Université de Nantes

Ecole Doctorale Biologie Santé

SPHERE (EA-4275)

bioStatistics, Pharmacoepidemiology and Human sciEnces REsearch

Habilitation à Diriger des Recherches

Yohann Foucher

Maître de Conférences en Biostatistique

Modélisation et pronostic de l'évolution de pathologies chroniques : Applications en transplantation rénale

Travail évalué devant le jury suivant

Hélène Jacqmin-Gadda	Directrice de Recherche, ISPED & Inserm U897 (Bordeaux)
Catherine Quantin	Professeur des Universités et Praticien Hospitalier, CHU de Dijon
Luc Frimat	Professeur des Universités et Praticien Hospitalier, CHU de Nancy
Bénédicte Stengel	Directrice de Recherche, Inserm U1018 (Paris)
Véronique Sébille-Rivan	Professeur des Universités et Praticien Hospitalier, EA4275 (Nantes)

Je dédie ce travail à ma grand-mère,
Gabrielle Foucher (1922-2012).

Remerciements

En premier lieu, je tiens à remercier Magali et Véronique. Grâce à elles, j'ai pu intégrer le paysage universitaire et hospitalier nantais. Toujours motivées, dynamiques et à l'écoute ; vos activités professionnelles sont impressionnantes tout en respectant vos vies personnelles et familiales. Je tenterais d'être à la hauteur et de respecter cet équilibre difficile.

Je n'oublie pas Jean-Pierre Daurès pour la formation, les opportunités et les conseils qu'il m'a offerts en thèse et qu'il m'offre toujours. Cette habilitation à diriger la recherche est la continuité des travaux que nous avions initiés ensemble. Merci.

Je tiens aussi à témoigner ma reconnaissance à Jean-Paul Soulillou. Il a toujours soutenu nos développements méthodologiques et statistiques, conscient de l'intérêt que ces derniers peuvent apporter à la recherche clinique et épidémiologique en transplantation. Cette vision est toujours avant-gardiste par rapport aux idées de nombreux jeunes cliniciens et épidémiologistes. La preuve d'un immense chercheur.

Merci bien sûr à toute l'équipe DIVAT-Biostat. Ce travail est aussi et avant tout le vôtre. Je pense à Katy, Marine, Florence, Sandra, Florent, Pascal, Philippe et Etienne. Nos échanges quotidiens dans une ambiance amicale sont très enrichissants. C'est un plaisir de travailler avec vous.

Christophe, même si tu ne fais pas partie de l'équipe, ta contribution est importante. Notre amitié nous rassemble avant tout. Un grand merci pour l'importation de fromages.

Un grand merci à toute l'équipe EA qui m'a accueilli. Il s'agit en particulier des statisticiens : Fanny, Myriam, Elodie, Emilie, Emmanuelle, Alice, Murielle, Mohand et Jean-Benoît. L'aventure ne fait que commencer.

Merci à Sophie, Nicolas, Richard et toute l'équipe U1064 pour leur bonne humeur et les échanges extrêmement riches que nous pouvons avoir ensemble (même sans venir aux réunions du lundi).

Merci à Hélène Jacqmin-Gadda, Catherine Quantin, Luc Frimat, Bénédicte Stengel et Véronique Sébille pour me faire l'honneur de constituer le jury.

Mes remerciements s'adressent enfin à mes proches. Mes parents m'ont toujours soutenu et encouragé. Céline, je te dois tout et bientôt plus. Je vous aime.

Table des matières

Préambule

Ce document présente ma candidature à l'Habilitation à Diriger des Recherches.

La première partie est dédiée au bilan des mes activités de recherche. Comme détaillé dans mon Curriculum Vitae, j'ai réellement commencé le développement de modèles originaux en Biostatistique lors de mon DEA à l'Université de Montpellier (IURC). Ce n'était pas mon choix initial, puisque j'avais réalisé un troisième cycle professionnel avec un DESS à l'Université de Bordeaux 2 (ISPED) quelques années plus tôt. Bien plus éloigné, je voulais initialement être pompier et j'avais choisi un IUT Hygiène, Sécurité et Environnement. Contrairement à l'évolution de pathologies chroniques, il semble donc très difficile de prédire une carrière professionnelle. Cependant, mon parcours explique mon activité de recherche plus appliquée que théorique, avec le souci d'un retour rapide des résultats aux patients.

Dans la seconde partie, en accord avec les recommandations de rédaction de ce type de document, j'ai choisi les 5 publications que je pense les plus représentatives de mon activité, celles aussi qui vont conditionner le futur proche.

J'essaye toujours de conserver du temps pour mes propres développements. Cependant, le constat est que "je" doit être remplacé par "nous". Il y a le "savoir faire", et maintenant le "savoir faire faire" qui n'est pas plus simple. C'est la troisième partie de ce document, avec la description des principaux projets collaboratifs, en particulier autour de la cohorte des patients transplantés. Ce domaine d'application offre le principal fil conducteur de nos projets et il nous permet d'avancer en cohérence. Des projets de collaborations vers d'autres pathologies sont cependant en cours de construction. Je pense que c'est le meilleur moyen pour offrir l'opportunité et l'espace aux lancements d'autres carrières au sein de l'équipe. L'ensemble de ce troisième volet prospectif est en anglais, en cohérence avec tous les travaux rédigés ou en cours de rédaction.

Liste des symboles

ANR Agence Nationale pour la Recherche

ARE Acute Rejection Episode

AT1R Angiotensin II Type 1 Receptor

AUC Area Under the ROC Curve

CIF Cumulative Incidence Function

CSM Composite Surrogate Marker

DEA Diplôme d'Etudes Approfondies

DESS Diplôme d'Etudes Supérieures Spécialisées

DGF Delayed Graft Function

DGFS Delayed Graft Function Score

DIVAT Données Informatisées VAlidées en Transplantation

DRCI Département de la Recherche Clinique et de l'Innovation

ESRD End Stage Renal Disease

HDR Habilitation à Diriger les Recherches

HLA Human Leukocyte Antigen

HMD Human Mortality Database

ISPED Institut de Santé Publique, d'Epidémiologie et de Développement

ITUN Institut de Transplantation - Urologie - Néphrologie

IURC Institut Universitaire de Recherche Clinique

KTFS Kidney Transplant Failure Score

MaKiT Mortality after Kidney Transplantation

MST Maîtrise des Sciences et Techniques

PHRC Programme Hospitalier de Recherche Clinique

PRO Patient Reported Outcome

REIN Réseau Epidémiologie et Information en Néphrologie

ROC Receiver Operating Characteristic

SPHERE bioStatistics, Pharmacoepidemiology and Human sciEnces REsearch

UNOS United Network for Organ Sharing

USRDS United State Renal Data System

Première partie

Bilan de mes activités de recherche

- Chapitre 1 -

Curriculum Vitae

- FOUCHER Yohann
- Lieu et date de naissance : 20 Mai 1979 à Nantes.
- Nationalité : Français.
- Statut : Maître de Conférences en Biostatistique.
- CNU : 85 - Sciences physicochimiques et ingénierie appliquée à la santé.
- Adresse : Université de Nantes - Faculté de Médecine - 1, rue Gaston Veil - BP 53508 - 44035 Nantes Cedex 1 - FRANCE.
- Téléphone : +33 240412824 & Fax : +33 240087411
- Email : Yohann.Foucher@univ-nantes.fr

1.1 Diplômes

2007 Thèse en Biostatistique à l'Université de Montpellier. Laboratoire : Institut Universitaire de Recherche Clinique. Titre : Modèle Semi-Markovien pour l'analyse de l'évolution de pathologie chronique. Directeur : Pr. Jean-Pierre Daurès.

2004 Master 2 Recherche en Biostatistique (ex. Diplôme d'Etudes Approfondies, DEA) à l'Université de Montpellier 1.

2002 Master 2 Professionnel en Statistique Appliquée aux Sciences Sociales et en Santé (ex. Diplôme d'Etudes Supérieures Spécialisées, DESS) à l'Université de Bordeaux 2 (ISPED).

2001 Master 1 en Santé Publique (ex. Maîtrise des Sciences et Techniques, MST) à l'Université de Bordeaux 2 (ISPED).

1999 Diplôme Universitaire de Technologie (DUT) en Hygiène Sécurité et Environnement à l'Université de Bordeaux 1.

1997 Baccalauréat Scientifique au Lycée Joubert (Ancenis, 44).

1.2 Expériences professionnelles

2010-2013 Maître de Conférences : SPHERE (EA 4275). CNU : 85 - Sciences physicochimiques et ingénierie appliquée à la santé.
Membre associé de l'INSERM U1064.
Consultant en Biostatistique/Méthodologie au CHU de Nantes - Département de la Recherche Clinique et de l'Innovation (DRCI) - Plateforme de Biométrie.

2008-2010 Enseignant vacataire à l'Université de Nantes - Faculté de Pharmacie.

2007-2010 Post-doctorant à l'Institut de Recherche en Transplantation - Urologie - Néphrologie (ITUN) à Nantes. Financement de la fondation CENTAURE. INSERM U1064.

2004-2007 Doctorant à l'Université de Montpellier I (IURC). Modèle Semi-Markovien pour l'analyse de l'évolution de pathologies chroniques. Directeur : Pr. Jean-Pierre Daurès. Financement par une bourse du ministère de l'enseignement et de la recherche.

2004-2007 Enseignant vacataire à l'Université de Montpellier I.

2002-2003 Biostatisticien au CHU de Nantes - Pôle d'Information Médicale, d'Evaluation et de Santé Publique.

1.3 Publications

1.3.1 Publications internationales avec comité de lecture

1. Danger R., Paul C., Giral M., Lavault A., **Foucher Y.**, Degauque N., Pallier A., Durand M., Castagnet S., Duong Van Huyen JP., Delahousse M., Renaudin K., Soulillou JP. and Brouard S. Expression of miR-142-5p in peripheral blood mononuclear cells from renal transplant patients with chronic antibody-mediated rejection. Plos One. 2013. In press. (IF 4.092)
2. Danger R., **Foucher Y.** Time dependent ROC curves for the estimation of true prognostic capacity of microarray data. Statistical Applications in Genetics and Molecular Biology. 2012 ; 11(6). (IF 1.517)
3. Combescure C., Daures J., **Foucher Y.** A literature-based approach to evaluate the predictive capacity of a marker using time-dependent summary receiver ope-

rating characteristics. Stat Methods Med Res. 2012. In press. (IF 2.443)

4. Trébern-Launay K., **Foucher Y.**, Giral M., Legendre C., Kreis H., Kessler M., Ladrière M., Kamar N., Rostaing L., Garrigue V., Mourad G., Morelon E., Soulillou JP., Dantal J. Poor long-term outcome in second kidney transplantation : a delayed event. PLoS One. 2012 ; 7(10) :e47915. (IF 4.092)

5. Brouard S., Pallier A., Renaudin K., **Foucher Y.**, Danger R., Devys A., Cesbron A., Guillot-Guegen C., Ashton-Chess J., Le Roux S., Harb J., Roussey G., Subra JF., Villemain F., Legendre C., Bemelman FJ., Orlando G., Garnier A., Jambon H., Le Monies De Sagazan H., Braun L., Noël C., Pillebout E., Moal MC., Cantarell C., Hoitsma A., Ranbant M., Testa A., Soulillou JP., Giral M. The Natural History of Clinical Operational Tolerance After Kidney Transplantation Through Twenty-Seven Cases. Am J Transplant. 2012 ; 12(12) :3296-307. (IF 6.394)

6. Thibault-Espitia A., **Foucher Y.**, Danger R., Migone T., Pallier A., Castagnet S., G-Gueguen C., Devys A., C-Gautier A., Giral M., Soulillou JP., Brouard S. BAFF and BAFF-R Levels Are Associated With Risk of Long-Term Kidney Graft Dysfunction and Development of Donor-Specific Antibodies. Am J Transplant. 2012 ;12(10) :2754-2762. (IF 6.394)

7. Michel L., **Foucher Y.**, Vukusic S., Confavreux C., de Sèze J., Brassat D., Clanet M., Clavelou P., Ouallet JC., Brochet B., Pelletier J., Labauge P., Lebrun C., Lepage E., Le Frere F., Jacq-Foucher M., Barriere P., Wiertlewski S., Laplaud DA. ; Club Francophone de la Sclérose En Plaques (CFSEP).Increased risk of multiple sclerosis relapse after in vitro fertilisation. J Neurol Neurosurg Psychiatry. 2012 ;83(8) :796-802. (IF 4.764)

8. **Foucher Y.**, Giral M., Soulillou JP., Daurès JP. Cut-off estimation and medical decision making based on a continuous prognostic factor : the prediction of kidney graft failure. Int J Biostat. 2012 ; 6 ;8(1). (IF 1.284)

9. **Foucher Y.**, Combescure C., Ashton-Chess J., Giral M. Prognostic markers : data misinterpretation often leads to overoptimistic conclusions. Am J Transplant. 2012 Apr ;12(4) :1060-1. (IF 6.394)

10. Degauque N., Boeffard F., **Foucher Y.**, Ballet C., Brouard S., Soulillou J.P. The blood of healthy individuals exhibits CD8 T cells with a highly altered TCR Vb repertoire but with an unmodified phenotype. PLoS One. 2011 ;6(6) :e21240. (IF 4.092)

11. Racapé M., Duong Van Huyen J.P., Danger R., Giral M., Bleicher F., **Foucher Y.**, Pallier A., Pilet P., Tafelmeyer P., Ashton-Chess J., Dugast E., Pettré S., Charreau B., Soulillou J.P., Brouard S. The involvement of SMILE/TMTC3 in endoplasmic reticulum stress response. PLoS One. 2011 ;6(5) :e19321. (IF 4.092)

12. Brouard S., Le Bars A., Dufay A., Gosselin M., **Foucher Y.**, Guillet M., Cesbron-Gautier A., Thervet E., Legendre C., Dugast E., Pallier A., Guillot-Gueguen C., Lagoutte L., Evanno G., Giral M., Soulillou J.P. Identification of a gene expression profile associated with operational tolerance among a selected group of stable kidney transplant patients. Transpl Int. 2011 Jun ;24(6) :536-47. (IF 2.921)

13. **Foucher Y.**, Giral M., Soulillou J.P., Daures J.P. Time-dependent ROC analysis for a three-class prognostic with application to kidney transplantation. Stat Med. 2010 Dec 30 ;29(30) :3079-87. (IF 2.328)

14. **Foucher Y.**, Daguin P., Akl A., Kessler M., Ladrière M., Legendre C., Kreis H., Rostaing L., Kamar N., Mourad G., Garrigue V., Bayle F., de Ligny B.H., Büchler M., Meier C., Daurès J.P., Soulillou J.P., Giral M. A clinical scoring system highly predictive of long-term kidney graft survival.Kidney Int. 2010 Dec ;78(12) :1288-94. (IF 6.105)

15. Giral M., **Foucher Y.**, Labrune Y., Karam G, Kessler M., Hurault de Ligny B., Büchler B., Bayle F., Meyer C., Daguin P., Soulillou JP. Kidney and recipient weight incompatibility : a cause of early proteinuria and reduced long-term graft survival. Journal of the American Society of Nephrology. J Am Soc Nephrol. 2010 Jun ;21(6) :1022-9. (IF 8.288)

16. Ashton-Chess J., Le Mai H., Jovanovic V., Renaudin K., **Foucher Y.**, Giral M., Moreau A., Dugast E., Mengel M., Racape M., Danger R., Usal C., Smit H., Guillet M., Gwinner W., Le Berre L., Dantal J., Soulillou JP., Brouard S. Immunoproteasome beta subunit 10 is increased in chronic antibody-mediated rejection. Kidney Int. 2010 May ;77(10) :880-90. (IF 6.105)

17. **Foucher Y.**, Giral M., Soulillou JP., Daures JP. A flexible semi-Markov model for interval-censored data and goodness-of-fit testing. Stat Methods Med Res. 2010 ;19(2) :127-45. (IF 1.768)

18. Rousseau V., **Foucher Y.**, Giral M., Soulillou J.P., Daurès J.P. Informative censoring in a multiplicative relative survival model : application for kidney transplant recipients. JP Journal of Biostatistics. 2009 ;3(3) :195-214. (missing IF)

19. Bruneau S., Le Berre L., Hervé C., Valanciuté A., Kamal M., Naulet J., Tesson L., **Foucher Y.**, Soulillou JP., Sahali D., Dantal J. Potential role of soluble ST2 protein in idiopathic nephrotic syndrome recurrence following kidney transplantation. American journal of kidney diseases. 2009 ;54(3) :522-32. (IF 5.152)

20. Ashton-Chess J., Dugast E., Colvin R.B., Giral M., **Foucher Y.**, Moreau A., Renaudin K., Braud C., Devys A., Brouard S., Soulillou J.P. Regulatory, effector, and cytotoxic T cell profiles in long-term kidney transplant patients. Journal of the American Society of Nephrology. 2009 ;20(5) :1113-22. (7.689)

21. **Foucher Y.**, Daguin P, Kessler M, Ladrière M, Legendre C, Kreis H, Durand D, Laveyssiere L, Mourad G, Garrigue V, Daurès JP, Soulillou JP., Giral M. How to evaluate the long-term prognostic capacity of pre-graft variables in transplantation ? Clinical Transplant. 2008 :113-8. (missing IF)
22. Asthon-chess J., Giral M., Mengel M., Renaudin K., **Foucher Y.**, Gwinner W., Braud C., Dugast E., Quillard T., Thebault P., Chiffoleau E., Braudeau C., Charreau B., Soulillou J.P., Brouard S. Tribbles-1 as a Novel Biomarker of Chronic Antibody-Mediated Rejection. Journal of the American Society of Nephrology. 2008 ;19(6) :1116-27. (IF 7.505)
23. Castelli C., Combescure C., **Foucher Y.**, Daurès J.P. Cost-effectiveness analysis in colo-rectal cancer using a semi-Markov model. Statistics in Medicine. 2007 Dec 30 ;26(30) :5557-71. (IF 1.547)
24. **Foucher Y.**, Giral M., Soulillou J.P., Daurès J.P. A semi-Markov model for multistate and interval-censored data with multiple terminal events. Application in renal transplantation. Statistics in Medicine. 2007 ; 26(30) :5381-93. (IF 1.547)
25. Mathieu E., **Foucher Y.**, Dellamonica P., Daurès J.P. A Parametric and Non Homogeneous Semi-Markov Process for HIV control. Methodology and Computing in Applied Probability. 2007 ;3 :389-397. (IF 0.753)
26. Giral M., Bertola J.P., **Foucher Y.**, Villers D., Bironneau E., Blanloeil Y., Karam G., Daguin P., Lerat L., Soulillou J.P. Effect of brain-dead donor resuscitation on delayed graft function : results of a monocentric analysis. Transplantation. 2007 ;83(9) :1174-81. (IF 3.641)
27. **Foucher Y.**, Saint-Pierre P., Puglièse P., Daurès J.P. A Semi-Markov Frailty Model for Multistate Survival Data : Illustration on HIV Disease. Far East Journal of Theoretical Statistics, 2006 ;19 :185-201. (missing IF)
28. **Foucher Y.**, Mathieu E., Saint-Pierre P., Durand J.F., Daurès J.P. A semi-Markov model based on Generalized Weibull distribution with an illustration for HIV disease. Biometrical Journal. 2005 ;47(6) :825-33. (IF 0.768)
29. Giral M., NGuyen J.M., Karam G., Kessler M., Hurault de Ligny B., Buchler M., Bayle F., Meyer C., **Foucher Y.**, Martin M.L., Daguin P., Soulillou J.P. Impact of graft mass in the clinical outcome of kidney transplants. Journal of the American Society of Nephrology. 2005 ;16(1) :261-8. (IF 7.240)

1.3.2 Publications nationales

1. Arnaud I., Elkouri D., N'Guyen JM., **Foucher Y.**, Karam G., Lepage JY., Billard M., Potel G., Lombrail P. Local guidelines and quality of antibiotic treatment

in urinary tract infections : a clinical audit in two departments of a university hospital. Presse Med. 2005 Dec 17 ;34 :1697-702. (IF 0.381)

2. Arnaud I., Elkouri D., N'Guyen JM., **Foucher Y.**, Karam G., Lepage JY., Billard M., Potel G., Lombrail P. Adequate prescription of antibiotic therapy for urinary tract infections in hospital : identifying and correcting non-observance of guidelines. Med Mal Infect. 2005 Mar ;35(3) :141-8. (IF 0.188)

1.3.3 Conférences internationales invitées

1. Dantan E., Lorent M., **Foucher Y.** Confusing correlation and prediction : a simple approach to evaluate marker accuracy for predicting time-to-event. 39th Annual Meeting of the European Group for Blood and Marrow Transplantation. London - April 2013.
2. Dantan E., Combescure C., Ashton-Chess J., Lorent M., Giral M., **Foucher Y.** How to evaluate the prognostic capacity of surrogate makers ? A methodological review in transplantation literature. Congress of the British Transplant Society - Glasgow 2012.
3. **Foucher Y.** Modern models in time-to-event analysis. Workshop at Ilam University of Medical Sciences (Iran). April 2010.

1.3.4 Conférences nationales invitées

1. **Foucher Y.** The modeling of the evolution of kidney transplant recipients Applications to the DIVAT cohort. International Society for Clinical Biostatistics - Montpellier (France), August 2009.
2. **Foucher Y.** Apport des modèles Markoviens dans la modélisation des patients en addictologie. Journées scientifiques de l'Université de Nantes. (France) June 2009.
3. **Foucher Y.** Développements autour des courbes ROC dépendantes du temps. Rencontres méthodologiques de la Direction Interrégionale de la Recherche Clinique (DIRC) Sud-Est, Marseille (France), March 2009.

