différence n'est cependant pas significative ($p = 0.1065$, test exact de Fisher). Le modèle « Parcimonieux » donnait le même résultat.
- On retrouve plus de chocs septiques en phase « Après » : 95,45 % contre 92,51 % en phase « Avant ». La différence n'est pas statistiquement significative ($p = 0.1233$, test exact de Fisher).
- Les valeurs non pathologiques de natrémie (entre 125 et 144 mEq/L) se rencontrent plus souvent en phase « Après » : 87,63 % des patients contre 83,65 % en phase « Avant ». En conséquence, le sous-score IGS II lié à la natrémie est plus fréquemment nul en phase « Après ». Cette différence n'est pas statistiquement significative ($p = 0.1434$, test exact de Fisher).

- Enfin, la différence standardisée supérieure à 10 % sur le type de séjour traduit simplement le fait que les deux seuls traumatismes non opérés de l'échantillon apparié se rencontrent en phase « Avant ».

Figure 31 Différences standardisées avant et après appariement sur le PS du modèle MI



4.5. Choix du modèle

Le Tableau 10 résume le nombre de patients appariés, l'équilibre moyen de l'échantillon apparié, et la valeur absolue des différences standardisées les plus importantes de chaque modèle.

Tableau 10 : Caractéristiques principales des trois modèles de PS

Caractéristique	Modèle MC	Modèle MP	Modèle MI
Patients appariés	840	870	748
Equilibre de l'échantillon*	2.39	2.08	2.16
Age ≥ 80	14.98 %	14.31 %	15.56 %
PaO ₂	10.46 %	8.03 %	4.42 %
PAS	6.92 %	14.44 %	6.95 %
McCabe	8.59 %	11.89 %	6.45 %
Score Rénal	9.17 %	10.49 %	12.64 %
Choc septique	8.21 %	5.72 %	12.39 %
Natrémie	3.46 %	9.28 %	11.39 %
Type de Séjour	6.91 %	9.61 %	10.37 %

* défini comme le rapport entre la moyenne de la valeur absolue des différences standardisées avant appariement et la moyenne après appariement

Le modèle MI discrimine le mieux les patients en fonction de leur phase de traitement. De ce fait, il apparie au final moins de patients, puisque les PS se concentrent plus sur les extrêmes. Le modèle MC, intégrant toutes les variables, permet d'obtenir le meilleur équilibre sur l'échantillon final.

Les trois modèles aboutissent à un échantillon qui n'est pas équilibré sur les plus de 80 ans.

Le modèle MC n'équilibre pas de surcroît sur la présence d'une ventilation.

Le modèle MP n'équilibre pas sur la pression artérielle systolique, le score de McCabe et la présence d'une défaillance rénale.

Le modèle MI n'équilibre pas sur la présence d'une défaillance rénale, sur la présence d'un choc septique, sur la natrémie et, dans une moindre mesure, sur le type de séjour.

Au vu de l'ensemble de ces caractéristiques, il paraît plus pertinent de travailler sur l'échantillon issu du modèle MC, en veillant à neutraliser les deux facteurs de confusion potentiels que sont l'âge et la présence de ventilation.

5. ANNEXES

5.1. Références

-
- ¹ Bernard GR, Macias WL, Joyce DE, Williams MD, Bailey J, Vincent JL. Safety assessment of drotrecogin alfa (activated) in the treatment of adult patients with severe sepsis. *Crit Care*. 2003 Apr;7(2):155-63. Epub 2003 Feb 28.
- ² MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AM. A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies. *Health Technol Assess*. 2000;4(34):1-154.
- ³ Klungel OH, Martens EP, Psaty BM, Grobbee DE, Sullivan SD, Stricker BH, Leufkens HG, de Boer A. Methods to assess intended effects of drug treatment in observational studies are reviewed. *J Clin Epidemiol*. 2004 Dec;57(12):1223-31.
- ⁴ Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41-55.
- ⁵ Cochran W. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* 1968;24:205-13.
- ⁶ Rosenbarn PR, Rubin DB. Reducting bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association* 1984;79:516-24.
- ⁷ Schafer, J.L. (1997), *Analysis of Incomplete Multivariate Data*, New York: Chapman and Hall.
- ⁸ Raghunathan TE, Lepkowski JM, Van Hoewyk J, Solenberger P. A Multivariate technique for Multiply Imputing Missing Values Using a Sequence of Regression Models. *Survey Methodology* 2001;27(1).