1.3.5 Présentations orales

1. Lorent M., Giral M., **Foucher Y.** Net time-dependent ROC curves : a new method for evaluating the accuracy of a marker to predict mortality related to end-stage renal disease in kidney transplant recipients. Francophone Society of Transplantation (Nantes), 2012.

2. Lino-Daniel M., Foucher Y., Daguin P., Phelizot C., Hourmant M. Identification des facteurs de risque de progression de l'insuffisance rénale dans une cohorte de patients ayant une clairance de créatinine comprise entre 30 et 60 ml/min. 14ème réunion commune SFD & SN. Genevia - 2013.
3. Trébern-Launay K., Giral M., **Foucher Y.** A multiplicative hazard regression model to compare the effect of factors associated with graft failure risk between first and second renal transplant. International Society for Clinical Biostatistics (Bergen), 2012.
4. Lorent M., Giral M., **Foucher Y.** Relative ROC curves : a solution for evaluating the accuracy of a marker to predict the cause-specific mortality. International Society for Clinical Biostatistics (Bergen), 2012.
5. Hegner B., Giral M., Dufay A., Van Huyen J.P.D., **Foucher Y.**, Renaudin K., Moreau A., Philippe A., Dechend R., Heidecke H., Cesbron A., Devys A., Soulillou J. P. Pre-transplant sensitization against angiotensin II Type 1 receptor is a novel independent risk factor of antibody-mediated tejection. Transpl. Int. Vol : 24 Suppl. : 3 Pages : 30-30 : Oct 2011.
6. Akl A, Giral M, Rousseau V, Kessler M, Ladriere M, Legendre C, Kreis H, Rostaing L, Kamar N, Mourad G, Garrigue V, Daures JP, Soulillou JP, **Foucher Y.** A relative survival model to assess the specific mortality related to renal transplantation. Transpl. Int. Vol : 24 Suppl. : 2 Pages : 204-204. Sep 2011. European Society for Organ Transplantation (ESOT) Congress.
7. Trébern-Launay K, **Foucher Y.**, Giral M., Legendre C., Kessler M., Kamar N., Garrigue V., Morelon E., Dantal J. Poor outcome of second kidney transplantation : a delayed event. Transpl. Int. Vol : 24 Suppl. : 2 Pages : 62-63. Sep 2011. ESOT Congress.
8. Mai H.L., Giral M., Launay K., **Foucher Y.**, Garrigue V., Legendre C., Kamar N., Kessler M., Morelon E., Hourmant M., Brouard S., Soulillou, J.P. Kidney after Nonrenal Organ Transplantation-Analysis of a 15-Year Multicenter Cohort. Am. J. Transplant. Vol : 11 Suppl. : 2 Pages : 148-148. Apr 2011. American Transplant Congress : Philadelphia.
9. Danger R., Lavault A., Giral M., Van Huyen J.P.D, **Foucher Y.**, Pallier A., Degauque N., Soulillou, J.P., Brouard S. miR-142-5p Over-Expression in Blood and Kidney Graft from Patients Exhibiting Chronic Antibody Mediated Rejection. Am. J. Transplant. Vol : 11 Suppl. : 2 Pages : 380-380. Apr 2011. American Transplant Congress : Philadelphia.
10. Brouard S., Pallier A., Renaudin K., Devys A., Cesbron A., Guillot-Gueguen C., **Foucher Y.**, Subra J.F., Villemain F., Legendre C., Thervet E., Bemelman F.J., Le Roux S., Roussey G., Orlando G., Gamier A., Jambon H., De Saeazan H.,

Le Monies Braun L., Noel C., Pillebout E., Moal M.C., Cantarell C., Hoitsma A., Ranbant M., Testa A., Danger R., Soulillou J.P., Giral M. Natural History of Clinical Operational Tolerance after Kidney Transplantation. Am. J. Transplant. Vol : 11 Suppl. : 2 Pages : 76-76. Apr 2011. American Transplant Congress : Philadelphia.

11. Roussey G., **Foucher Y.**, Guest G., Ranchin B., Maisin A., Novo R., Andre J., Cloarec S., Guyot C. Influence of Corticosteroid in Renal Transplantation.15th Congress of the International Pediatric Nephrology Association, 2010.
12. Devys A., Lino M., **Foucher Y.**, Mc Ilroy A., Soulillou J.P., Cesbron A., Blancho G. Would preformed HLA antibodies detected by single antigen assay predict acute humoral rejection or needlessly contraindicate transplantation ? 24th European Immunogenetics and Histocompatibility Conference, 2010.
13. Michel L., **Foucher Y.**, De Seze J., Vukusic S., Confavreux C., Brassat D., Ouallet J., Brochet B., Wiertlewski S., Laplaud D. Increase of Relapse Rate in Patients with Multiple Sclerosis after In-Vitro Fertilization Is Related to the Use of LHRH Agonists : A Multicenter Retrospective Study in France. 62nd Annual Meeting of the American-Academy-of-Neurology, 2010.
14. Victorri-Vigneau C., **Foucher Y.**, Huet M., Guillou-Landreat M., Sebille V., Jolliet P. Use of an original statistical methodology to identify factors linked with a bad observance in the opiate maintenance treatment follow up. Fundamental and clinical pharmacology, 2009.
15. Ashton-Chess J., Dugast E., Colvin RB., Giral M., **Foucher Y.**, Moreau A., Renaudin K., Braud C., Devys A., Brouard S., Soulillou JP. Regulatory, Effector, and Cytotoxic T Cell Profiles in Long-Term Kidney Transplant Patients. 8th American Transplant Congress (Toronto), 2009.
16. **Foucher Y.**, Giral M., Soulillou JP., Daurès JP. Time-dependent ROC analysis for a three-class prognostic. International Society for Clinical Biostatistics (Prague), 2009.
17. Giral M., **Foucher Y.**, Labrune Y., Karam G., Kessler M., Hurault de Ligny B., Buchler M., Bayle F., Meyer C., Daguin P., Soulillou JP. Kidney and recipient weight incompatibility : a cause of early proteinuria and reduced long-term graft survival. American Transplant Congress. (Boston), 2009.
18. **Foucher Y.**, Giral M., Soulillou J.P., Daurès JP. An adaptation of time-dependent ROC curves to construct a prognosis test based on the repetition of a surrogate marker : Application on the kidney graft failure and the creatinine clearance. International Society for Clinical Biostatistics (Copenhaguen), 2008.
19. Giral M., **Foucher Y.**, Daurès JP., Soulillou JP. New tool to analyse confounding

factors in kidney graft outcome : A DIVAT data base study. American Transplant Congress (Toronto), 2008.

20. Asthon-chess J., Jovanovic V., **Foucher Y.**, Dugast E., Giral M., Renaudin K., Heslan M., Soulillou JP., Brouard S. The immunoproteasome subunit beta 10 as a novel peripheral blood and intragraft biomarker of chronic antibody mediated allograft rejection in clinical transplantation. American Transplant Congress (Toronto), 2008.
21. Allain-Launay E., Roussey-Kesler G., Guest G., Ranchin B., Maisin A., Andre J., Cloarec S., **Foucher Y.**, Guyot C. Pediatric renal transplantation : report from a French pediatric database. European Society for Paediatric Nephrology. (Montpellier), 2008.
22. **Foucher Y.**, Soulillou JP., Daurès JP., Giral M. Longitudinal analysis of kidney transplant recipients with multi-state model. American Transplant Congress (San Francisco), 2007.
23. **Foucher Y.**, Giral M., Soulillou JP., Daurès JP. A semi-Markov model for interval-censored data and multiple events : Application to the evolution of kidney transplant recipients. 27th Annual Conference of the International Society for Clinical Biostatistics. (Genevia), 2006.
24. Castelli C., Combescure C., **Foucher Y.**, Daurès JP. Cost-effectiveness analysis in colorectal cancer using a semi-Markov model. 27th Annual Conference of the International Society for Clinical Biostatistics. (Genevia), 2006.
25. **Foucher Y.**, Giral M., Soulillou JP., Daurès JP. A semi-Markov Model with Interval Censoring and Non-Proportional Hazards. 23th International Biometric Conference. (Montréal), 2006.
26. Asthon-chess J., Giral M., Braud C., Baeten D., Heslan M., Chiffoleau E., **Foucher Y.**, Soulillou J.P., Brouard S. Tribbles homolog 1, a peripheral blood marker of chronic rejection in clinical renal allotransplantation identified by transcriptome profiling : role and correlation with renal histology and function. World transplant congress, 2006.
27. Harzallah K., Frimat L., Giral M., **Foucher Y.**, Soulillou JP., Dantal J. Induction Therapy Is Not Associated with Increasing Risk of Cancer After Renal Transplantation. The Federation of Clinical Immunology Societies (FOCIS), 2006.
28. Cuzin L., **Foucher Y.**, Agher R., Barone M., Billaud E., Daurès JP., Dellamonica P., Druart P., Salmi D., Puglièse P. Safe and effective switch to a triple nucleosidic reverse transcriptase inhibitors regimen in a successfully pre-treated French HIV-infected population. IAS Conference. 2005.
29. Mathieu E., **Foucher Y.**, Dellamonica P., Daurès JP. Parametric and Non Homogeneous semi-Markov Process for HIV Control. Applied Stochastic Models and

Data Analysis (ASMDA), 2005.

1.4 Logiciels

1.4.1 Packages R

R propose un langage et un environnement logiciel libre d'accès (*open source*) pour les calculs statistiques et graphiques. R offre une grande variété d'outils statistiques, auxquels peuvent s'ajouter des éléments complémentaires, appelés *packages*. Tous les packages sont disponibles à l'adresse suivante : `www.divat.fr/en/softwares`.

ROCt Il s'agit d'un ensemble de fonctions R pour calculer des courbes ROC (Receiver Operating Characteristic) dépendantes du temps à partir de l'estimateur de Kaplan-Meier et de celui des k plus proches voisins. Ces deux estimateurs sont développés pour les modèles de survie traditionnels (toutes causes confondues) et pour les modèles de survie nette (mortalité en excès liée à une pathologie).

SROCt Il s'agit d'un ensemble de fonctions R pour calculer des courbes ROC dépendantes du temps à partir de données de survie agrégées, c'est à dire à partir des courbes de survie publiées dans différents articles en fonction des niveaux d'un marqueur.

prognosticROC L'objectif des courbes ROC pronostiques est d'évaluer les qualités pronostiques d'un marqueur binaire en fonction des survies calculées dans chacun des deux groupes. Contrairement aux courbes ROC dépendantes du temps, qui sont basées sur les notions de sensibilité et de spécificité, les courbes ROC pronostiques sont basées sur les valeurs prédictives, découlant directement des courbes de survie. L'aire sous la courbe représente la probabilité que le temps de survie soit meilleur dans un groupe que dans l'autre. Ce travail est réalisé en collaboration avec le Dr. Christophe Combescure (Hopitaux Universitaires de Genève).

ROC632 Les biopuces peuvent être utilisées pour construire une signature pronostique. Le problème principal est le sur-ajustement dû à un nombre de variables explicatives très grand par rapport au nombre d'individus. Nous avons développé l'estimation des courbes ROC dépendantes du temps en les adaptant à l'algorithme de ré-échantillonnage par Bootstrap 0,632+. Cette méthode permet de corriger assez efficacement la sur-estimation de l'aire sous la courbe, sous réserve d'un nombre raisonnable de marqueurs mesurés par rapport au nombre de sujets et du taux de censure à droite.

MetaSurv En Epidémiologie, il est maintenant assez bien reconnu qu'une courbe de survie est plus informative qu'un "simple" rapport de risque (*hazard ratio*).

Cependant, dans les méta-analyses, la quantité la plus populaire est le rapport de risque moyen à partir de tous les rapports observés (pondération en fonction des intervalles de confiance). Avec ce package, nous proposons d'estimer des courbes de survie moyennes à partir d'un ensemble de courbes de survie observées. Ce travail est réalisé en collaboration avec le Dr. Christophe Combescure (Hopitaux Universitaires de Genève).

MRsurv Il s'agit de fonctions R qui permettent de réaliser des modèles multiplicatifs de survie relative en l'absence de tables de mortalité. Ce modèle permet de comparer l'effet de facteurs de risque entre deux groupes issus de deux échantillons observés.

1.4.2 Calculateurs en ligne

Tous les scores prédictifs ou les développements mis en place pour l'aide à la prise de décisions des cliniciens sont disponibles sous forme d'applications pour smartphones, tablets et ordinateurs : `www.divat.fr/en/online-calculators`.

KTFS Le KTFS (*Kidney Transplant Failure Score*) est un indicateur composite basé sur 8 paramètres disponibles chez tous les patients au premier anniversaire de la greffe rénale. En entrant les valeurs des paramètres, l'application renvoie la valeur du score et fournit une série d'interprétations.

EVALBIOM La distance entre les courbes de survie est souvent mal interprétée en terme de capacités pronostiques. A partir de cette application, il est possible d'entrer les effectifs à la *baseline* ainsi que les probabilités de survie à un certain temps pour deux groupes de patients. Elle renvoie, pour ce temps donné, les indicateurs objectivement interprétables comme la sensibilité, la spécificité, les valeurs prédictives, etc.

DGFS Le DGFS (*Delayed Graft Function Score*) est un indicateur basé sur 5 variables collectées en pré-greffe. Il permet d'évaluer le risque de retard au démarrage du greffon. L'application fonctionne comme pour le KTFS.

1.5 Brevets

KTFS Ce score, calculé à un an post transplantation, permet d'évaluer le risque d'échec de greffe (retour en dialyse) jusqu'au huitième anniversaire de la greffe. Numéro d'enregistrement : 0959043. Titre : *Method and device for determining a risk of graft rejection.*

1.6 Financements obtenus pour des projets propres

1. Elaboration d'un score pronostique de la mortalité liée à la transplantation rénale pour l'aide à l'allocation des greffons. Etude épidémiologique observationnelle multicentrique de la survie relative des patients transplantés rénaux en prenant en compte une population comparable de patients dialysés en attente de transplantation à partir des cohortes prospectives DIVAT et REIN (PI M. Giral, PHRC national 2011).
2. Etude randomisée ouverte d'évaluation d'une prise en charge par TELECONSULTATION versus un suivi STANDARD de patients Transplantés Rénaux en fonction d'un score de risque précoce d'échec de greffe (PI A. Meurette, PHRC national 2011).
3. Construction d'un marqueur composite de substitution de la survie à long terme : application à la transplantation rénale. (ANR Jeune Chercheur Jeune Chercheuse, 2010)

1.7 Etudes cliniques comme méthodologiste

De manière complémentaire au laboratoire SPHERE (EA-4275) où les chercheurs y développent leurs propres activités de recherche, j'aide les chercheurs et les cliniciens à construire les études adaptées à leurs objectifs, en particulier pour l'obtention de financements. Cette activité est principalement réalisée au sein de la plateforme de Biométrie, dirigée par le Pr. V. Sébille-Rivan, au sein de la Direction de la Recherche Clinique et de l'Innovation (DRCI).

Voici une liste de projets ayant obtenus un financement :

1. Etude prospective, multicentrique, randomisée, ouverte et contrôlée évaluant l'intérêt d'un système d'oxygénation à haut débit lors de la pré oxygénation pour l'intubation du malade en insuffisance respiratoire aigue hypoxémique (PHRC inter-régional, 2012).
2. Performance diagnostique des signes cliniques chez les patients adultes suspects de méningites dans les Services d'Accueil des Urgences. Etude prospective multicentrique (PHRC local, 2012).
3. Etude de phase 1 évaluant la tolérance du Trioxyde d'Arsenic (ATO) par voie intraveineuse dans le Lupus Systémique (LES) évolutif (Medsenic, 2012).
4. Impact de l'absence de corticothérapie sur l'évolution de la fonction rénale et

sur la progression de la fibrose du greffon, quantifiée par méthode numérique, chez des patients transplantés rénaux – Etude multicentrique, randomisée, ouverte, prospective de phase IV (Astellas, 2011).

5. Etude prospective, multicentrique, randomisée, en double-aveugle, contrôlée en groupes parallèles, visant à évaluer la balance bénéfice-risque du sevrage progressif d'un inhibiteur de la calcineurine (Tacrolimus) chez des patients transplantés rénaux depuis plus de 4 ans et cliniquement sélectionnés (PHRC national 2010).
6. Etude prospective multicentrique ouverte d'évaluation de l'efficacité et de la sécurité de la neuromodulation du système nerveux végétatif sympathique dans la prise en charge du syndrome de vessie douloureuse (PHRC national 2009).
7. Etude prospective multicentrique génétique et physiopathologique de la dysrégulation du complément au cours des pertes fœtales à répétition (PHRC régional 2009).
8. Evaluation de l'efficacité de la transplantation pancréatique par rapport au traitement de base par l'insuline chez des patients diabétiques de type 1 (PHRC national 2009).
9. La validation de biomarqueurs par ProtoArrays pour la mise au point d'un test de diagnostic de la glomérulopathie d'allogreffe et de la néphropathie chronique du greffon rénal (ANR 2009).
10. Extension de la base DIVAT aux items chirurgicaux de la transplantation (PHRC national 2009).
11. Performances diagnostiques de biomarqueurs mesurés à un an (TRIB1, FoxP3, etc.) pour remplacement de la biopsie (DHOS 2008).
12. Identification de nouveaux virus sur puce olégonucléotidique panvirale chez les patients transplantés (PHRC national 2007).

- Chapitre 2 -

Encadrement d'étudiants

2.1 Étudiants en thèse de science

L'ensemble des étudiants inscrits en thèse sur les projets de transplantation rénale sont dirigés par le Pr. Magali Giral. Je suis co-encadrant à hauteur de 50%.

2.1.1 Katy Trébern-Launay

Le Dr. Katy Trébern-Launay a débuté sa thèse en 2010, grâce à un soutien de la fondation CENTAURE. L'objectif de son travail est l'étude de l'évolution des patients greffés une seconde fois d'un rein par rapport à ceux greffés pour la première fois. C'est aussi cette application qui motive l'utilisation originale de modèles de survie relative.

Plusieurs communications orales ont été associées à au travail de Katy Trébern-Launay :

- Trébern-Launay K, Foucher Y., Giral M., Legendre C., Kessler M., Kamar N., Garrigue V., Morelon E., Dantal J. Poor outcome of second kidney transplantation : a delayed event. Transpl. Int. Vol : 24 Suppl. : 2 Pages : 62-63. Sep 2011. ESOT Congress.
- Trébern-Launay K., Giral M., Foucher Y. A multiplicative hazard regression model to compare the effect of factors associated with graft failure risk between first and second renal transplant. International Society for Clinical Biostatistics (Bergen), 2012.

De plus, un premier article a été publié :

- Trébern-Launay K., Foucher Y., Giral M., Legendre C., Kreis H., Kessler M.,

Ladrière M., Kamar N., Rostaing L., Garrigue V., Mourad G., Morelon E., Soulillou JP., Dantal J. Poor long-term outcome in second kidney transplantation : a delayed event. PLoS One. 2012 ; 7(10) :e47915.

Dans ce travail, nous montrons un plus mauvais pronostic des secondes greffes, cet excès de risque apparaissant après quelques années de greffe. Avec l'amélioration de la qualité et de l'espérance de vie des patients re-transplantés par rapport aux dialysés, il est clair que cet excès de risque parait négligeable et ne remet pas en question la stratégie de greffe. En revanche, pour améliorer l'allocation des greffons, il peut apparaître important de comparer les profils des patients à risque. Nous développons les modèles multiplicatifs de survie relative pour permettre une comparaison performante de ces deux populations (voir le chapitre 11 pour plus de détails). Ce travail est soumis :

– Trébern-Launay K., Giral M., Dantal J. Foucher Y. A multiplicative-regression model to compare the effect of factors associated with the time to graft failure between first and second renal transplant. BMC Medical Research Methodology.

2.1.2 Marine Lorent

Marine Lorent a commencé sa thèse de science en décembre 2011. L'objectif de son travail est de modéliser et pronostiquer la mortalité post-transplantation rénale des receveurs. Des scores pronostiques ont déjà été publiés, comme celui de Hernandez et al. [67]. L'originalité de son projet est de travailler sur la notion de survie nette, c'est à dire la survie si la seule cause de décès possible est liée à la maladie (transplantation, insuffisance rénale, etc). Pour cela, nous avons développé le concept de courbes ROC nettes (voir le chapitre 10 pour plus de détails). Deux communications orales ont été associées à ce travail :

– Lorent M., Giral M., **Foucher Y.** Net time-dependent ROC curves : a new method for evaluating the accuracy of a marker to predict mortality related to end-stage renal disease in kidney transplant recipients. Francophone Society of Transplantation (Nantes), 2012.
– Lorent M., Giral M., **Foucher Y.** Relative ROC curves : a solution for evaluating the accuracy of a marker to predict the cause-specific mortality. International Society for Clinical Biostatistics (Bergen), 2012.

De plus, un papier est soumis :

– Lorent M., Giral M., Foucher Y. Net time-dependent ROC curves : a solution for evaluating the accuracy of a marker to predict disease-related mortality. Statistics in Medicine.

L'ensemble de ce projet est financé par l'ANR CSM (*Composite Surrogate Marker*, Jeunes Chercheurs et Jeunes Chercheuses, 2011).

2.1.3 Florence Gillaizeau

Florence Gillaizeau a débuté sa thèse de science en septembre 2012. L'objectif est de poursuivre les développements sur les modèles multi-états en transplantation rénale. Un des axes de développements est d'inclure le concept de survie nette dans ces modèles semi-Markoviens. Son projet est financé par le PHRC national MaKiT (*Mortality after Kidney Transplantation*, PROG/11/85, 2011).

Pour initier ce projet, Florence Gillaizeau a développé un premier modèle semi-Markovien pour évaluer l'effet d'un anticorps préformé à la greffe (AT1R) sur l'évolution du patient transplanté. Une communication orale est associée à ce travail au congrès de la Société Française de Transplantation à Nantes en 2012. Un article est soumis :

– Gillaizeau F., Giral M., Dantan E., Dragun D., Soulillou JP., Foucher Y. Multi-state analysis of kidney transplant recipients outcome : a semi-Markov model for studying the role of pre-transplant sensitization against Angiotensin II Type 1 receptor. Journal de la Société Française De Statistique.

2.2 Etudiants en Master 2

2.2.1 Florent Le Borgne (Biostatistique, Bordeaux, 2012)

L'objectif de ce stage est l'étude des facteurs de risque d'un retard au démarrage du rein greffé (*Delayed Graft Function*, DGF). Ce DGF est défini par le besoin d'au moins une dialyse dans la première semaine qui suit la greffe. Au-delà de cet objectif, nous avons proposé un score pronostique, plus simple que la référence publiée aujourd'hui (voir la section 9.1 pour plus de détails).

2.2.2 Pascal Rigouin (Biostatistique, Bordeaux, 2010)

Pascal Rigouin a initié pour la première fois les modèles de survie relative appliqués aux transplantés rénaux au sein de l'équipe SPHERE. La mortalité attendue était estimée à partir des patients du réseau SIMS-REIN en Ile de France (Pr. Paul Landais). L'objectif était de modéliser les différences entre les facteurs de risque de mortalité des patients en liste d'attente et ceux transplantés. Les résultats comportaient de nom-

breuses limites, mais ce sont eux qui ont justifié le PHRC national MaKiT (PHRC, PROG/11/85, 2011).

2.2.3 Katy Trébern-Launay (Epidémiologie, Nantes, 2009)

Pendant ce stage, Katy Trébern-Launay a initié son travail de thèse.

2.2.4 Yann Labrune (Mathématique appliquée, Vannes, 2008)

L'objet du stage était l'étude du pronostic des patients transplantés en fonction du poids de leur rein. Yann Labrune a ainsi pu montrer un épuisement lié à l'inadéquation de la masse du rein par rapport au poids du receveur. Cette inadéquation augmente le risque de retour en dialyse après trois ans de greffe (période d'épuisement). Ce travail a été publié en 2010 :

– Giral M., Foucher Y., Labrune Y., Karam G, Kessler M., Hurault de Ligny B., Büchler B., Bayle F., Meyer C., Daguin P., Soulillou JP. Kidney and recipient weight incompatibility : a cause of early proteinuria and reduced long-term graft survival. Journal of the American Society of Nephrology. J Am Soc Nephrol. 2010 Jun ;21(6) :1022-9.

2.2.5 Thomas Moyon (Mathématique appliquée, Vannes, 2007)

Il s'agissait d'une analyse globale des facteurs de risque d'échec de greffe à partir de la cohorte DIVAT.

- Chapitre 3 -

Résumé des activités passées

3.1 Une continuité de travaux autour de DIVAT

Mes premières activités de recherche en Biostatistique ont commencé à l'Institut Universitaire de Recherche Clinique (IURC), Université de Montpellier. J'y ai réalisé mon stage de DEA qui s'est poursuivi par une thèse de science (bourse du ministère pour l'enseignement et la recherche). L'objectif principal était de développer les modèles multi-états dans une approche semi-Markovienne et de montrer leurs intérêts en transplantation rénale à partir de la cohorte de patients transplantés rénaux DIVAT. Ce domaine d'application découlait d'une première collaboration avec Magali Giral sur l'étude du pronostic des receveurs en fonction du poids de leur greffon. Ce travail s'était déroulé dans les meilleures conditions lors de mon premier passage au CHU de Nantes comme ingénieur d'étude.

Cette collaboration et les premiers résultats ont permis de justifier la suite de ma thèse en post-doctorat à l'ITUN (INSERM U1064), grâce au soutien de la fondation CENTAURE (`www.fondation-centaure.fr`). Après deux années, j'ai rejoint officiellement l'équipe SPHERE (EA 4275), dirigée par le Pr. V. Sébille-Rivan, en conservant des liens étroits avec l'INSERM.

3.2 La présentation de la cohorte DIVAT

Cette cohorte n'est pas un simple registre car elle rassemble l'ensemble des critères cliniques et biologiques utiles à la prise en charge et au suivi médical du patient transplanté rénal. Il s'agit d'un outil de recherche clinique. Ce réseau comprend les centres de

Nantes, Nancy, Toulouse, Lyon, Paris Necker, Paris Saint-Louis et Montpellier. Plus de 15000 patients ont été inclus depuis l'initiation du réseau. Aujourd'hui, 40% des patients greffés rénaux français sont inclus dans le réseau en 1990. Les données sont saisies de manière prospective par un assistant de recherche clinique spécifiquement formé dans chaque centre. Des audits croisés sont réalisés entre les centres tous les ans. Un contrôle de cohérence automatique est réalisé toutes les semaines et un rapport est envoyé aux attachés de recherche clinique.

A la greffe, les données du donneur sont collectées : âge, sexe, système HLA, relation avec le receveur, hypertension, cause de décès, pression artérielle, médicaments, sérologies et données biologiques (urée, protéinurie et créatinémie). Parallèlement, les données du receveur sont : date de transplantation, âge, sexe, sérologie, poids, taille, antécédents médicaux (cancer, cardiovasculaire, etc.), temps d'ischémie froide, maladie initiale, méthode de dialyse, greffes précédentes, date de première dialyse, système HLA, cross-match, immunisation (*Panel Reactive Antibody*, PRA) et les dialyses post-greffes.

Pendant le suivi, qui est obligatoire à 3 mois, à 6 mois et à tous les anniversaires de la greffe, les paramètres suivants sont collectés : poids, taille, pression artérielle, créatinémie, protéinurie, cholestérol, hémoglobine, traitements immunosuppresseurs et antihypertenseurs. D'autres paramètres sont collectés continuellement : épisodes de rejet aigu, infections, complications, retours en dialyse, décès. Le suivi est interrompu au retour en dialyse.

DIVAT a plus récemment été couplé à une biocollection (DIVAT biocoll.). Elle rassemble les urines, le sérum, la plasma et les cellules PBMC des donneurs et des receveurs au sein du Centre de Ressources Biologiques (CRB) du CHU de Nantes. Elle permet des travaux en routine sur les biomarqueurs en s'appuyant sur les compétences de l'équipe du Dr. S. Brouard de l'unité de recherche U1064 de l'INSERM. Cette recherche translationnelle a permis d'identifier de nombreux biomarqueurs potentiels : sanguins phénotypiques et transcriptionnels.

3.3 Les modèles semi-markoviens

En médecine, la quasi-totalité des travaux qui s'intéressent à la survenue d'événements sont basés sur le modèle à risque proportionnel de Cox [23] et la probabilité de survie est décrite par l'estimateur non-paramétrique de Kaplan-Meier [81]. C'est le cas en transplantation rénale où l'objectif principal est d'améliorer la survie du patient et de son greffon. Plus précisément, l'identification en pré-greffe des facteurs de risque de la perte des greffons permet aux experts de définir les règles de répartition et d'adapter la surveillance des patients après la greffe. La littérature dans ce domaine est très

importante mais repose souvent sur une méthodologie statistique peu adaptée.

L'estimateur de Kaplan-Meier et/ou le modèle de Cox permettent l'analyse du temps d'apparition d'un événement unique. Deux choix de variables à expliquer sont souvent rencontrés : le délai entre la greffe et le retour en dialyse (les décès avec un greffon fonctionnel sont alors des censures à droite) ou le délai entre la greffe et le premier échec observé (retour en dialyse et décès du patient avec un greffon fonctionnel). La première solution est critiquable puisqu'un nombre non-négligeable de décès est dû à la greffe ou à sa prise en charge. La surestimation de la survie qui en découle n'est pas le seul problème. Le modèle de Cox suppose en effet une indépendance entre le processus de censure et le temps de l'événement étudié. Cette hypothèse est difficilement acceptable puisque le niveau de la fonction du greffon est lié à de nombreuses comorbidités. Le second choix est à l'inverse le plus pessimiste. Il considère tous les décès comme liés à la transplantation alors que beaucoup d'entre eux sont indépendants.

Nous avons ainsi montré que l'approche multi-états permettait d'améliorer la modélisation de la survie. Nous avons en particulier choisi l'approche semi-Markovienne, pour laquelle les risques de transition entre les états dépendent principalement du temps passé dans l'état. Nous avons initialement développé ce modèle pour l'évolution des patients atteints du VIH [49, 50, 97].

L'application de ces modèles pour étudier l'évolution des patients transplantés rénaux a été ensuite proposée pour gérer les transitions censurées par intervalle [45]. Grâce à cette méthode, plus de facteurs de risque sont identifiés et leur interprétation est d'autant plus direct et qualitative que l'effet est modéliser spécifiquement sur chaque transition. Le modèle semi-Markovien permet aussi de prendre en compte une structure d'aggravation de l'état du greffon en se basant sur le débit de filtration glomérulaire.

Nous avons également utilisé ce modèle pour améliorer l'évaluation médicoéconomique de traitements de pathologies chroniques [19]. Deux prises en charge différentes de patients atteints d'un cancer du côlon ont été étudiées. Leur efficacité a été modélisée selon une approche multi-états : rémission, rechute, décès. Les coûts de ces prises en charge, conditionnés sur l'état de santé, permettent une modélisation plus précise par rapport à la méthode classique (rapport entre l'efficacité globale et les coûts marginaux).

Dans la continuité de ces travaux, nous avons proposé un test permettant d'évaluer le respect de l'hypothèse semi-Markovienne selon laquelle la fonction de risque dépend du temps passé dans l'état en cours [46]. Dans ce même travail, nous proposons une amélioration de la gestion de la censure par intervalle.

3.4 La survie relative

Même si ces approches multi-états dépassent certaines limites du modèle de Cox, comme la séparation explicite des transitions vers le décès ou vers le retour en dialyse, elles n'offrent pas de solution dans la confusion entre les décès indépendants ou liés à la transplantation. Un modèle spécifique à certaines causes de décès n'est pas adapté puisqu'il n'est pas possible de définir explicitement les décès dus ou non à la pathologie. Plus récemment, nous avons donc adapté les modèles de survie relative pour pouvoir indirectement supprimer la part de la mortalité indépendante de la transplantation à partir de tables de mortalité de la population générale. Nous montrons ainsi que la surmortalité relative à la transplantation est indépendante du sexe du receveur. Autrement dit, le fait d'être transplanté n'aggrave pas l'écart de mortalité observé entre hommes et femmes dans la population générale. L'inverse est observé pour l'âge du receveur à la greffe puisqu'il semble que l'effet de l'âge observé dans la population générale est aggravé après la transplantation. Ce travail est en cours de ré-écriture et sera bientôt soumis. Nous avons aussi intégré cette approche dans un modèle à risques compétitifs grâce au travail de thèse de Vanessa Rousseau [131] : le retour en dialyse et le décès dû à la transplantation sont en compétition et la fonction de risque vers le décès dû à la transplantation est estimée relativement à la population générale.

La limite majeure de ces analyses porte sur la population de référence qui n'est pas forcément la population générale, mais plutôt les patients qui restent en dialyse. Modéliser cette mortalité attendue des patients sur liste d'attente d'un greffon est un des objectifs du PHRC national MaKiT.

3.5 Les courbes ROC dépendantes du temps

Indépendamment des limites méthodologiques des modèles statistiques utilisés en transplantation pour l'analyse des temps d'événements, beaucoup d'études s'intéressent à identifier des marqueurs non-invasifs qui permettraient de mieux stratifier les patients selon leur risque de faire tel ou tel événement. Les marqueurs cliniques non-invasifs couramment utilisés comme variables pronostiques sont la créatinémie ou le débit de filtration glomérulaire. Il existe une littérature importante démontrant leur forte corrélation avec la survie du greffon et/ou du patient. Cependant, on retrouve la plupart du temps une inadéquation entre la question posée et la méthode statistique utilisée : les conclusions sont basées sur des tests statistiques évaluant la probabilité que le lien observé soit dû au hasard (p-value). Il existe une confusion entre les notions de corrélation et prédiction, nous avons d'ailleurs rappeler ce problème méthodologique dans une lettre

à l'éditeur de l'*American Journal of Transplantation* [42].

La capacité pronostique d'un marqueur peut être évaluée par l'aire sous la courbe ROC (Receiver Operating Characteristic). Les notions de sensibilité et de spécificité mesurent mieux les qualités de discrimination d'un outil diagnostique. Cependant, les courbes ROC traditionnelles ne prennent pas en compte la dimension temporelle des données : les trajectoires peuvent être censurées ou tronquées et un événement n'a pas la même importance selon son temps d'apparition. Pour prendre en compte cette complexité supplémentaire, nous avons utilisé et adapté les courbes ROC dépendantes du temps [63]. Nous avons ainsi évalué les capacités pronostiques des principaux facteurs de risque pré-greffe connus en transplantation rénale. Même si certaines variables sont fortement corrélées à la survie du greffon et/ou du patient ($p<0{,}0001$), aucune ne semble constituer un marqueur à partir duquel une décision médicale peut être prise [44].

Nous avons proposé un outil pronostique d'échec de greffe rénale alliant plusieurs paramètres [43]. Il est calculé un an après la greffe. L'aire sous la courbe est égale à 0,78, capacité pronostique supérieure aux marqueurs classiquement utilisés (créatinémie, évolution de la créatinémie et filtration glomérulaire estimée). Un brevet est déposée (0959043).

Nous avons aussi développé la théorie des courbes ROC dépendantes du temps pour permettre le calcul de seuils de décision [48]. L'objectif est de définir un test binaire permettant de classer les patients à haut risque des patients à faible risque. Cette valeur minimise une fonction de coût proportionnelle à la probabilité de faux positifs et négatifs. Pour rester proche de la problématique clinique, cette fonction de coût intègre aussi le poids d'un faux positif par rapport à celui d'un faux négatif. Le chapitre 6 détaille ce travail.

Pour rejoindre notre problématique en transplantation, nous avons développé les courbes ROC dépendantes du temps lorsque deux événements sont en compétition : retour en dialyse et décès avec un greffon fonctionnel [47]. Le pronostic est alors composé de trois classes : retour en dialyse, décès avec un greffon fonctionnel et vivant avec un greffon fonctionnel. Notre modèle utilise le principe semi-markovien : les courbes ROC paramétriques peuvent être soit spécifiques à chaque événement à pronostiquer, soit marginales (quel que soit l'événement). Nous avons ainsi démontré que la fonction rénale (formule MDRD) constitue un marqueur acceptable du risque de retour en dialyse mais il n'est pas pronostique du risque de décès (toutes causes confondues). Ce travail est détaillé dans le chapitre 7.

Plus récemment, nous avons proposé une méthode pour construire une courbe ROC à partir de courbes de survie déjà publiées [21]. L'objectif est de pouvoir évaluer les capacités pronostiques d'un marqueur à partir de méta-analyses. Notre méthode est

d'abord validée à partir des données de DIVAT (comparaison des estimations obtenues à partir des données individuelles et celles faites à partir des données agrégées pour chaque centre). La méthode est ensuite appliquée sur les données de la littérature pour évaluer si le marqueur de prolifération tumorale Ki-67 est un marqueur pronostique de la survie des patientes atteintes d'un cancer du sein. Ce travail est détaillé dans le chapitre 5.

Enfin, nous avons adapté l'algorithme de bootstrap 0,632+ à l'estimation des courbes ROC dépendantes du temps [25]. Comme détaillé dans le chapitre 4, l'objectif est de construire et valider une signature pronostique à partir de biopuces lorsque les données sont censurées et/ou tronquées. Cet algorithme permet d'éviter les résultats trop optimistes dus à la sur-paramétrisation (*overfitting*, lorsque le nombre de variables largement supérieur au nombre d'individus). Pour réduire cette dimension, nous utilisons une pénalisation de la vraisemblance du modèle de Cox de type Lasso.

3.6 Un soutien méthodologique à l'équipe U1064

Comme abordé précédemment, par exemple avec le KTFS associé à une aire sous la courbe ROC à 8 ans post greffe de 0,78, les paramètres cliniques ne permettent pas à eux seuls de prédire l'évolution future d'un patient. La solution est de prendre en compte de nouveaux biomarqueurs ou d'identifier de nouveaux facteurs de risque cliniques. Il s'agit d'un des objectifs de l'équipe INSERM U1064 dirigée par Sophie Brouard. Ma contribution consiste en un soutien statistique et méthodologique. Les principaux travaux ont consisté à étudier le rôle du protéasome [13], de FOXP3 [12], de SMILE [125], du répertoire Vβ [28] ou de BAFF [144] comme marqueurs potentiellement impliqués dans le diagnostic ou le pronostic des patients transplantés rénaux.

Deuxième partie

Cinq publications représentatives

- Chapitre 4 -

Time dependent ROC curves for the estimation of true prognostic capacity of microarray data

Danger R. and Foucher Y.
Statistical Applications in Genetics and Molecular Biology. 2012 Nov 22 ;11(6).

4.1 Introduction

The use of microarray technology has revolutionized the identification of molecular signatures for the prediction of patient outcome, notably in the cancer field [150, 130, 139]. In such time-to-event data, the samples (tissue, blood, urine, etc.) are collected under the same experimental conditions (event-free) and the individuals are followed up in order to identify features predictive of the event. Depending on the follow-up interval, the time of the event may be observed or not (right-censored). The features selection method must take this type of incomplete data into account.

One solution is to firstly select the features according to initial risk(s) factor(s) collected at the same time as the sample and to secondly use time-to-event methods in order to validate the predictive properties of these features [130, 5, 73]. Nevertheless, these risks factors are not necessarily observed and some features may be predictive independently from known risk factors. Few authors have proposed using survival-specific methods in order to directly select the best predictive features. The simplest method is the Kaplan and Meier estimator [81] associated with the LogRank statistic. Even though this type of non-parametric method is attractive, the feature expression has to be categorized. In order to calculate the cut-off values, Jenssen et al. computed the LogRank statistic for all features and for all possible cut-off values for each feature (2 or 3 classes) [75]. The

inherent problem of dimension is then dramatically exacerbated by the repetition of the tests for all the cut-off values. The Cox proportional hazards model should be used for the regression of time according to quantitative factors such as feature expression, but some reductions of the dimension need to be performed since the number of features is much higher than the number of individuals.

To manage this issue, a simple method is the univariate selection of features based on traditional inference tests. Bovelstad et al. found the Cox model with lasso penalty very reliable, as it is also a variable selection method with considerably better prediction performance [15]. Li and Gui proposed method that combines partial Cox regression with partial least squares [92]. However, Schumacher et al. [133] demonstrated, using the 0.632+ estimator of the prediction error, that the approach of Li and Gui performed worse than the Kaplan-Meier benchmark value, ignoring all covariate information [133]. In accordance with the comparative study by Bovelstad et al. [15], Schumacher et al. found that the Cox model with lasso penalty appeared to have the most potential [133]. The authors also demonstrated that the 0.632+ estimator of the prediction error is an interesting indicator for evaluating and comparing different models. It takes into account overfitting without splitting the available data into training and validation sets. The 0.632+ estimator of prediction error [133] or the corresponding summary measure by integrating over the time are thus useful tools for comparisons and validations of survival models in the context of overfitting [122]. Nevertheless, three limitations should be considered. First, the tuning parameter is defined using the full data (5-fold cross-validation), while the selection of the model complexity has to be included in each bootstrap iteration in order to avoid overoptimistic results [136]. Second, even though the prediction error can be used for model comparisons, it is not a meaningful indicator for biologists or clinicians. Third, the prediction error is based on the regression residuals and therefore depends on the incidence of the event, but the sample may not necessarily represent the targeted population.

In this paper, we propose a 0.632+ estimator of the area under the time-dependent Receiver Operating Characteristic (ROC) curve [63]. The method is designed for the analysis of censored and/or truncated survival data. In Section 2, we describe the 0.632+ approach. In Section 3, we perform simulations to test the adequacy of the method. We apply this method in Section 4 to the DLBCL (diffuse large-B-cell lymphoma) dataset published by Rosenwald et al. [130]. All of these analyses use the Cox model with lasso penalty.

4.2 Methods

4.2.1 Cox model with lasso penalty

Let X be the vector of the P feature expressions with $X = (X_1, ..., X_P)$ and $x_j = (x_{j1}, ..., x_{jP})$ the associated observations for the individual j ($j = 1, ..., N$). N represents the number of individuals. Let T_j be the time of the event occurrence for the individual j and C_j be the corresponding time of last follow-up (right-censoring time). If $C_j < T_j$, T_j is not observed and the only available information is that the event occurs after C_j ($\Delta_j = 0$). If $C_j \geq T_j$, T_j is observed ($\Delta_j = 1$). The time of the last observation of the individual j is Y_{1j}, with $Y_{1j} = min(T_j, C_j)$. We also consider that T_j can be left-truncated. Y_{1j} is then observable only if $T_j \geq Y_{0j}$, where Y_{0j} is the left truncation random variable. The Cox models assumes that the instantaneous hazard at time t for the jth individual is

$$h(t|x_j) = h_0(t)exp(\eta_j)$$

where $h_0(t)$ is the unknown baseline hazard function and η_j is the prognostic index equals to $x'_j\beta$, where $\beta = (\beta_1, ..., \beta_P)$ are the regression coefficients vector. The classical estimation is based on the maximum partial likelihood $l(\beta)$ [23]. The lasso shrinks the regression coefficients towards zero by penalizing by the sum of their absolute value [145] :

$$\hat{\beta} = \arg\max\{l(\beta) - \lambda \sum_{p=1}^{P} |\beta_p|\}$$

where the tuning parameter λ is positive.