5.2. Index des tableaux

Tableau 1 : Méthodes employées dans les études observationnelles revues par Klungel et al. ³	9
Tableau 2 : Mode d'entrée en réanimation en fonction de la phase de recrutement	19
Tableau 3 : Délai entre l'apparition du sepsis et l'entrée en réa en fonction de la phase de recrutement	19
Tableau 4 : Caractéristiques du bilan médical avec une différence standardisée supérieure à 10 %	20
Tableau 5 : Caractéristiques des composantes du score IGS II	24
Tableau 6 : Caractéristiques des composantes du score LODS	26
Tableau 7 : Répartition des patients par quintiles de PS en fonction de la phase de recrutement - MC	31
Tableau 8 : Répartition des patients par quintiles de PS en fonction de la phase de recrutement - MP	35
Tableau 9 : Répartition des patients par quintiles de PS en fonction de la phase de recrutement - MI	40
Tableau 10 : Caractéristiques principales des trois modèles de PS	44

5.3. Index des illustrations

Figure 1 Structure hiérarchique des observations de l'étude PREMISS	7
Figure 2 Proportion de valeurs manquantes par variable	13
Figure 3 Répartition du nombre de valeurs manquantes par patient	14
Figure 4 Nombre moyen de valeurs manquantes si la variable est manquante	15
Figure 5 Moyenne conditionnelle de valeurs manquantes et pourcentage de valeurs manquantes	16
Figure 6 Différences standardisées calculées sur les caractéristiques du patient à son bilan initial	17
Figure 7 Distribution de l'âge des patients en fonction de la phase de recrutement	18
Figure 8 Délai entre l'entrée à l'hôpital et l'entrée en réanimation et phase de recrutement	18
Figure 9 Différences standardisées calculées sur les caractéristiques du patient à son bilan médical	20
Figure 10 Différences standardisées calculées sur les caractéristiques biologiques du patient	21
Figure 11 PaO ₂ , Score de Glasgow et Diurèse en fonction de la phase de recrutement	22
Figure 12 Bilirubine, Fréquence cardiaque et Prothrombine en fonction de la phase de recrutement	22
Figure 13 Urée, Bicarbonatémie et Créatinine en fonction de la phase de recrutement	23
Figure 14 Différences standardisées calculées sur les composantes de l'IGS II	24
Figure 15 Différences standardisées calculées sur les composantes du LODS	25
Figure 16 Différences standardisées calculées avant redressement des biais	27
Figure 17 Répartition des propensity scores en phase « Avant » et en phase « Après » - Modèle MC	30
Figure 18 Densité de probabilité des propensity scores en phase « Avant » et en phase « Après » - MC	31
Figure 19 Différences standardisées des patients de phase « Avant » en fonction de l'inclusion dans MC	32
Figure 20 Différences standardisées des patients de phase « Après » en fonction de l'inclusion dans MC	33
Figure 21 Différences standardisées avant et après appariement sur le PS du modèle MC	34
Figure 22 Répartition des propensity scores en phase « Avant » et en phase « Après » - MP	36
Figure 23 Densité de probabilité des propensity scores en phase « Avant » et en phase « Après » - MP	36
Figure 24 Différences standardisées des patients de phase « Avant » en fonction de l'inclusion dans MP	37
Figure 25 Différences standardisées des patients de phase « Après » en fonction de l'inclusion dans MP	38
Figure 26 Différences standardisées avant et après appariement sur le PS du modèle MP	39
Figure 27 Répartition des propensity scores en phase « Avant » et en phase « Après » - MI	41
Figure 28 Densité de probabilité des propensity scores en phase « Avant » et en phase « Après » - MI	41
Figure 29 Différences standardisées des patients de phase « Avant » en fonction de l'inclusion dans MP	42
Figure 30 Différences standardisées des patients de phase « Avant » en fonction de l'inclusion dans MP	43
Figure 31 Différences standardisées avant et après appariement sur le PS du modèle MI	44