4.2.2 Time dependent ROC curve

The question is to evaluate the predictive capacity of the estimated prognostic index $\hat{\eta} = x\hat{\beta}$. As explained by Molinaro et al. [100], the goal is to build a rule implementing the information from $\hat{\eta}$ in order to predict T. Suppose that if $\hat{\eta}$ is higher than a critical value c, it predicts the occurrence of the event before τ. τ is the prognostic time. Respecting the definition of Heagerty et al. [63], the false negative rate (FNR) and false positive rate (FPR) for a prognosis up to time τ are respectively $FNR_\tau(c) = P(\hat{\eta} \leq c|T \leq \tau)$ and $FPR_\tau(c) = P(\hat{\eta} > c|T > \tau)$. The prognostic capacities of $\hat{\eta}$ can be summarized by the time dependent ROC curve and the area under this curve (AUC). $1 - FNR_\tau(c)$ is plotted against $FPR_\tau(c)$ for all the thresholds $c : \{FPR_\tau(c), 1 - FNR_\tau(c), c \in \Re\}$.

These probabilities can be non-parametrically estimated :

$$\widehat{FNR}_\tau(c) = \{\hat{G}(c) - \hat{S}(-\infty, \tau) + \hat{S}(c, \tau)\}/\{1 - \hat{S}(-\infty, \tau)\} \quad (4.1)$$

$$\widehat{FPR}_\tau(c) = \hat{S}(c, \tau)/\hat{S}(-\infty, \tau) \quad (4.2)$$

where $\hat{G}(c)$ is the empirical distribution function of $\hat{\eta}$ in c, which is equal to

$$N^{-1} \sum_{j=1}^{N} \delta(\hat{\eta}_j < c).$$

$\delta(a)$ equals 1 if a is true and 0 otherwise. $S(c, \tau)$ is the bivariate survival function of $\hat{\eta}$ and T. Because $h_0(t)$ is unknown, $S(c, \tau)$ can be estimated using the Akritas estimator [4] :

$$\hat{S}(c, \tau) = N^{-1} \sum_{j=1}^{N} \hat{S}(\tau | \hat{\eta} = \hat{\eta}_j) \delta(\hat{\eta}_j > c) \quad (4.3)$$

The estimation of the conditional survival probability, $\hat{S}(\tau | \hat{\eta} = \hat{\eta}_j)$, is based on the nearest neighbors. Let K_{jl} be the indicator that a patient l is eligible in the neighbors of the patient j according to $\hat{\eta}_j$: $K_{jl} = \delta(|\hat{G}(\hat{\eta}_j) - \hat{G}(\hat{\eta}_l)| < \pi)$. π represents the proportion of included neighbors. From these definitions, the estimation of conditional survival respects the same principle as Kaplan and Meier :

$$\hat{S}(\tau | \hat{\eta} = \hat{\eta}_j) = \prod_{y_{1j} \leq \tau} \{1 - d(y_{1j})/R(y_{1j})\}, \quad (4.4)$$

where $d(y_{1j})$ and $R(y_{1j})$ are respectively the number of events and the number of patients at risk at time y_{1j}. Formally :

$$d(y_{1j}) = \sum_{l=1}^{N} K_{jl} \delta(y_{1j} = y_{1l}) \Delta_l \quad (4.5)$$

$$R(y_{1j}) = \sum_{l=1}^{N} K_{jl} \delta(y_{1j} \leq y_{1l}) \quad (4.6)$$

This estimator is valid if the data are not left-truncated : $y_{0j} = 0$, $\forall j = 1, ..., N$. The left-truncation can be included by correcting the number (4.6) of at risk patients [109] :

$$R(y_{1j}) = \sum_{l=1}^{N} K_{jl} \delta(y_{0j} \leq y_{1l} \leq y_{1l}) \quad (4.7)$$

4.2.3 0.632+ estimator

As explained by Ransohoff [127], molecular markers are set to revolutionize the process of evaluating prognosis and diagnosis. However, the data overfitting is often associated with high expectations but also with disappointment when original results cannot be reproduced. The best method is to validate the results in a independent and large dataset. Because it is often impossible (cost, availability of samples, etc.), a common approach is to split the data into a training and test set. Molinaro et al. demonstrated the limitations of this method based on theoretical arguments and simulation results [100]. Previous studies have found that the bootstrap-based resampling methods are superior to cross-validation and leave-one-out cross-validation. As recent studies in this field of microarray-based prediction [133, 100, 7, 51], we focused on the 0.632+ bootstrap estimator. This methodology was recently proposed by Adler and Lausen for estimating ROC curves when data are complete [2], which is not adapted for time-to-event analysis with censored/truncated data.

Consider B bootstrap samples ($b = 1, ..., B$) of size N with replacement. Respecting the same principle as the cross-validation, the observations included in the bootstrap sample are used for training and for estimating the regression parameters $\hat{\beta}^b$. The definition of the model complexity also has to be included in each bootstrap iteration in order to avoid overoptimistic results. Schumacher et al. [133] found that selection criteria, such as bayesian information criterion, fail to compute the tuning parameter λ and they used 5-fold cross-validation [110]. But the authors were limited by the computational burden and the λ value was obtained using the full dataset. We proposed the use of the algorithm recently reported by Goeman [56] that efficiently estimates the Cox model with lasso penalty. We also used 5-fold cross-validation to estimate the tuning parameter $\hat{\lambda}^b$ of each sample b ($b = 1, ..., B$).

Based on the estimations $\hat{\beta}^b$ and $\hat{\lambda}^b$, the FNR (4.1) and the FPR (4.2) can be calculated using the data not included in the bootstrap sample, respectively $\widehat{FNR}_\tau^b(c)$ and $\widehat{FPR}_\tau^b(c)$. The *bootstrap cross-validation* rates are obtained by the average :

$$\widehat{FNR}_\tau^{BCV}(c) = B^{-1}\sum_{b=1}^{B}\widehat{FNR}_\tau^b(c) \tag{4.8}$$

$$\widehat{FPR}_\tau^{BCV}(c) = B^{-1}\sum_{b=1}^{B}\widehat{FPR}_\tau^b(c) \tag{4.9}$$

The corresponding ROC curve is $\{\widehat{FPR}_\tau^{BCV}(c), 1 - \widehat{FNR}_\tau^{BCV}(c), c \in \Re\}$ and $\widehat{AUC}_\tau^{BCV}$ is the corresponding bootstrap cross-validation estimation of the AUC at time τ. If N is sufficiently large ($N > 40$), the probability that an individual appears in the training

set is $1-(1-1/N)^N \approx 0.632$. The proportion of 0.368 is composed of completely independent data from the replicated data included in the training set, which causes an underestimation of the prognostic capacity. To correct for this, Efron proposed the 0.632 estimator [34] :

$$\widehat{FNR}_\tau^{632}(c) = 0.368\,\overline{FNR}_\tau(c) + 0.632\,\widehat{FNR}_\tau^{BCV}(c) \quad (4.10)$$

$$\widehat{FPR}_\tau^{632}(c) = 0.368\,\overline{FPR}_\tau(c) + 0.632\,\widehat{FPR}_\tau^{BCV}(c) \quad (4.11)$$

where $\overline{FNR}_\tau(c)$ and $\overline{FPR}_\tau(c)$ are the *apparent* rates. These rates are calculated similarly to (4.8) and (4.9), but using the B training sets. More precisely, based on $\hat{\beta}^b$ and $\hat{\lambda}^b$, the apparent FNR and FPR can be calculated using only the data included in the bootstrap sample, respectively $\overline{FNR}_\tau^b(c)$ and $\overline{FPR}_\tau^b(c)$:

$$\overline{FNR}_\tau(c) = B^{-1}\sum_{b=1}^{B}\overline{FNR}_\tau^b(c) \quad (4.12)$$

$$\overline{FPR}_\tau(c) = B^{-1}\sum_{b=1}^{B}\overline{FPR}_\tau^b(c) \quad (4.13)$$

The 0.632 ROC curve is $\{\widehat{FPR}_\tau^{632}(c), 1-\widehat{FNR}_\tau^{632}(c), c\in\Re\}$ and $\widehat{AUC}_\tau^{632}$ is the corresponding 0.632 estimation of the AUC at time τ.

The rates (4.10) and (4.11) may be associated with overestimations if the apparent estimations are very small when overfitting data. Efron and Tibshirani improved the correction with the 0.632+ estimator [36]. The no-information rates associated with FNR and FPR may be estimated using all the data and considering the independence between $\hat{\eta}$ and T : $\hat{\gamma}_{N,\tau}(c) = 1-\hat{\gamma}_{P,\tau}(c) = \hat{G}(c)$. These no-information probabilities are used to define the overfitting rates :

$$\hat{r}_{N,\tau}(c) = \{\widehat{FNR}_\tau^{BCV}(c) - \overline{FNR}_\tau(c)\}/\{\hat{\gamma}_{N,\tau}(c) - \overline{FNR}_\tau(c)\} \quad (4.14)$$

$$\hat{r}_{P,\tau}(c) = \{\widehat{FPR}_\tau^{BCV}(c) - \overline{FPR}_\tau(c)\}/\{\hat{\gamma}_{P,\tau}(c) - \overline{FPR}_\tau(c)\} \quad (4.15)$$

We assigned these rates to 0 for negative values and to 1 for values higher than 1. The 0.632+ estimations of the false negative and positive rates are thus defined by :

$$\begin{aligned}\widehat{FNR}_\tau^{632+}(c) &= \{1-\psi(\hat{r}_{N,\tau}(c))\}\,\overline{FNR}_\tau(c)\\ &+ \psi(\hat{r}_{N,\tau}(c))\,\widehat{FNR}_\tau^{BCV}(c) \quad (4.16)\\ \widehat{FPR}_\tau^{632+}(c) &= \{1-\psi(\hat{r}_{P,\tau}(c))\}\,\overline{FPR}_\tau(c)\\ &+ \psi(\hat{r}_{P,\tau}(c))\,\widehat{FPR}_\tau^{BCV}(c) \quad (4.17)\end{aligned}$$

where $\psi(x) = 0.632/(1-0.368x)$. The corresponding 0.632+ ROC curve is $\{\widehat{FPR}_\tau^{632+}(c), 1 - \widehat{FNR}_\tau^{632+}(c),\ c \in \Re\}$ and $\widehat{AUC}_\tau^{632+}$ is the corresponding 0.632+ estimation of the AUC at time τ. This method has been implemented in an R package `ROC632` available at www.divat.fr/en/softwares or upon request from authors.

4.3 Simulation studies

4.3.1 Methods

Different values of N were used : 60, 125 and 250. The times-to-event were simulated using a Weibull PH model. The feature expressions were obtained by assuming independent standard normal distributions. The censoring times were simulated independently respecting Exponential distributions. The prognostic time was fixed at 6 months. The parameters of the Exponential distributions were defined to obtain three different censoring rates at 6 months : 30, 50 and 70%. We distinguished the 3 following scenarios :

#1 **Total overfitting.** Among 750 features, no feature is associated with the time-to-event, β = (0, 0, ..., 0). We fixed the shape and scale parameters of the Weibull distribution at respectively 4 and 55. In this scenario, the true value of the AUC is 0.5.

#2 **No overfitting.** Among 3 features, 2 features are associated with the time-to-event, β = (log(1.2), -log(1.2), 0). We fixed the shape and scale parameters of the Weibull distribution at respectively 4 and 148. In this scenario, the true value of the AUC is obtained by using the apparent estimator. Indeed, there is no overfitting because only 3 features are analyzed by using at least 60 individuals.

#3 **High overfitting.** Among 750 features, 2 features are associated with the time-to-event, β = (log(1.2), -log(1.2), 0, ..., 0). We fixed the shape and scale parameters of the Weibull distribution at respectively 4 and 148. In this scenario, the overfitting is high since the number of features is much higher than the number of individuals. Nevertheless, the only two significant features have a similar effect than the two significant features in the second scenario. Thus, the true values of the AUC are similar to the apparent estimations in the second scenario (where there is no overfitting).

For each possible combination of the overfitting levels, sample sizes and censoring rates, 250 samples were simulated. $B = 100$ bootstrap iterations were used for each simulated sample. Because Simon et al. recently described a cross-validation (CV) based method for estimating time-dependent ROC curves [137], we also added this estimator to the comparisons.

4.3.2 Results

The results are presented in Table (4.1). As expected in the first scenario, the apparent estimator overestimated the AUC with values between 0.69 and 0.80. The 0.632 estimator appeared also to be overoptimistic. In contrast, the BCV and 0.632+ estimations were similar to the true value of 0.50. These conclusions can be made regardless of the sample size and the censoring rate.

In the second scenario, the BCV, the 0.632 and the 0.632+ estimations were similar to the true apparent estimations if the sample size was large ($N = 250$). These adequate estimations were also observed for $N = 125$ with 30% of censoring. However, the three methods underestimated the prognostic capacity for lower sample size and/or higher censoring rates. Theses underestimations were greater for the BCV.

In the third scenario, one can notice that the 0.632+ estimations were similar to the true AUC for $N = 250$ with 30 or 50% of censoring. These adequate estimations were also observed for $N = 125$ with 30% of censoring. For the worst situations (lower sample sizes and/or higher censoring rates), the 0.632+ estimator underestimated the prognostic capacity. This underestimation was higher for the BCV estimator, but lower for the 0.632 estimator.

Regardless the scenario, the CV estimations were associated with the more important standard deviations.

4.4 Application to the DLBCL study

In this study, Rosenwald et al. [130] evaluated tumor samples from 240 DLBCL patients treated with anthracycline-based therapy. The full dataset was split into training (n=160) and test (n=80) sets. They confirmed the existence of the two DLBCL subgroups described previously, GCB-like and ABC-like. The overall survival was significantly different among the subgroups, with 5-year survivals of 60% for the GCB-like and 35% for ABC-like subgroups. An additional third subtype was described with a 5-year survival of 39%. These groups were first obtained by a clustering analysis using all the features.

Several authors have reanalyzed these data. In particular, Segal concluded that the use of the Cox model with lasso penalty appears to be the best approach as it balances between prediction and interpretation [135]. Time-dependent ROC curves were used in the test set. Gui and Li used the same splitting and ROC-based validation [58]. Because splitting the data worsens the overfitting and does not use the available information

Table 4.1: *Mean and standard deviation (between brackets) of the time dependent AUC at 6 months obtained from the 250 simulated samples for each combination of overfitting level, sample size and censoring rate. B =100 bootstrap replications are performed for computing the apparent, the bootstrap cross-validation (BCV), the 0.632, the 0.632+ and the 5-fold cross-validation (CV) estimations.*

Level *	N	cens. rate	Apparent	BCV	0.632	0.632+	5-fold CV
0/750	60	0.3	0.743(0.103)	0.518(0.071)	0.608(0.059)	0.528(0.080)	0.523(0.192)
		0.5	0.781(0.097)	0.504(0.058)	0.614(0.053)	0.513(0.065)	0.495(0.177)
		0.7	0.745(0.108)	0.500(0.058)	0.596(0.058)	0.510(0.063)	0.489(0.181)
	125	0.3	0.782(0.071)	0.514(0.050)	0.623(0.040)	0.527(0.057)	0.485(0.138)
		0.5	0.741(0.082)	0.498(0.054)	0.595(0.047)	0.508(0.062)	0.492(0.149)
		0.7	0.701(0.118)	0.502(0.058)	0.581(0.059)	0.514(0.066)	0.483(0.143)
	250	0.3	0.799(0.035)	0.508(0.039)	0.625(0.027)	0.521(0.046)	0.497(0.086)
		0.5	0.777(0.062)	0.505(0.046)	0.612(0.039)	0.516(0.054)	0.472(0.111)
		0.7	0.693(0.121)	0.505(0.059)	0.579(0.061)	0.519(0.067)	0.518(0.145)
2/3	60	0.3	0.846(0.050)	0.789(0.063)	0.810(0.057)	0.808(0.058)	0.822(0.118)
		0.5	0.818(0.082)	0.742(0.092)	0.770(0.086)	0.765(0.089)	0.740(0.142)
		0.7	0.749(0.119)	0.656(0.110)	0.690(0.110)	0.677(0.116)	0.650(0.150)
	125	0.3	0.842(0.041)	0.818(0.044)	0.827(0.042)	0.826(0.042)	0.835(0.061)
		0.5	0.840(0.049)	0.803(0.058)	0.816(0.054)	0.815(0.055)	0.795(0.089)
		0.7	0.798(0.076)	0.732(0.086)	0.756(0.081)	0.751(0.084)	0.722(0.103)
	250	0.3	0.838(0.032)	0.825(0.033)	0.830(0.033)	0.830(0.033)	0.836(0.039)
		0.5	0.836(0.033)	0.817(0.037)	0.824(0.035)	0.824(0.035)	0.814(0.050)
		0.7	0.828(0.047)	0.795(0.054)	0.807(0.051)	0.806(0.051)	0.774(0.074)
2/750	60	0.3	0.902(0.039)	0.646(0.083)	0.751(0.058)	0.677(0.090)	0.725(0.144)
		0.5	0.869(0.071)	0.591(0.091)	0.702(0.071)	0.614(0.102)	0.629(0.195)
		0.7	0.827(0.086)	0.534(0.072)	0.649(0.061)	0.549(0.083)	0.554(0.168)
	125	0.3	0.905(0.032)	0.783(0.051)	0.831(0.036)	0.819(0.048)	0.833(0.062)
		0.5	0.890(0.041)	0.729(0.073)	0.791(0.053)	0.765(0.075)	0.779(0.094)
		0.7	0.834(0.087)	0.619(0.086)	0.702(0.070)	0.647(0.097)	0.645(0.140)
	250	0.3	0.898(0.023)	0.811(0.027)	0.845(0.022)	0.843(0.025)	0.832(0.037)
		0.5	0.903(0.025)	0.789(0.037)	0.833(0.027)	0.825(0.034)	0.813(0.050)
		0.7	0.881(0.052)	0.729(0.060)	0.785(0.047)	0.766(0.059)	0.757(0.079)

* Number of significant features versus total number of features.

efficiently, we used the 0.632+ estimation of the time-dependant ROC curves. Applying the same strategy as previous authors, the missing data were replaced according to the mean expression level of the nearest 8 genes. Respecting the previous simulation results, the prognostic times were $\tau = 2, 3, ..., 15$ years in order to maintain the censoring rates within 52.5%. $B = 100$ bootstrap iterations were used.

The results are presented in Figure (4.1). The AUCs obtained using the 0.632+ estimator were between 0.70 and 0.65 (from 2 to 14 years). The area decreased with prognostic time, illustrating that long-term failures are often more difficult to predict. The overfitting was high with an apparent AUC around 0.95. In contrast, the prognostic capacity appeared to be underestimated when using the BCV estimator. The 0.632+ estimations were less optimistic than the 0.632 estimations. The AUCs obtained in the test set by Segal [135] were more like those of the 0.632+ estimations.

Figure (4.2) presents the previous estimations by *(i)* re-estimating the tuning parameter at each bootstrap iteration, *(ii)* estimating the tuning parameter on the full data

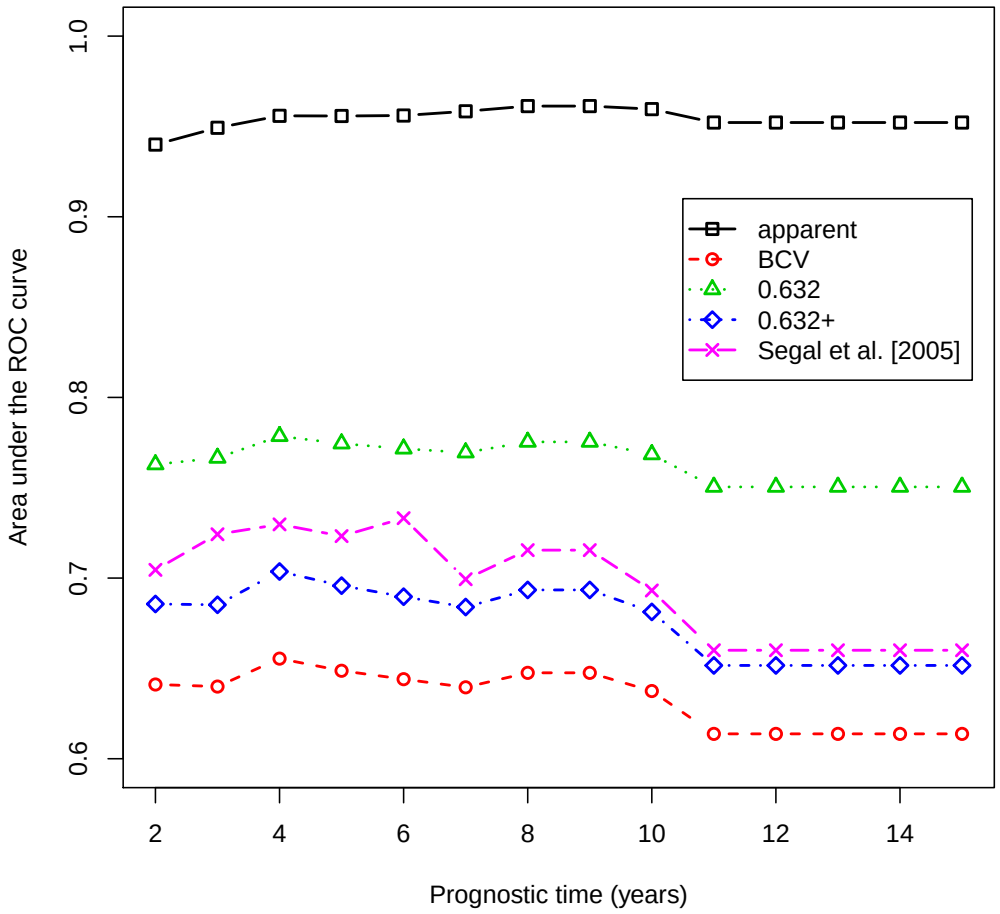

Figure 4.1: *AUC according to the prognostic times and the different estimators. Results obtained by analyzing the DLBCL data, by using the Cox model with lasso penalty and by re-estimating the tuning parameter for each bootstrap iteration (5-fold cross-validation).*

set and re-using the value at each bootstrap iteration to select features, and *(iii)* using the same vector of features (initially defined with the tuning parameter on the full data set) at each bootstrap iteration. As expected, the highest apparent AUC was obtained with the first approach : the model building is completely re-performed at each iteration with the maximum adjustment to the training set. Moreover, the third approach gave a significant overestimation of the AUC obtained by resampling (BCV, 0.632 or 0.632+) because all steps of model building are not performed within each bootstrap sample : the selection of features is ignored. The associated overoptimistic results have already been demonstrated [136].

Interestingly for the BCV and 0.632+ methods, the AUC obtained by using re-sampling were closed for *(i)* and *(ii)*, while overestimation was expected in the second approach, i.e. when the tuning parameter was fixed among the bootstrap iterations but the features selection at each iteration was retained. This similarity was not observed for the 0.632 method with an overestimation in *(ii)*.

4.5 Discussion and Conclusions

In this paper, we propose a 0.632+ estimator of the time-dependent ROC curve to exceed the latter limitations. First, ROC-based methodologies are well-accepted in the community of biologists and physicians. The AUC represents the ability of a prognostic factor to correctly distinguish patients who will develop events in the future from those who will not. In the application to DLBCL data, for example, we estimated the AUC at 0.68 for a prognosis up to 10 years (Figure 4.1). Thus, a patient who will die before 10 years has a 68% chance of having a score higher than a patient who will be alive at this time. Even if the 0.632+ computation may be difficult to understand, the interpretation of the resulting AUC is straightforward. Second, the time-dependent FNR and FPR are conditional distributions of the prognostic index assuming the distribution of the time-to-event). These indicators do not depend on the incidence of the event. The AUC is an invariant prognostic indicator among populations with different incidences. Third, the tuning parameter is re-estimated at each bootstrap iteration according to the recent efficient algorithm of Goeman [56]. Thus, the full data set is never used.

We validated the proposed methodology by simulating three overfitting levels (total, high, low) according to different sample sizes and censoring rates. If the available information is sufficient ($N = 250$ with less than 50% of censoring or $N = 125$ with less than 30% of censoring), the 0.632+ estimators appeared to be similar to the true expected areas under the curve, regardless of the overfitting level. In contrast, the 0.632 estimator appeared overoptimistic when all the features are independent to the time-to-event

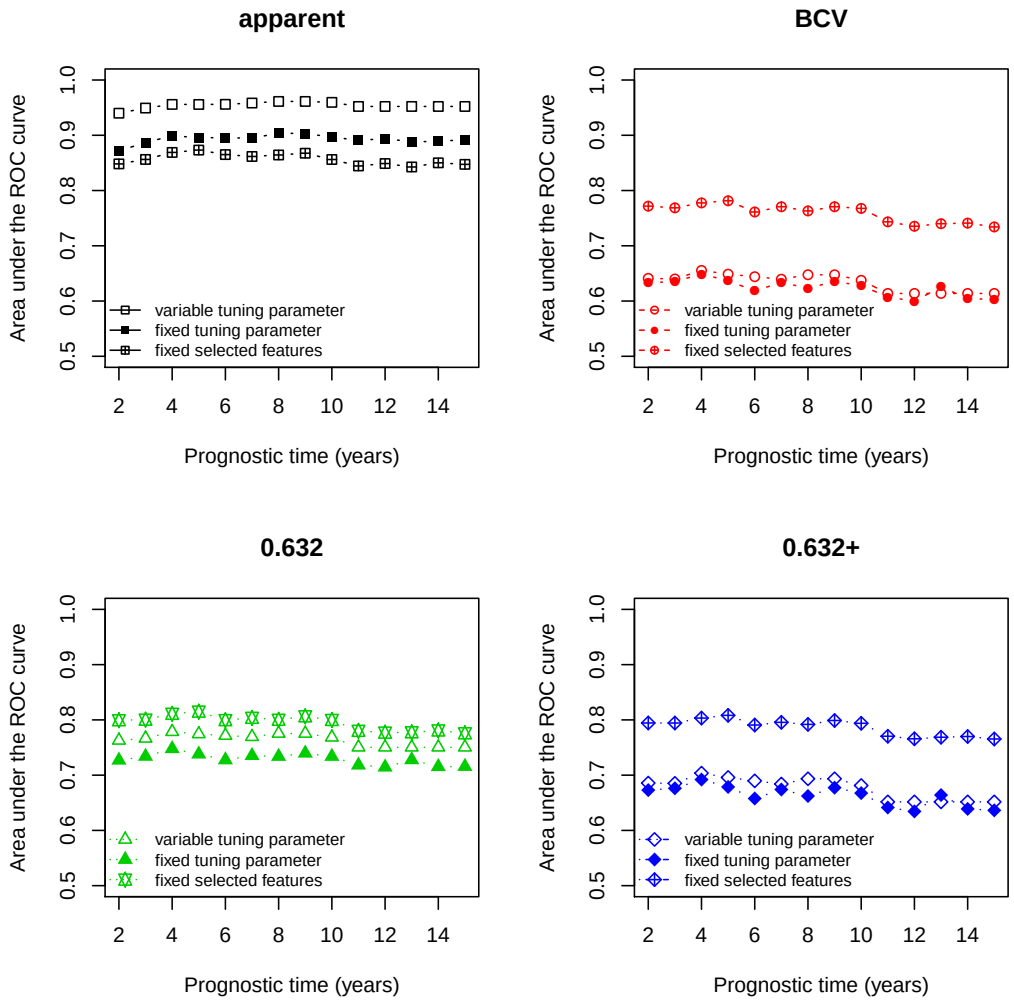

Figure 4.2: *AUC according to the prognostic times and the different estimators. Three approaches are compared : the tuning parameter is re-estimated at each bootstrap iteration, the tuning parameter is initially estimated on the full data set and the value is used at each iteration for the features selection, and the tuning parameter and the selected features are both initially defined on the full data set.*

distribution. Moreover, the BCV estimator underestimated the true prognostic capacity when the overfitting was high. Finally, the CV estimator was associated to very large standard deviations. Based on these results and because the overfitting level is unknown in practice, the 0.632+ appears to have the best behavior. Nevertheless, we also have to notice that a minimum of 125 individuals are required with less than 30% of censoring or a minimum of 250 individuals for a censoring rate up to 50%. It is not possible to compute all the possible situations, but it is clear that these methods should not be used for small sample sizes (less than 125 individuals), in opposition to the current literature in medicine or in biology.

We applied the methodology to the DLBCL data [130], which includes 240 patients with about 50% of censoring for the longer prognostic time. The AUCs obtained by using the 0.632+ estimator were between 0.70 and 0.65. This illustrates the utility of this signature to predict mortality up to 15 years, but it also illustrates that this signature alone is not sufficient for medical decision-making. Indeed, a patient who will die before 10 years has a 32% chance of having a score lower than a patient who will be alive at this time. The areas obtained using the 0.632+ estimator were very similar to the revisited analysis by Segal [135], who split the sample for independent validation. Regardless of the similar results to Segal, our approach has several advantages by using all the data. First, the split is associated with a increasing type II error in the features selection step. Second, a particular split of the data is possible. If very different training and test sets are obtained, the results may be under optimistic. Third, it avoids the temptation to re-perform the analysis with different splits so as to choose the best results.

Again using the DLBCL data, we evaluated the impact of not including all the steps of the model building in the algorithm. Overlaps between the bootstrap training and test sets can cause an overestimation of the prediction capacity [76]. The reference approach is to re-estimate the tuning parameter and to select the features at each bootstrap iteration. Nevertheless, the 0.632+ estimator appeared to be robust when the tuning was estimated on the full data set and re-used at each step for the features selection. Even though this limit was acknowledged by Schumacher et al. [133], their assumption seems reasonable on this real data set. This assumption is associated to an important time-saving. Nevertheless, in agreement with Simon et al. and because results from DLBCL data are not generalizable, the complete re-estimation of the model at each iteration remains less open to criticism [136].

The global approach (resampling, calibration, validation) can be affected by the model misspecification, as all the regression analysis. Given the high dimension data, it is not realistic to test every assumption for every feature. The choice of the survival model was not treated in this paper ; only the Cox model with lasso penalty was used. The three main assumptions are the hazard proportionality, the log linearity and the condi-

tionally independent censoring given features. Future comparisons of different modeling approaches by using the 0.632+ time dependent ROC curves will be interesting. These may include the ridge Cox regression [151], the generalization of partial least squares [92], the random survival forests [72] or the structure-based variable selection for survival data [88]. This is beyond the scope of this paper, as here we focused on the definition and the validation of the 0.632+ estimator for time dependent ROC curves. However, the algorithm has been formulated in a general way and can be subsequently applied to other models. Moreover, the proposed estimator of the time-dependent ROC curve is non-parametric.

In summary, we demonstrated in this paper that the 0.632+ estimator of time-dependent ROC curve is useful to estimate the predictive accuracy of prognostic signature based on high-dimensional data. Simulations illustrated that it outperforms the other approaches, especially the cross-validation solution recently proposed by Simon et al. [137]. The application on DLBCL data constitutes an example of the simple interpretation of the results in term of medical decision making. We propose an R package entitled `ROC632` and available at www.divat.fr/en/softwares. This algorithm includes all the steps of the model building.

Acknowledgement

This work was supported by the RTRS "The Fondation de Co-operation Scientifique - CENTAURE", by grant from the French National Agency of Research ANR-11-JSV1-0008-01. We also thank Dr. J Ashton-Chess for editing the manuscript.

- Chapitre 5 -

A literature-based approach to evaluate the predictive capacity of a marker using time-dependent Summary Receiver Operating Characteristics (SROCt)

Combescure C., Daurès JP. and Foucher Y.
Statistical Methods in Medical Research. 2012. Online First.

5.1 Introduction

The correlation between potential prognostic markers and disease evolution is a major issue of modern clinical research. Given the abundance of papers devoted to the evaluation of the prognostic performances of biomarkers, methods to synthesize the results from the literature are needed. Since most of the published studies are based on survival analyses, several authors have performed meta-analyses using survival differences. For example, Parmar et al. [111] and Williamson et al. [155] suggested preferentially summarizing time-to-event data from different trials using an overall Hazard Ratio (HR), i.e. the usual risk indicator in survival analysis. Tierney et al. [146] subsequently provided a step-by-step guidance on how to calculate this overall HR. Their method is adapted to study the risk differences among well-defined groups (e.g. two treatments, presence/absence of exposure), but is limited when the study focuses on a continuous marker. First, different cut-offs may be used to define the groups of comparison and one can expect the HRs between groups to depend on the cut-offs choices. Consequently, the assumption that the observed HRs among studies are derived from a single common HR is misguided. Moreover, the number of groups may also vary across studies, making meta-analysis of HRs no longer possible.

Another approach, consisting of summarizing the survival curves through a meta-analytic regression model [27], was recently extended to incorporate random effects [10]. More flexible than the meta-analysis of HRs, the cut-offs may be introduced as covariates. However, the results of such survival models are difficult to interpret in terms of prognostic capacity. The reason for this is that the HR or the survival probabilities do not deal with the decision errors based on the marker. This indicator is thus insufficient to enable a conclusion on predictive capacity. For instance, Buckley et al. [16] demonstrated that the C-Reactive protein level is independently associated with the incidence of coronary heart disease by pooling hazard ratios. Nevertheless, the authors also acknowledged that *the clinical implication of this finding is less clear, because the pooled risk ratio does not necessarily measure the usefulness of C-Reactive protein level in reclassifying intermediate-risk persons.* As such, the terms "correlation" and "prediction" are often confused [85]. Heagerty et al. [63] underlined this issue in prognostic studies and developed the time-dependent Receiver Operating Characteristics (ROCt) theory. The data interpretation is similar to usual ROC curves, but the methodology has the advantage of taking the survival process into account.

The aim of this paper was to develop alternative method for literature-based meta-analysis of predictive markers. The aggregated information, available in the published studies, was used for the estimation of summary ROCt (SROCt) curves. The next section describes the method used and section 3 applies this method to a multi-centric cohort of kidney transplant patients (DIVAT). Each center was considered as one published study and the SROCt curves were obtained from the corresponding aggregated data. The results were then compared with the ROCt curves obtained from the individual data. This analysis constitutes a validation step, since the meta-analysis based on individual data can be considered as the gold standard [108, 32]. Section 4 applies this method to the meta-analysis already published by Azambuja et al. [26], the aim being to evaluate whether KI-67 can be considered as a good prognostic marker for breast cancer survival.

5.2 Estimation of the SROCt Curve

5.2.1 Principles of the method

Let (X, T) be respectively the biomarker and the time-to-event variables, where $X \in \Re$ and $T > 0$. X is measured at the time origin. Without loss of generality, we assume that the risk of failure increases with X. As regards to the paper by Heagerty et al. [63] in the Kaplan-Meier based estimation (the authors also proposed alternative approach based on nearest neighbor estimation), the sensitivity and the specificity at

cut-off x for a prognosis up to time t are defined as follows :

$$se(x|t) = \{Pr(X > x) - Pr(T > t, X > x)\}/Pr(T \leq t) \quad (5.1)$$

$$sp(x|t) = 1 - \{Pr(X > x, T > t)/Pr(T > t)\} \quad (5.2)$$

The sensitivity is the probability that the biomarker is higher than x for a patient with failure before time t, i.e. $se(x|t) = Pr(X > x|T \leq t)$. The specificity is the probability that the biomarker is lower than x for a patient with no failure before time t, i.e. $sp(x|t) = Pr(X \leq x|T > t)$. The assessment of these probabilities are based on the joint probability $Pr(X > x, T > t)$. In papers focusing on the performances of a prognostic factor, the results are usually summarized by survival curves after categorization of the biomarker and/or by HRs derived from a survival regression model. The published HRs are not helpful in our context because they cannot address the joint probability. Survival curves provide more information and were used for estimation of SROCt curves.

The distribution of the marker may depend of the study. Let $f_k(x)$ be the probability density function of the marker X in the kth study. We assumed a common underlying distribution for all studies. The variability between studies was modeled with a random variable $\omega : f_k(x) = f(x|\omega_k)$, where ω_k is the value of ω for the kth study. The probability of observing a marker between a and b for a patient included in the kth study is :

$$p_k(a, b) = \int_a^b f(x|\omega_k)dx \quad (5.3)$$

Heterogeneous values of ω_k would mean the distribution of the marker varies substantially across the studies. We denoted $S_k(t|x)$ the survival function of T in the kth study for a given value of the marker x. Similarly to the marker distribution, the time-to-event distribution was modeled as follows : $S_k(t|x) = S(t|\nu_k, x)$, where ν_k is the value for the kth study of the random variable ν. There is no reason to expect a correlation among studies between the distribution of the marker and the distribution of the failure time according to this marker. ω and ν were thus assumed to be independent. For the kth study, the bivariate probability, that no failure occurs until t and that the marker lies in between a and b, is :

$$S_k(t, a, b) = \int_a^b S(t|\nu_k, x)f(x|\omega_k)dx \quad (5.4)$$

The sensitivity and specificity at a cut-off x for a prognosis up to time t can be derived by introducing Expressions 5.3 and 5.4 to Equations 5.1 and 5.2 :

$$se_k(x|t) = \{p_k(x, +\infty) - S_k(t, x, +\infty)\}/\{1 - S_k(t, -\infty, +\infty)\} \quad (5.5)$$

$$sp_k(x|t) = 1 - \{S_k(t, x, +\infty)/S_k(t, -\infty, +\infty)\} \quad (5.6)$$

These probabilities can be obtained regardless of the study by replacing the random variables by their mean. The overall SROCt curve for a prognosis up to time t is the plot of the sensitivity as a function of one minus the specificity for the possible cut-offs x.

5.2.2 Estimation procedure

Let K equal the number of studies of size N_k $(k = 1, ..., K)$. From published studies, one can expect to obtain the marker intervals defining the groups, the number of individuals in each group and the survival curve for each group. The biomarker may be categorized in different ways. The number of groups in the kth study is denoted G_k. The gth group $(g = 1, ..., G_k)$ of size N_{kg} is defined by the marker interval $[c_{k,g-1}, c_{k,g}[$ with $c_{k,0} = -\infty$ and $c_{k,G_k} = +\infty$. The observed probabilities of being in the gth group equal N_{kg}/N_k and the corresponding observed survival probabilities are measured every year, month or day depending on the time scale used (Kaplan-Meier survival curves are usually provided).

The estimation procedure consists of two steps : the parameters of the marker distribution are first estimated and then the parameters of the survival according to the marker are assessed. We assume that an appropriate bijective function exists, termed $h(.)$, which follows a Gaussian distribution :

$$f(h(x)|\omega_k) = \frac{1}{\sigma\sqrt{2\pi}} exp\Big(\frac{-(h(x) - \mu - \omega_k)^2}{2\sigma^2}\Big) \tag{5.7}$$

with $\sigma > 0$, $\mu \in \Re$ and $\omega_k \sim N(0, \sigma_\omega)$. μ and σ are the mean and the standard deviation of $h(X)$. The normality of $h(x)$ was adapted for both of the following applications and this assumption will hold true in a lot of analyses. However, other distributions can also be used. We used a flexible survival model in order to avoid any restrictive assumptions. The hazard function associated with $S(t|\nu_k, h(x))$ was defined as a piecewise constant function with a specific association of $h(X)$ at each interval $[\tau_0, \tau_1[$, $[\tau_1, \tau_2[$,... and $[\tau_{L-1}, \tau_L[$, with $\tau_0 = 0$ and $\tau_L = +\infty$:

$$\lambda(t|\nu_k, h(x)) = exp\Big(\nu_k + \sum_{l=0}^{L}(\beta_{l,1} + \beta_{l,2}h(x)) \times \mathbb{1}\{t > \tau_l\}\Big) \tag{5.8}$$

in which $\beta_{l,1}, \beta_{l,2} \in \Re$ are the regression parameters specific to the interval l and $\nu_k \sim N(0, \sigma_\nu)$.

To assess the Non-Linear Mixed Models 5.7 and 5.8, we used the `R` package `nlme` [124, 116] based on maximization of the restricted log-likelihood. We propose `R` package `SROCt`

is available at `www.divat.fr/en/softwares/sroct`. Each observation was weighted by the corresponding number of at-risk individuals (optional argument `weights`, see section 2.4 for more details). The integrals involved in Equations 5.3 and 5.4 were computed using the 30-points Gauss-Legendre quadrature.

The area under the estimated SROCt curve was obtained by the trapezoidal rule. It is impossible to develop an analytic formulation of the confidence interval for these areas. We proposed a simulation procedure (1000 iterations). At each iteration, the parameter values were simulated according to their estimated values and standard deviations. The corresponding areas under the SROCt curve were then computed. The $2.5th$ and $97.5th$ percentiles of the empirical distribution represented the 95% confidence interval ($CI_{95\%}$). We did not perform non-parametric bootstrap resampling because we needed to resample the center random effects, as well as the within-group effects.

5.2.3 Checking the validity of the models

Goodness-of-fit was evaluated by plotting the predicted values of the regressions 5.3 and 5.4 versus the observed probabilities. Moreover, the residuals of the model 5.4 were also plotted against the survival time. If extreme residuals were identified, the corresponding studies were removed from the analysis to test the robustness of the results.

5.2.4 Extracting the data from published

One can collect the size of each group at baseline, but in contrast to the good practice for survival plots described by Pocock et al. [117], many papers do not present the number of at-risk patients over time. We thus assumed that the number of at-risk patients is unknown during follow-up and we adapted the methodology detailed in the appendix of the paper by Parmar et al. [111] in order to calculate these numbers. Briefly, suppose that we want to calculate the number of censored observations between t_e and t_s in the gth group of the kth study. Let t_{kg}^{max} be the corresponding maximum follow-up time and $R_{kg}(t_e)$ the number of at-risk patient at the beginning of the previous interval. By assuming that the rate of censoring is constant and non-informative, the number of censoring observations should be close to $R_{kg}(t_e)(t_s - t_e)/(t_{kg}^{max} - t_s)$. Because the sample size at baseline and the changes in the survival probabilities are observed, the number of at-risk individuals can be recursively calculated from this baseline. The survival probabilities can be objectively extracted from a digitalized picture by using the `R` packages `ReadImages` and `digitize`, as proposed by Poisot [121].

5.3 Aggregated versus individual estimations

In order to validate the proposed method, we analyzed real data from a French multi-centric cohort of kidney transplant recipients (DIVAT). The medical objective was to evaluate whether 1-year serum creatinine (Cr) is a good predictive marker of graft failure. Cr is a breakdown product and is removed from the body by the kidneys. If kidney function is abnormal, blood Cr levels increase. Deaths with a functional kidney or returns to dialysis were considered as graft failures. We considered a sub-population of 4195 adult patients and who had received a first kidney graft between January 1996 and Jun 2008. Five centers participated : Nantes (n=1375), Nancy (n=797), Paris Necker (n=1031), Toulouse (n=507) and Montpellier (n=485). A total of 511 graft failures were observed (346 returns to dialysis and 165 deaths with a functional kidney).

We first performed the usual methodology by Heagerty et al. [63] based on the individual data regardless of the transplantation center. This approach is non-parametric, based on the Kaplan-Meier estimator of survival and on the empirical distribution of the marker. In order to calculate confidence intervals for the areas under the curves, 1000 bootstrap samples were performed. The $2.5th$ and $97.5th$ percentiles of the empirical distribution represented the limits of the 95% non-parametric bootstrap confidence interval.

Based on the same database, we constructed an aggregated dataset to perform a meta-analysis on 5 published mono-centric studies. In order to consistently remain as close as possible to published papers, we differentially chose the groups according to the 1-year Cr value, which resulted in 1 center with 5 groups $(Q_{20\%}, Q_{40\%}, ..., Q_{80\%})$ [1], 2 centers with 3 groups ($Q_{10\%}, Q_{90\%}$ and $Q_{33\%}, Q_{66\%}$) and 2 centers with 2 groups ($Q_{50\%}$). We chose 4 intervals to model the hazard function 5.8 : $\tau_1 = 2$, $\tau_2 = 4$ and $\tau_3 = 6$ years. The estimations of the parameters involved in the models 5.7 and 5.8 are presented in Table (5.1).

These detailed aggregated data are available at `www.divat.fr/en/softwares/sroct` for non-R users or in the `SROCt` package at the same address. The figures with the corresponding survival curves are also downloadable at this address.

The SROCt curves for a prognosis up to 3, 5 and 9 years after the Cr measurement are presented with continuous lines in Figure (5.1). The areas under these curves respectively equaled 0.70 ($CI_{95\%} = [0.61, 0.76]$), 0.69 ($CI_{95\%} = [0.61, 0.73]$) and 0.66 ($CI_{95\%} = [0.59, 0.70]$). They were similar to those obtained by the non-parametric analysis of individual data : 0.72 ($CI_{95\%} = [0.68, 0.75]$), 0.70 ($CI_{95\%} = [0.66, 0.72]$) and

1. $Q_{\alpha\%}$: $\alpha\%-$quantile of the 1-year Cr value.

Table 5.1: *Estimation of the parameters and their standard errors for the analysis of kidney transplant recipients (standard errors of random effects : $\sigma_\omega = 0.03$, $\sigma_{\upsilon_k} = 0.34$).*

	Estimation	Standard Error
μ	4.86	0.01
σ	-1.19	0.02
$\beta_{0,1}$	-16.11	5.54
$\beta_{0,2}$	2.46	1.07
$\beta_{1,1}$	0.47	11.24
$\beta_{1,2}$	-0.04	2.17
$\beta_{2,1}$	3.76	10.81
$\beta_{2,2}$	-0.72	2.12
$\beta_{3,1}$	3.46	9.82
$\beta_{3,2}$	-0.71	1.97

0.64 ($CI_{95\%} = [0.61, 0.68]$). Nevertheless, one can also note that the confidence intervals for the aggregated estimation were larger. This illustrates the loss of precision when analyses are performed on summarized data. The Figure (5.2) illustrates the validity of the model with very small distances between fitted and observed values. One can also note the homogeneity of residuals according to the survival time.

5.4 The prognostic capacity of KI-67 for patients with breast cancer

KI-67 is a marker of the proliferative activity of breast cancer, but its prognostic capacity is still unclear. In their recent meta-analysis, de Azambuja et al. [26] concluded that KI-67 positivity conferred a worse survival. This work focused on the 35 evaluable studies of the relationship between KI-67 and the overall survival. All of these studies provided a hazard ratio, but only 23 described survival curves according to the level of KI-67. The aggregated data are exhaustively available at `www.divat.fr/en/softwares/sroct`. Survival probabilities were measured every year.

We used the same regression for the survival time (Equation 5.8). However, according to the higher number of studies and the higher follow-up, we chose 5 time intervals for modeling the hazard function with $\tau_4 = 8$ years. The estimated parameters are presented in the first columns of Table (5.2). The corresponding SROCt curves for a prognosis up to 3, 5 and 9 years are presented in Figure (5.3). The corresponding areas equaled 0.63 ($CI_{95\%} = [0.57, 0.71]$), 0.64 ($CI_{95\%} = [0.56, 0.74]$) and 0.65 ($CI_{95\%} = [0.52, 0.77]$). The left side of the confidence intervals were higher than the non-informative value of 0.50. Therefore, KI-67 appeared to be a significant prognostic marker. Nevertheless, the

prognostic capacity was limited with important error rates in term of medical decision making (areas far from the perfect value at 1).

The goodness-of-fit of the models was checked and the graphs are provided in the Figure (5.4). The studies by Trihia et al. [149] and Railo et al. [126] were associated with higher residuals in comparison to the other studies. We thus performed the same analysis without the corresponding observations. The updated parameters are presented in the last columns of Table (5.2) and the goodness-of-fit in the Figure (5.5). The updated areas respectively equaled 0.65 ($CI_{95\%} = [0.61, 0.69]$), 0.67 ($CI_{95\%} = [0.63, 0.73]$) and 0.68 ($CI_{95\%} = [0.64, 0.78]$). Because the extreme residuals were removed, the confidence intervals were smaller. Nevertheless, the values of the areas were closed to the ones obtained when all the studies are included. Therefore, the clinical conclusions seemed similar.

Table 5.2: *Estimation of the parameters and their standard errors for the analysis of KI-67 (standard errors of random effects : $\sigma_\omega = 0.54$ and $\sigma_{\upsilon_k} = 0.71$ for the analysis with all studies ; $\sigma_\omega = 0.44$ and $\sigma_{\upsilon_k} = 0.70$ for the robustness analysis).*

	Analysis with all the studies		Robustness analysis	
	Estimation	Standard Error	Estimation	Standard Error
μ	2.56	0.12	2.65	0.10
σ	-0.26	0.08	-0.21	0.02
$\beta_{0,1}$	-4.31	0.67	-5.72	0.69
$\beta_{0,2}$	0.51	0.22	0.75	0.21
$\beta_{1,1}$	0.02	1.39	1.57	0.96
$\beta_{1,2}$	0.10	0.47	-0.23	0.31
$\beta_{2,1}$	-0.57	2.51	-1.88	1.20
$\beta_{2,2}$	0.07	0.80	0.54	0.36
$\beta_{3,1}$	0.51	3.36	1.60	1.54
$\beta_{3,2}$	-0.24	1.11	-0.62	0.52
$\beta_{4,1}$	-1.66	3.02	-1.26	1.36
$\beta_{4,2}$	0.48	1.00	0.56	0.49

5.5 Discussion

We developed a method to assess time-dependent summary ROC curves from published results. This method has three clear advantages over classical meta-analyses based on HRs [111], survival curves [10] or standard summary ROC curves [102, 153]. First, even if the interpretation of the hazard ratio is in fact one of the key point of basic survival, it is difficult to interpret in terms of prognostic capacity. While HR focuses on group comparisons, the SROCt curve is representative of prognostic capacity across the

possible cut-offs of a marker. SROCt curves constitute an appropriate way to represent and quantify the prognostic capacity of a marker. Moreover, they enable the incorporation of more than 2 groups from the original studies ; all of the available information is thus used. Second, meta-analyses based on HRs or survival curves require a homogeneous definition of groups across the studies included. Beyond the heterogeneity caused by varying cut-offs, pooling such HRs or survival curves is not meaningful and should be avoided. The SROCt approach specifically takes into account the number of groups and their corresponding cut-offs. The proposed method is also adaptable to binary tests (the cut-off of the marker is fixed) or covariates may be added (in the marker or in the time distributions in order to avoid identifiability issue). Third, classical summary ROC curves are also limited to single time-points while the follow-up time may change across published studies. As such, summary ROC curves are more appropriate for diagnostic studies [153].

The proposed method is generic and applicable to a large panel of marker distributions and survival models. Of note, the survival model we describe is highly flexible in comparison to traditional approaches. This avoids a level of arbitrariness for the results. The goodness-of-fit analysis justifies our choice for both applications. As for every regression analysis, the results should not be interpreted if the goodness-of-fit plots indicate inadequacy between observed and fitted values. Because the number of works is significant for a meta-analysis, this information should be used to drive the modeling. For instance, Dowsett et al. [30] specified that *statistical analyses should take account of the log-normal distribution generally followed by KI-67 measurements*. Indeed, this distribution fully corresponds to the observed data, as illustrated by Figures (4A) and (5A). Although, the computation can be complex, the results are easily interpretable by researchers in clinical epidemiology, who are familiar with the ROC principle. Moreover, the practical use of the recommended procedures is enhanced since an `R` package is available.

However, there are limitations to the proposed method. First, as for any ROC curves, our approach is not appropriate if the relationship between the biomarker and the time-to-event is not monotonic. This problem can be detected from the published curves : the survival should decrease (or increase) with the level of the biomarker. The results of the KI-67 meta-analysis also illustrated that this problem may be detected by plotting the observed bivariate probabilities against the fitted values (Figure 4-B). Second, the variance of the area under the SROCt curve is difficult to compute. It needs simulations of the estimated parameters according to their variance. Another limitation is the measurement of heterogeneity, which is often measured by the I-squared method [68] when a single-value outcome is meta-analyzed. The problem is more complex when the outcome is a curve. Here, the heterogeneity is related to the variance of the random effect (σ^2_ω

and σ_ν^2). If the structures of the marker distribution and/or the survival models differ dramatically between studies, our method would not be appropriate (but no method would be). The last limitation is the calculation of the expected number of at-risk patients during the follow-up. The method assumes that the censoring rate is constant and non-informative. It makes necessary assumptions, because many papers only provide Kaplan-Meier curves without the number of at-risk patients over time, nor the confidence intervals of the curves. In this paper, we have adapted the approach proposed by Parmar et al. [111]. Nevertheless, other methods exist and can be used as alternatives [31, 155, 59].

The examples we presented illustrate the utility of our method. In the first example, the individual patient data from a multi-centric cohort were aggregated using randomly selected cut-offs. The aim was to compare the time-dependent ROC curves obtained from the individual data with those obtained from the aggregated data. After this validation, the aim of the second example was to show the feasibility of the method based on real aggregated data. The results were concordant with those from the published meta-analyses but were more informative. The meta-analysis demonstrated the worse overall survival for breast cancer patients who were positive for KI-67. But this is the first time that a meta-analysis has demonstrated the small capacity of KI-67 to predict overall survival.

Acknowledgement

We thank Roche Laboratories for its supporting the DIVAT data bank. This work is part of the French Transplantation Research Network (RTRS) supported by the "Fondation de Cooperation Scientifique CENTAURE" and by grant from the French National Agency of Research ANR-11-JSV1-0008-01 . We also thank Dr. J Ashton-Chess for editing the manuscript. We wish to thank members of the DIVAT centers included in this study : K. Launay, P. Daguin, M. Giral, M. Kessler, M. Ladrière, H. Kreis, C. Legendre, JP. Soulillou, L. Rostaing, N. Kamar, G. Mourad and V. Garrigue.

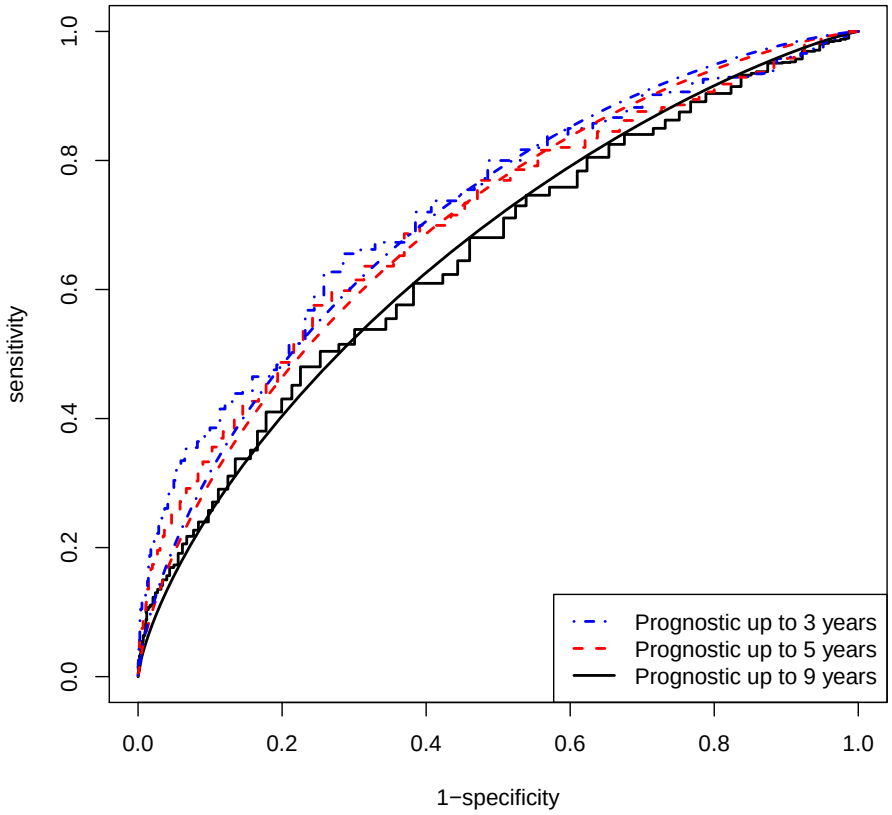

Figure 5.1: *SROCt Curves based on aggregated (continuous lines) and ROCt curves based on individual data (step lines) to evaluate the capacity of 1-year Cr to predict death or return to dialysis up to 3, 5 and 9 years after its measurement.*

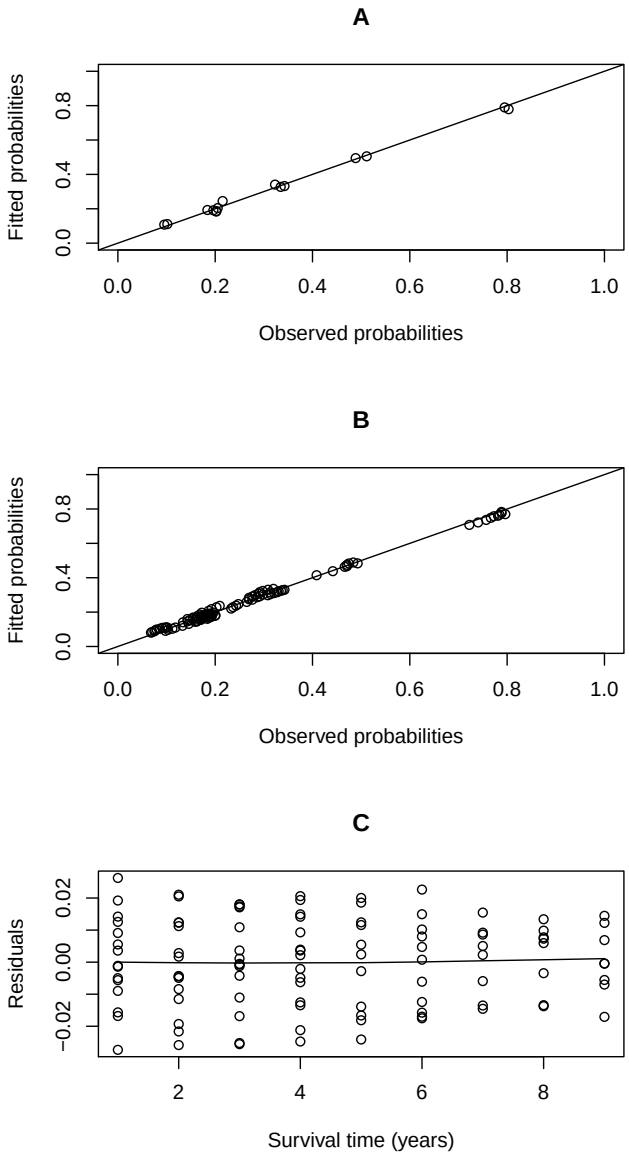

Figure 5.2: *Check of the goodness-of-fit for the creatinine study.* ***(A)*** *The observed proportions of patients in the gth interval of the study k versus the fitted proportions ($p_k(a,b)$, equation 3).* ***(B)*** *The observed bivariate probabilities versus the fitted bivariate probabilities ($S_k(t,a,b)$, equation 4).* ***(C)*** *The residuals of the bivariate probabilities ($S_k(t,a,b)$, equation 4) versus the times (t).*

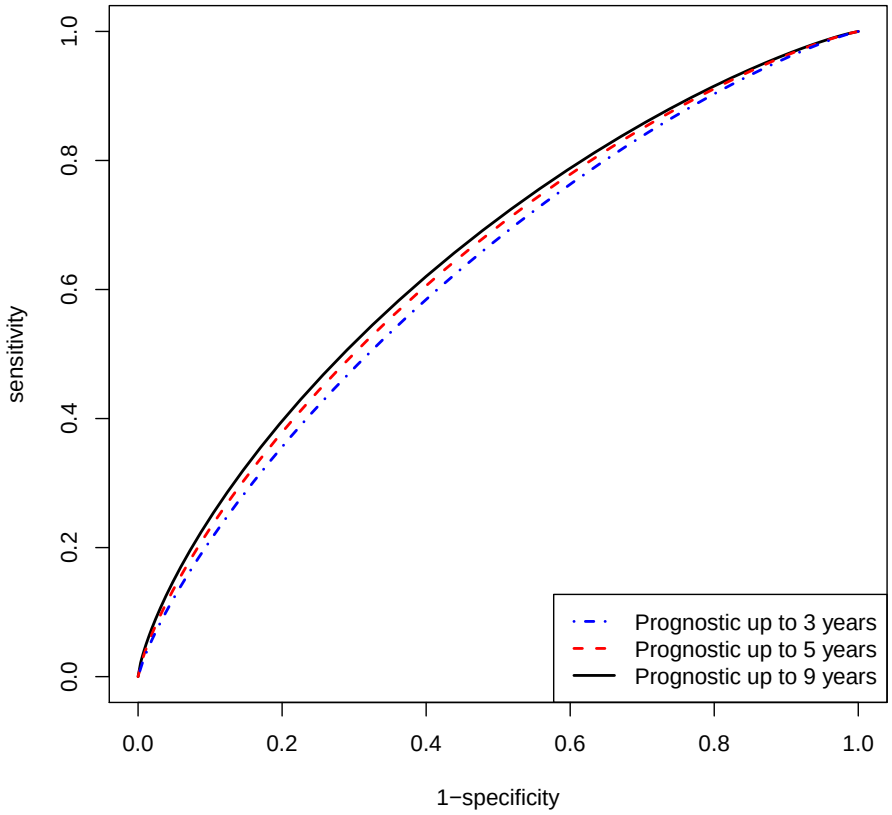

Figure 5.3: *SROCt Curves based on the 23 published studies to evaluate the capacity of KI-67 to predict the overall survival of patients with breast cancer up to 3, 5 and 9 years.*

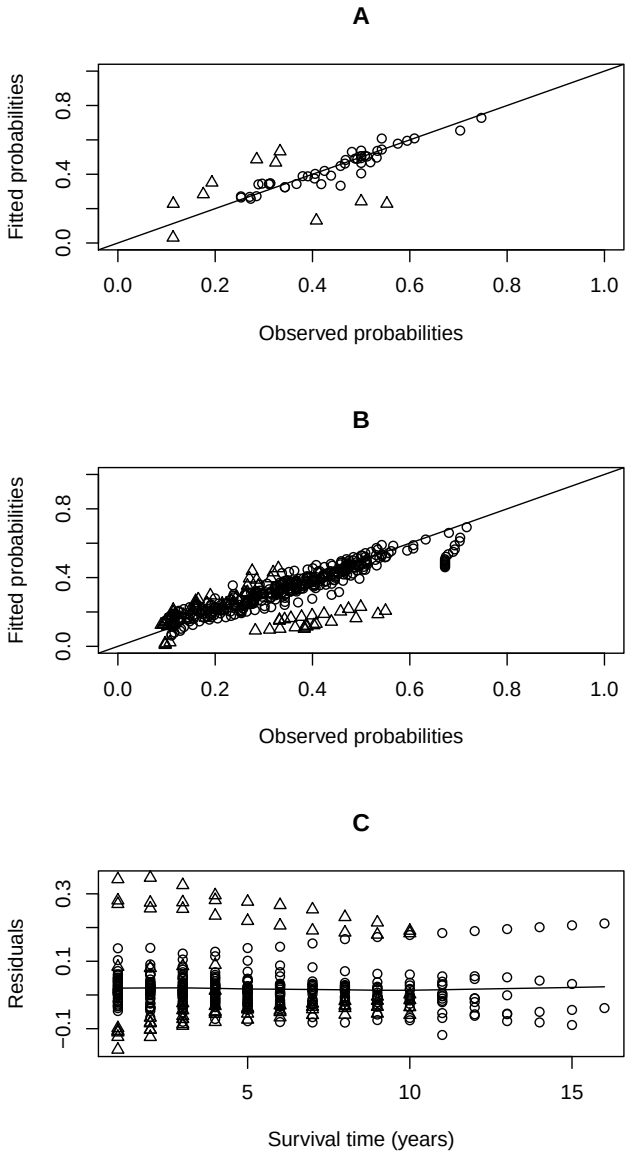

Figure 5.4: *Check of the goodness-of-fit for the KI-67 study. The points 'Δ' represent the residuals associated with the studies of Trihia et al. (2003) and Railo et al. (1993). The points 'o' represent the residuals associated with the other studies.* ***(A)*** *The observed proportions of patients in the gth interval of the study k versus the fitted proportions ($p_k(a,b)$, equation 3).* ***(B)*** *The observed bivariate probabilities versus the fitted bivariate probabilities ($S_k(t,a,b)$, equation 4).* ***(C)*** *The residuals of the bivariate probabilities ($S_k(t,a,b)$, equation 4) versus the times (t).*

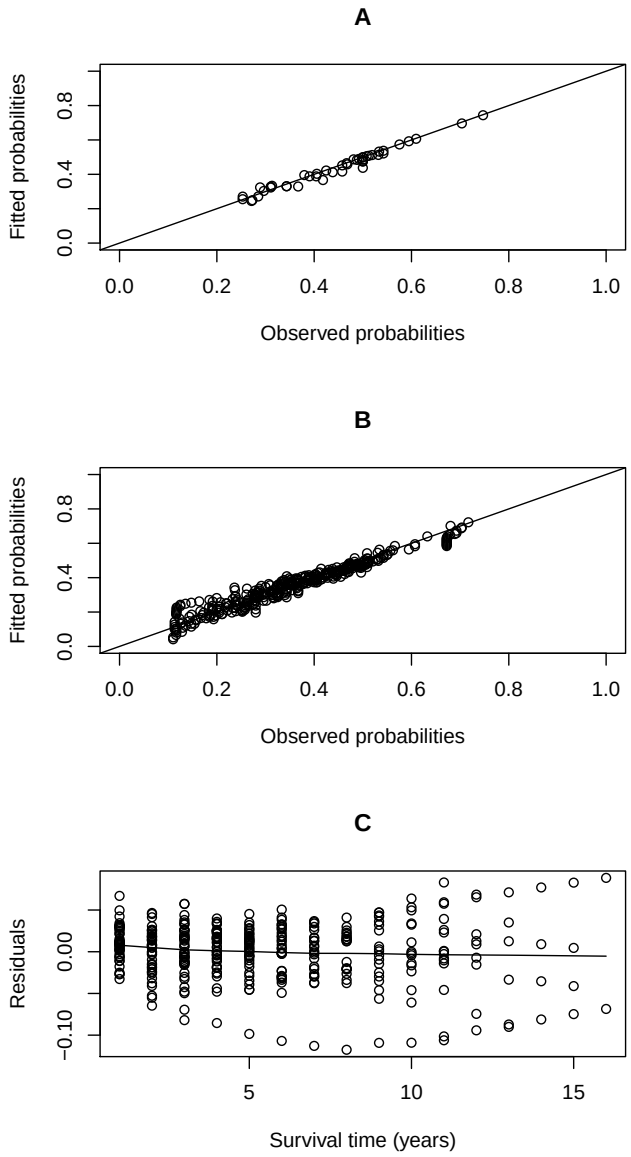

Figure 5.5: *Check of the goodness-of-fit for the KI-67 study without the 3 studies with extreme residuals (Trihia et al., 2003 and Railo et al., 1993).* ***(A)*** *The observed proportions of patients in the gth interval of the study k versus the fitted proportions ($p_k(a,b)$, equation 3).* ***(B)*** *The observed bivariate probabilities versus the fitted bivariate probabilities ($S_k(t,a,b)$, equation 4).* ***(C)*** *The residuals of the bivariate probabilities ($S_k(t,a,b)$, equation 4) versus the times (t).*

- Chapitre 6 -

Cut-off estimation and medical decision making based on a continuous prognostic factor : the prediction of kidney graft failure.

Foucher Y., Giral M., Soulillou J.P. and Daurès J.P. International Journal of Biostatistics. 2012 Jan 6 ;8(1).

6.1 Introduction

Cut-off determination is a frequent issue for statisticians. In clinical or epidemiological studies, the effect of continuous risk factors are often analyzed respecting a dose-effect relationship [134]. For example in kidney transplantation, numerous studies demonstrated that creatinine clearance is highly correlated with long-term graft survival [104, 54, 61]. However, it is difficult for clinicians to apply these results in practice as no cut-off is estimated for a decision. Moreover, from a statistical point of view, it may be important to categorize continuous covariates when a dose-effect assumption does not hold.

In order to determine such cut-offs in survival analyses with censored follow-up, a widely used method is to define a grid and to retain the cut-off associated with the highest difference between survival curves. However, with such a procedure, investigators are confronted with the problem of multiple testing. Therefore, authors such as Le Blanc and Crowley have considered tree methods [90]. These procedures split data by maximizing the difference in survival between groups, which is commonly measured by the log-rank test. Contal and O'Quigley have also proposed the maximization of a statistic, which bears the advantage of avoiding cut-offs located near the extremes [22].

The most recent publication on this topic is based on the generalized maximally selected statistics [69]. The latter authors have proposed an algorithm for a unified treatment of different kinds of maximally selected statistics enabling a large number of cut-offs. The procedure leads to a maximally selected chi-square test, as published by Miller and Siegmund [99]. The common characteristic of these methods is their independence to their application. However, risk assessment in medical practice has to take into account the consequences of the clinician decision.

In this paper, we propose an alternative method for cut-off estimation based on a decision-making framework. Depending on the application, this method reflects the impact of prognostic errors (monetary cost, medical gravity, social consequences, etc.) and takes into account the desired time of the prognostic.

In Section 2, we describe the adaptation of time-dependent ROC curves, initially defined by Heagerty et al. [63], for the estimation of cut-offs. We define in detail the ROC analysis because the evaluation of the prognostic capacity of a marker is of prime importance before the determination of a particular cut-off. Two non-parametric methods are proposed, based on the Kaplan-Meier estimator and on the Akritas nearest neighborhood estimator. In Section 3, we propose comparing the recent method of Hothorn and Zeileis [69] with the new approaches by simulations. Section 4 applies the methods to the analysis of kidney transplant recipients. The clinical objectives are the evaluation of the prognostic capacity of the 1-year creatinine clearance (CrCl) and the definition of the optimal cut-off to discriminate two groups according to their risk of failure. Finally, Section 5 discusses the new method and its benefits and limitations.

6.2 Methods

6.2.1 Framework

Using the counting process notation, let $D(t) = 1$ if the failure occurred before time t (i.e., $T \leq t$) and $D(t) = 0$ otherwise (i.e., $T > t$). All patients are free of failure at the beginning of the study ($D(0) = 0$). We consider X, measured at $t = 0$, as a prognostic marker of the failure time T. By convention, suppose that high values of X are associated with a high risk of failure. The prognostic test is defined as positive (patient at risk of failure), if the prognostic marker is higher than a cut-off c. The methodology associated with this type of prognostic analysis has been recently developed by Heagerty et al. [63, 64]. The sensitivity is thus the probability of a positive prognostic test among patients with failure before time t, i.e. $P(X > c|D(t) = 1)$. The specificity is the probability of a negative prognostic test among patients free of failure before time

t, i.e. $P(X \leq c|D(t) = 0)$. The ROC (Receiver Operating Characteristic) curve of a prognostic at time t represents the sensitivity in function of one minus the specificity for the different cut-offs c. The ROC curve is monotone non-decreasing for each t. The accuracy of the marker to predict the failure is measured by the area under the curve (AUC), independently of the cut-off. In order to find the optimal cut-off, we define a cost function, $C(c|t)$, which represents the total cost associated with the prognostic test based on the cut-off c at time t. One can distinguish the number of false positives (patients with a positive test but free of failure at time t) and the number of false negatives (patients with a negative test but with a failure before time t), respectively $n_{(fp)}(c|t)$ and $n_{(fn)}(c|t)$. As a result, the cost function represents the sum of these errors weighted by their respective costs, $C_{(fp)}$ and $C_{(fn)}$. Thus, we have :

$$C(c|t) \propto k \times n_{(fp)}(c|t) + n_{(fn)}(c|t), \tag{6.1}$$

where $\propto$ means "proportional to" and where $k = C_{(fp)}/C_{(fn)}$ is the relative importance of a false positive according to a false positive. Depending on the application, this function $C(c|t)$ can be determined such that it reflects the monetary impact, the medical gravity, the social impact, etc. The optimal cut-off is calculated to minimize this cost function for a given time of prognosis t.

6.2.2 Estimation using the Kaplan-Meier estimator

In the framework above, the sensitivity of the prognostic test with cut-off c at time t represents the probability of having a positive test $\{X > c\}$, given that a failure occurs before time t. Based on the work of Heagerty et al. [63], the non-parametric estimation of this probability, $se_{KM}(c|t)$, is :

$$se_{KM}(c|t) = \{1 - S_{KM}(t|X > c)\}\{1 - \bar{G}_X(c)\}/\{1 - S_{KM}(t)\}, \tag{6.2}$$

where $S_{KM}(t|X > c)$ is the Kaplan-Meier estimator of the survival probability at time t conditional on $\{X > c\}$ and $\bar{G}_X(c)$ is the empirical distribution function. Respectively, the estimator of the specificity of a prognostic test with cut-off c at time t, $sp_{KM}(c|t)$, is :

$$sp_{KM}(c|t) = S_{KM}(t|X \leq c)\bar{G}_X(c)/S_{KM}(t) \tag{6.3}$$

The ROC curve of a prognostic at time t, is the sensitivity $se_{KM}(c|t)$ plotted in function of one minus the specificity $sp_{KM}(c|t)$ for all the possible cut-off c. In order to find the

optimal cut-off, $\tilde{c}$, the cost function (6.1) can be developed as follow :

$$\begin{aligned} C(c|t) &\propto kP(X > c, D(t) = 0) + P(X \leq c, D(t) = 1) \\ &\propto kP(T > t|X > c)P(X > c) + P(T \leq t|X \leq c)P(X \leq c), \end{aligned}$$

and can be non-parametrically estimated :

$$C_{KM}(c|t) \propto kS_{KM}(t|X > c)\{1 - \bar{G}_X(c)\} + \{1 - S_{KM}(t|X \leq c)\}\bar{G}_X(c) \tag{6.4}$$

6.2.3 Estimation using the Akritas estimator

Heagerty et al. [63] also proposed to estimate $P(X > c, D(t) = 0)$ using the nearest neighborhood estimator initially proposed by Akritas [4]. Respecting the same notations, let $S_{\lambda_n}(c, t)$ denote the estimation of this bivariate survival probability, where $2\lambda_n \in (0, 1)$ represents the percentage of observations that is included in each neighborhood. This estimator ensures a monotone ROC curve in contrast to the Kaplan-Meier approach. Moreover, the Kaplan-Meier estimator will be biased if censoring is dependent on marker, whereas the Akritas estimator will be robust in this situation. The sensitivity and specificity become :

$$se_{\lambda_n}(c|t) = \{(1 - \bar{G}_X(c)) - S_{\lambda_n}(c, t)\}/\{1 - S_{\lambda_n}(-\infty, t)\} \tag{6.5}$$

$$sp_{\lambda_n}(c|t) = 1 - S_{\lambda_n}(c, t)/S_{\lambda_n}(-\infty, t) \tag{6.6}$$

The ROC curve of a prognostic at time t is similarly obtained using the probabilities (6.5) and (6.6) instead of (6.2) and (6.3). The cost function can be estimated by :

$$C_{\lambda_n}(c|t) \propto (k + 1)S_{\lambda_n}(c, t) + \bar{G}_X(c) \tag{6.7}$$

6.2.4 Computation details

We chose the trapezoid method to calculate the area under the curve. In order to calculate the 95% confidence intervals ($CI_{95\%}$) of the optimal cut-offs and of the areas under the curves, 1999 bootstrap replications were performed and percentile intervals were calculated [35].

6.3 Results from simulations

Both approaches were compared to the Hothorn and Zeiles method [69], which identifies a global cutoff over all failure time and is based on the maximization of the difference between the survival curves (generalized maximally selected statistics). The false positive and negative errors are not taken into account. The Hothorn and Zeiles method is thus completely independent of the application and the decision consequences. Therefore, the present comparisons are only relevant if no assumption is made about the prognostic time and the consequences. We arbitrarily chose $k = 1$ (no preference regarding the minimization of the false positive or the false negative errors). We also arbitrarily chose a prognostic time equal to half the maximum observed follow-up time.

We simulated artificial samples for different sample sizes (N=25, 50, 100 and 200). Let us suppose a variable Z simulated assuming a standard normal. This variable was used in the 4 following scenarios in order to define the value of X :

1. The times-to-event are simulated according to a proportional hazard model with a Weibull distribution (scale and shape parameters respectively equal to 1.5 and 0.5). z is the observation of a random variable Z which is a transformation of X. We assume that the regression coefficient β equals 0.4 (relative risk equals 1.5). A single cut-off is fixed at 0.5 with $Z = 1$ if $X > 0.5$ and with $Z = 0$ otherwise. The times of censoring are uniformly distributed between 0 and 15.
2. The times of events are simulated similarly. However, we fix a cut-off in 0.5 with $X = Z$ if $Z > 0.5$ and with $X = 0$ otherwise. In contrast to the scenario #1, the hazard function does not jump at 0.5. The function is constant before 0.5 and proportionately increases with the values above 0.5.
3. If $Z \leq 0.5$ the times of events are simulated according to a Weibull law (scale and shape parameters respectively equal to 1.5 and 0.5), but if $Z > 0.5$ the Exponential distribution is used (scale and shape parameters respectively equal to 1.5 and 1.0). Thus, a single cut-off exists at 0.5 but the PH assumption does not hold.
4. The scenario is equivalent to (1) but with $X = Z$. There is no cut-off.

1000 simulations were performed per scenario and per sample size. The results are presented in Table (6.1). Globally, the results were very similar regardless of the method. Below a sample size of 50 individuals, no method was reliable, but the proposed methodology (Kaplan-Meier or Akritas) offered estimations closer to the true cut-off in comparison to the Hothorn and Zeileis method. Regardless of the scenario, the variability of the estimations using cost function seemed to be lower than those obtained by the Hothorn and Zeileis method. If no cut-off exists (Scenario #4), the proposed method estimated cuts-offs close to 0, which separated the sample into two balanced groups. In

Size	H. & Z.	K. & M.	Akritas
		Scenario #1	
N = 25	-0.06 (1.07)	0.10 (0.97)	0.11 (0.98)
N = 50	0.13 (1.09)	0.28 (0.83)	0.27 (0.84)
N = 100	0.30 (0.82)	0.36 (0.73)	0.34 (0.69)
N = 200	0.41 (0.51)	0.38 (0.46)	0.38 (0.43)
		Scenario #2	
N = 25	0.02 (1.20)	0.21 (0.98)	0.18 (1.08)
N = 50	0.25 (1.08)	0.36 (0.90)	0.33 (0.89)
N = 100	0.49 (0.81)	0.41 (0.73)	0.41 (0.69)
N = 200	0.57 (0.50)	0.49 (0.50)	0.49 (0.50)
		Scenario #3	
N = 25	0.24 (1.05)	0.16 (0.85)	0.17 (0.80)
N = 50	0.36 (1.01)	0.32 (0.63)	0.32 (0.60)
N = 100	0.44 (0.77)	0.40 (0.44)	0.40 (0.40)
N = 200	0.49 (0.37)	0.43 (0.26)	0.43 (0.24)
		Scenario #4	
N = 25	-0.28 (1.09)	0.05 (0.83)	-0.04 (0.81)
N = 50	-0.35 (1.06)	0.00 (0.76)	0.01 (0.74)
N = 100	-0.32 (0.91)	-0.01 (0.70)	-0.03 (0.67)
N = 200	-0.31 (0.80)	0.01 (0.61)	0.02 (0.60)

Table 6.1: *Estimations of the cut-off values according to the scenarios and the sample sizes. Based on 1000 simulations, the medians (and the corresponding inter-quartile intervals) are reported.*

this scenario, the Hothorn and Zeileis method offered lower estimations.

6.4 Kidney transplant survival data

6.4.1 Data description

The data were extracted from the DIVAT data bank from Nantes Hospital (France), which is a prospective cohort of kidney transplant recipients. Biological and clinical data have been recorded since 1990. Specialized clinical research assistants who were independent to the medical team, computerized the pre- and post-transplant parameters of each patient transplanted in the center. Recorded data are submitted to an annual medical cross-audit with a level of error below 1%.

In this paper, we consider a subpopulation of 839 patients more than 18 years of age and who received a kidney transplant between January 1996 and September 2006. Death or a return to dialysis mean that the graft failed. The value of the creatinine clearance

(CrCl) one year after the transplant is the marker of interest to predict long-term graft survival. A low CrCl value is associated with a higher risk of graft failure. Note that there is no reason to justify that a cut-off exists with a discontinuity of the risk at this value. However, clinicians have to take decisions to classify the patients according to their risk of failure. Usually, clinicians consider a CrCl above 40 ml/min as an indication of a poor prognosis. A cut-off of 40 ml/min is obtained as half of the lower CrCl limit for a healthy person (80 ml/min). When one kidney is transplanted to a patient, one expects a CrCl of half the normal values. To our knowledge, no quantitative study has been performed to justify this threshold according to the risk of failure. Based on the Kaplan-Meier analysis of graft survival, Hariharan et al. have shown that a creatinine level of more than 1.5 mg/dL is associated with a poor graft outcome, compared to levels below that value [61]. The focus of interest is whether this cut-off is optimal for discriminating two groups of patients according to their risk of graft failure.

Our objective was to determine the optimal cut-off of the CrCl value collected 1 year after the transplantation for predicting graft survival. The origin of the study ($t = 0$) is thus at 1 year after the transplantation and concerns only patients with a functional kidney at 1 year. Return to dialysis, death and censored patient within the first year of transplantation were not included in the analysis. The survival time of interest is thus the time between the first anniversary of the transplantation and the graft failure or the death of the patient. In the following developments, we will principally choose a prognostic time of up to 4 or 8 years after the CrCl measurement ($t = 4, 8$). Different formulae are available to calculate the CrCl. We used the Modification of Diet in Renal Disease version [91]. The mentioned data are provided for reanalysis and verification (http ://www.divat.fr/).

6.4.1.1 Study of the prognostic accuracy

Figure 6.1 shows the ROC curves based on the Kaplan-Meier or on the Akritas estimators. It illustrates the ability of the CrCl to predict a failure up to 4 and up to 8 years after the first anniversary of the transplantation. In agreement with Heagerty et al. [63], both estimators gave similar results. Using the Akritas version, the AUC at 4 years was 0.79 ($CI_{95\%} = [0.70, 0.85]$) versus 0.73 ($CI_{95\%} = [0.69, 0.87]$) at 8 years. Using the Kaplan-Meier estimator, the AUC at 4 years was 0.80 ($CI_{95\%} = [0.69, 0.87]$) versus 0.70 ($CI_{95\%} = [0.50, 0.82]$) at 8 years. Regardless of the estimators and as expected, the CrCl was a better prognostic variable for a short-term prognostic.

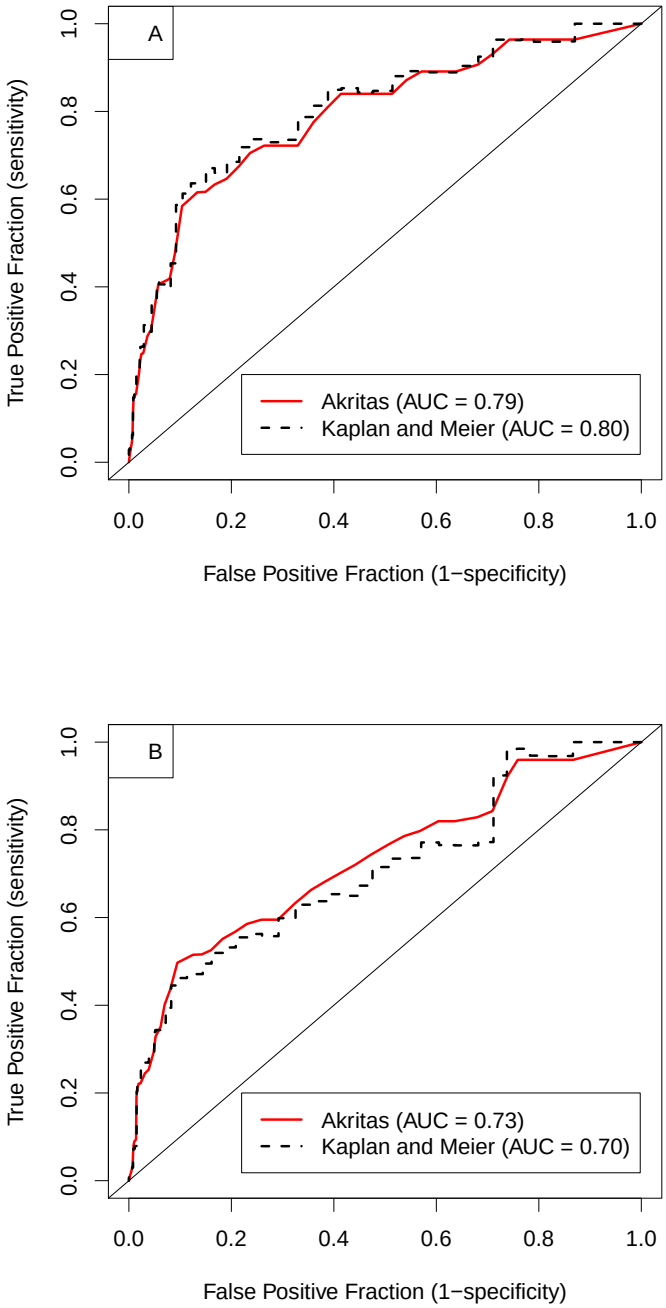

Figure 6.1: *The ROC curves estimated for evaluating the capacity of the CrCl to predict a graft failure or the death of the transplanted recipient up to 4 years (A) and 8 years (B)*

6.4.1.2 Determination of an optimal cut-off

If the objective of the cut-off is to discriminate all the events with a minimum of errors, then k should be equal to 1 (the cost of a false negative is equivalent to the cost of a false positive). However, in such a case, the cut-off can appear to be disconnected from the real medical issue. In our application, the medical consequences of a false negative are much more serious than the consequences of a false positive. The relative weight k is assumed at 1/9. We used the approach of Vickers and Elkin to determine this ratio [152]. Consider that the follow-up of the at-risk patients is more frequent (every 3 months) in comparison with the follow-up of the risk-free patients (every year). If the probability of failure is close to 1, all clinicians will decide on an intensive follow-up. If the probability of failure is close to 0, all clinicians will decide on a less intensive follow-up. After discussions, clinicians defined the disease probability for which the decision is unsure at 10%. In other words, if the probability of a graft failure is below 10%, they accept to increase the length of the intervals between two visits. Above 10%, they accept to decrease this length. Their decision appears unsure at about 10% and Vickers and Elkin demonstrated that the relative harms of a false positive and a false negative is thus equal to 10/(100-10) [152].

Figure 6.2 represents the optimal cut-offs for $k = 1/9$, for both estimators and for all the prognostic times. Consider a prognostic up to 4 years after the CrCl measurement, which corresponds to the time origin, 1 year after the transplantation. The optimal cut-off equals 31.4 ml/min ($CI_{95\%} = [26.7, 33.9]$, Akritas estimator). For a prognostic up to 8 years after the CrCl measurement, the optimal cut-off also equals 31.4 ml/min ($CI_{95\%} = [27.7, 34.1]$, Akritas estimator). One can see that these estimations are different from the usual cut-off of 40 ml/min. However, the cut-offs do not vary according to the prognostic time. The results based on the Kaplan-Meier estimator are very similar.

We did not compare our results with the Hothorn and Zeileis method, because we chose the specific value of k and the specific time of prognosis. The comparison would not be relevant.

6.5 Discussion

In medicine, it is useful to determine a decision cut-off of prognostic markers. We have proposed a method in order to define such thresholds from time-dependent data where the failure may be censored.

We first proposed to validate the methodology using simulated data and comparing the results with the recent method of Hothorn and Zeileis [69]. This approach is today

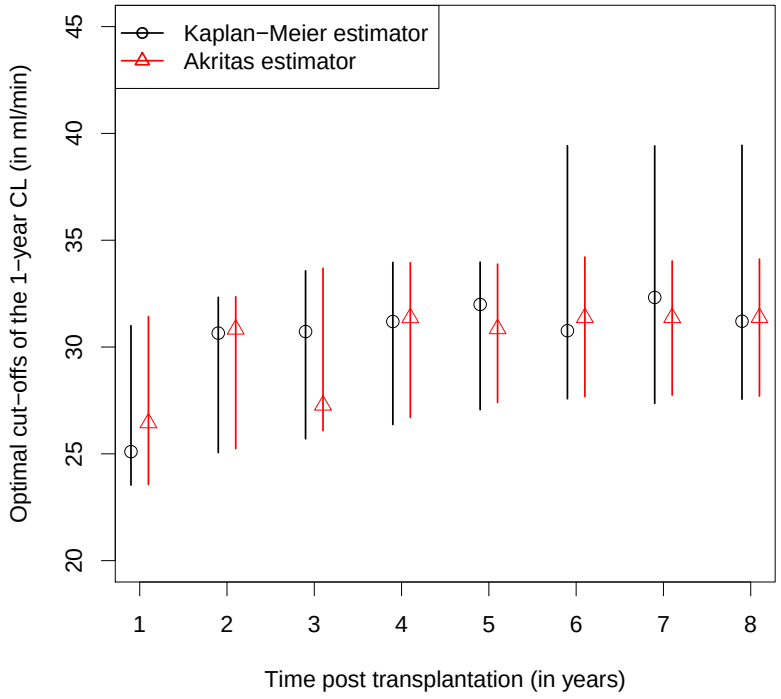

Figure 6.2: *Optimal cut-offs depending on both estimators (Kaplan-Meier and Akritas) for* $k = 1/9$*. The cut-off estimations are associated with the corresponding 95% confidence interval.*

a reference in traditional statistical analysis : the principal objective is the study of the correlation between risk factor and time-to-event and the estimation of cut-off associated with the lower p-value (null hypothesis that the hazard ratio equals 1). With the present approach, we distinguish the risk factor analysis from the prognostic analysis where the cut-off estimation needs to be adapted to the medical question. Therefore, for a relevant comparison of the methods, no assumption was made about the consequences of the prognostic decision. The objective of these simulations was to validate our methodology in this context. Regarding the simulation results, this objective was achieved. However, this does not mean that the proposed methodology is better. As explained, the methodologies should not be applied in the same context.

In accordance with Gail and Pfeiffer [52], we believe that existing methods are irrespective of the intended application. Our proposed approach is more adequate in the context of medical decision making. Two possible modulations can be considered in order to obtain a cut-off close to the expectations of clinicians. Firstly, we can calculate the predictive accuracy and the cut-off value according to the required time of the prognosis. This distinction is of prime importance since a marker could be informative for early events, but not so useful for a long-term prediction. Our application illustrated this statement. Moreover, the fixing of the prognostic time is very important since the gravity of error may be different if it appears just after the decision or if it appears a long time after. From a statistical point of view, dichotomization of the survival time of the proposed method may lead to a loss of information. However, regarding the practice of the medical decision, determining the prognostic time is important.

Secondly, the concrete consequences of decision errors based on the cut-off can be taken into account. For example in transplantation, clinicians prefer to wrongly predict a future failure as opposed to wrongly predict the survival of the graft. This second error is only associated with a more intensive but useless follow-up. The priority is thus to minimize the number of false negatives ($k < 1$). The weights of both errors should be defined according to experts. The methodology proposed by Vickers and Elkin can be used to help this definition [152]. Even if no idea about the weights results from the discussion, the non-informative choice is always possible ($k = 1$).

The simulations demonstrated the good capacity of the methods to estimate the existing cut-off. One can ask the relevance of these estimations when no cut-off exist (scenario #5). However, in the medical practice, decisions have to be made even if no cut-off exists. This is the case in our application : the risk of graft failure decreases continuously with an increase in CrCl. Nevertheless, clinicians have to make decisions daily based on this marker.

A limitation of the proposed methodology is that no adjustment is possible to take

into account confounding factors. The cut-off value can vary according to other determinants. However, based on the proposed nonparametric methodology, a solution is to perform a stratified analysis. To avoid the traditional limitations associated with stratified analysis, it may be interesting to develop a multivariate approach, semiparametric method can be considered.

Finally, it may also be of interest to generalize the method to multiple variables and to consider the marker as time-dependent. We are in the process of working on this type of extension, in particular based the recent paper by Zheng and Heagerty [161].

Acknowledgement

We thank the Roche Laboratory for its support of the DIVAT data bank and Dr J. Ashton-Chess for editing the manuscript. This work is a part of the French Transplantation Research Network (RTRS) supported by the CENTAURE foundation.

- Chapitre 7 -

Time-dependent ROC analysis for a three-class prognostic with application to kidney transplantation

Foucher Y, Giral M, Soulillou JP and Daurès JP.
Statistics in Medicine. 2010 Dec 30 ;29(30) :3079-87.

7.1 Introduction

ROC graphs have long been used in signal detection theory and are now popular in the study of continuous markers for a binary diagnostic [162]. Given a decision threshold, the sensitivity and specificity are the basic measurements of the accuracy of the corresponding diagnostic test. As the threshold shifts, so do the sensitivity and specificity. The receiver operating characteristic (ROC) curve is a plot of the sensitivity versus one minus the specificity, for all possible cut-off points.

Even through this method is very useful for many applications, two significant limitations can be observed. Firstly, the disease status is considered as a fixed outcome. This approach cannot thus be used for a long-term prognostic with possible censored times. Heagerty, Lumley and Pepe proposed recently an adaptation, known as a time-dependent ROC curve [63]. Based on the estimators of Kaplan-Meier or Akritas, this method takes into account the censoring of the outcome. Since the writing of the first paper, Zheng and Heagerty have also proposed certain considerable extensions, particularly with the analysis of longitudinal markers [160, 161].

Secondly, only one binary disease can be studied using classic methodology. However, in numerous medical applications, the outcome of the diagnosis is more complex. For example in cancerology, the remission of the disease may be stable or may progress

either until relapse or death. Mossman has described a method for expanding the scope of ROC analysis and for describing diagnostic accuracy when there are three possible outcomes [103]. We can also cite the recent works of Heckerling [65] or He et al. [62], but little attention has been given to extending for time-dependent outcomes. The fact that the problem of competing risks is a complex issue in survival analysis. There have been extensive discussions in the literature on the statistical analysis of competing risk data since the first work of Prentice et al. [123] until today with the paper by Li, Elashoff and Li [93] for example. But only Pepe et al. [112] proposed a brief review paper with one part on competing risk and time dependent ROC analysis. Very recently, Saha and Heagerty [132] have also developed time-dependent ROC curves in the presence of competing risks.

In this article, we propose an extension of the ROC concept to the prognostic of three alternative evolutions. The prognostic marker is measured at the beginning of the follow-up for a long-term prediction, which is possibly censored. We use an extension of the competing risk models with the semi-markovian approach. The interest of this method is the division of the process modeling into two parts : the time of occurrence of the failure and the type of the failure [113, 49]. As we demonstrate in the following developments, this division is very useful to compute the sensitivity and the specificity associated with each type of failure. We can define two alternative ROC curves with cause-specific or marginal estimations.

These developments are motivated by the analysis of kidney transplant recipients. The objective of this application is to evaluate the accuracy of the 1-year creatinine clearance (CrCl) for the prediction of a return to dialysis or the death of a patient. This study has already been conducted by Kaplan, Schold and Meier-Kriesche [80], who concluded that CrCl does not have a sufficient predictive value to serve as a reliable predictive test for graft loss or patient death. However, these authors used the basic ROC method for the binary diagnostic test without censoring, which is unsuitable for this application where three prognostic alternatives are time-dependent.

7.2 Methods

7.2.1 The semi-markovian regression

Let T the time of the first event between two competitive failures X ($X = \{1, 2\}$) and Y the marker value at the origin of the follow-up ($Y \in \Re$). Let P_i denote the probability that the first failure is i ($i = 1, 2$). Since we can observe two possible failures, we have $P_1 = 1 - P_2 \in [0, 1]$. The logistic function can be used to insure this property :

$P_1 = exp(\alpha)/\{1+exp(\alpha)\}$ and $P_2 = 1/\{1+exp(\alpha)\}$, $\forall \alpha \in \Re$. In practice, the probability P_i $(i = 1, 2)$ represents the proportion of patients who fail from cause i when the time tends to ∞. If a large part of the sample is right-censored, these probabilities P_i should not be interpreted.

These probabilities do not deal with failure times. Thus, according to the semi-markovian property [113], we define the survival function specific to the transition into the failure i, $S_i(t) = P(T > t|X = i)$. Based on these definitions and the Bayes' theorem, we can deduce the well-known net survival of competing risk models : $P(T > t, X = i) = P_i\, S_i(t)$. Moreover, in order to accommodate the right-censoring, the marginal survival is deduced from the law of total probability : $P(T > t) = \sum_i P_i\, S_i(t)$.

Under the proportional hazard assumption, $S_i(t|Y = y) = S_{0i}(t)^{exp(\beta_i y)}$, where $S_{0i}(t)$ is the baseline survival function specific to the failure i and β_i is the regression parameter associated with the marker Y. In the following developments, we will note $Z_i = \beta_i Y$, the score associated with the failure i. Notice that the risk of failure i increases with Z_i. The main advantage of this regression is to allow the inclusion of a vector of variables into Y and thus to obtain a composite score. Let t_j denote the time of the last observation of the subject j $(j = 1, ..., N)$ and $z_{ij} = \beta_i y_j$ denotes the subject's score associated with the failure i $(i = 1, 2)$. Thus, the log likelihood of such a sample is defined by :

$$\begin{aligned} log\mathcal{V} &= \sum_{j=1}^{N} \left\{ \sum_{i=1}^{2} \delta_{ij} \Big\{ log(P_i) + log(\lambda_{0i}(t_j)) + z_{ij} - exp(z_{ij})\Lambda_{0i}(t_j) \Big\} \right. \\ &+ \left. \Big(1 - \sum_{i=1}^{2} \delta_{ij}\Big) log\Big(\sum_{i=1}^{2} P_i S_{0i}(t_j)^{exp(z_{ij})} \Big) \right\} \end{aligned} \tag{7.1}$$

where δ_{ij} equals 1 if the end of the follow-up consists of failure i for the subject j and 0 for a right-censoring. Based on the second part of this function, the methodology accommodates censoring. λ_{0i} and Λ_{0i} are the baseline hazard and cumulative hazard functions derived from S_{0i}. Log-Minus-Log survival plots are used to determine if baseline hazard functions are proportional. For more details and explanations on this approach, you can see the work of Foucher et al. [49].

7.2.2 Evaluations of the prognostic performances

7.2.2.1 Sensitivity and specificity

The cause-specific definition - Based on the previous notations, we suppose that $\{Z_i > c_{i\tau}\}$ corresponds to the subgroup at risk of failure i before the time τ. $c_{i\tau}$ represents the threshold of this prognostic test. The sensitivity (se) and the specificity

(sp) of the score Z_i for the prediction of the failure i, i.e., $P(Z_i > c_{i\tau}|T \leq \tau, X = i)$ and $P(Z_i \leq c_{i\tau}|T > \tau, X = i)$ are equal to :

$$se_i(c_{i\tau}|\tau) = \int_{c_{i\tau}}^{\infty}(1 - S_i(\tau|z_i))g_i(z_i)dz_i \Big/ \int_{-\infty}^{\infty}(1 - S_i(\tau|z_i))g_i(z_i)dz_i \quad (7.2)$$

$$sp_i(c_{i\tau}|\tau) = \int_{-\infty}^{c_{i\tau}} S_i(\tau|z_i)g_i(z_i)dz_i \Big/ \int_{-\infty}^{\infty} S_i(\tau|z_i)g_i(z_i)dz_i \quad (7.3)$$

where $g_i(z_i)$ is the probability density function of the score z_i. In practice, the sensitivity represents the probability of being at risk given that a failure i occurs before time τ. Respectively, the specificity represents the probability of not being at risk given that a failure i does not occur before time τ. The choice of τ needs to be made with caution, within the observation period in which the number of patients remains sufficient, in order to avoid any prediction or expected result.

The exact time of failure may be censored before time τ, therefore one cannot estimate these quantities by simple proportions (i.e. se would be the proportion of positive tests among patients with a failure i before time τ). These equations are based on the previous semi-Markov model (demonstrations are detailed in Appendix A) that deals with censored trajectories. Saha and Heagerty [132] recently proposed an alternative definition of specificity where the control group is $\{T > \tau\}$, i.e. patients without any failure before τ. Based on this definition, the aim is to evaluate the capacity of the marker to discriminate the patients with failure i before τ from those with no failure up to τ. The aim of the present cause-specific approach is rather than the discrimination from the patients without the failure i up to τ.

The corresponding cause-specific ROC curve for a prognostic at time τ, called $ROC_i(\tau)$, plots the se (7.2) in relation to one minus the sp (7.3) for the different cutpoints $c_{i\tau}$. The accuracy of the marker to predict the presence/absence of the failure i is measured by the area under the $ROC_i(\tau)$ curve, called $AUC_i(\tau)$. $AUC_i(\tau)$ is calculated using the trapezoid method. The confidence intervals of $AUC_i(\tau)$ are obtained from 1000 bootstrap replications. The $2.5th$ and $97.5th$ percentiles of the empirical distribution represent the limits for the 95% bootstrap percentile confidence interval ($CI_{95\%}$).

The marginal definition - In order to evaluate the performances of both scores to globally predict a failure, we also define the marginal se and sp. The first is the probability that at least one of the two scores is higher than its respective cutpoint, given that one of the two failures occurs before the prognostic time. The second is the probability that both scores are less than or equal to their respective threshold, given that no failure occurs before the prognostic time. These marginal definitions allow to evaluate if the marker Y is capable to predict at least one of the failures. We demonstrate

(Appendix A), that these quantities can be respectively estimated by :

$$se(c_\tau|\tau) = 1 - \left\{ \sum_{i=1}^{2} P_i \int_{-\infty}^{\omega_i} (1 - S_i(\tau|z_i)) g_i(z_i) dz_i \right\} \times \left\{ \sum_{i=1}^{2} P_i \int_{-\infty}^{\infty} (1 - S_i(\tau|z_i)) g_i(z_i) dz_i \right\}^{-1} \quad (7.4)$$

$$sp(c_\tau|\tau) = \left\{ \sum_{i=1}^{2} P_i \int_{-\infty}^{\omega_i} S_i(\tau|z_i) g_i(z_i) dz_i \right\} \Big/ \left\{ \sum_{i=1}^{2} P_i \int_{-\infty}^{\infty} S_i(\tau|z_i) g_i(z_i) dz_i \right\} \quad (7.5)$$

where $c_\tau = (c_{1\tau}, c_{2\tau})$, $\omega_1 = min(c_{1\tau}, \gamma^{-1} c_{2\tau})$ and $\omega_2 = min(\gamma c_{1\tau}, c_{2\tau})$, with $\gamma = \beta_1/\beta_2 > 0$. These estimators are thus valid if β_1 and β_2 have the same sign, i.e., the marker Y is a protective or a risk factor for both failures. This case is the more frequent in medical applications. However, if Y has an opposite effect between both failures ($\gamma < 0$), the corresponding equations are presented in Appendix C.

The marginal ROC curve, $ROC(\tau)$, and the its area under, $AUC(\tau)$ are obtained similarly to those obtained previously, but using the marginal se (7.4) and sp (7.5) instead of the se (7.2) and sp (7.3) of the score i.

7.2.2.2 Positive and negative predictive values

The main advantage of se and sp is their independence from the incidences of the failures. However, it is also important to study the real relevance of the prognostic marker in a given population, particularly when the study is conducted using a representative sample. For this purpose, the positive and negative predictive values (ppv and npv) can be informative.

The cause-specific definition - The ppv associated with the failure i for a prognostic at time τ is the probability that this failure occurs before τ for patients with a positive test $\{Z_i > c_{i\tau}\}$. The npv is the probability that this failure does not occur before τ for patients with a negative test $\{Z_i \leq c_{i\tau}\}$. We demonstrate in Appendix B that :

$$ppv_i(\tau|c_{i\tau}) = \int_{c_{i\tau}}^{\infty} (1 - S_i(\tau|z_i)) g_i(z_i) dz_i \Big/ \int_{c_{i\tau}}^{\infty} g_i(z_i) dz_i \quad (7.6)$$

$$npv_i(\tau|c_{i\tau}) = \int_{-\infty}^{c_{i\tau}} S_i(\tau|z_i) g_i(z_i) dz_i \Big/ \int_{-\infty}^{c_{i\tau}} g_i(z_i) dz_i \quad (7.7)$$

The marginal definition - We also define the marginal ppv and npv, which globally evaluate the performances of both scores. The marginal ppv is the probability that a failure occurs before the prognostic time for patients with at least one positive test.

The marginal npv is the probability that no failure occurs before the prognostic time for patients with both negative tests. If γ is positive (Appendix B), these quantities are equal to :

$$ppv(\tau|c_\tau) = \left\{ \sum_{i=1}^{2} P_i \int_{\omega_i}^{\infty} (1 - S_i(\tau|z_i)) g_i(z_i) dz_i \right\} \Big/ \left\{ \int_{\omega_1}^{\infty} g(z_1) dz_1 \right\} \quad (7.8)$$

$$npv(\tau|c_\tau) = \left\{ \sum_{i=1}^{2} P_i \int_{-\infty}^{\omega_i} S_i(\tau|z_i) g_i(z_i) dz_i \right\} \Big/ \left\{ \int_{-\infty}^{\omega_1} g(z_1) dz_1 \right\} \quad (7.9)$$

7.3 Analysis of kidney transplant recipients

The clinical objective is to study the accuracy of the 1-year CrCl for predicting the evolution of kidney transplant recipients until the $6th$ anniversary of transplantation (intermediate prognostic) and until the $10th$ anniversary of the transplantation (long-term prognostic). The origin of the follow-up ($t = 0$) being the first anniversary of transplantation, the prognostic times τ are equal to 5 and 9 years. At any time during the follow-up, a patient can occupy one of the following three states : stable with a functional kidney, returned to dialysis ($X = 1$) or died with a functional kidney ($X = 2$). The data are extracted from the DIVAT data bank, a French prospective study of kidney transplant recipients. The sample is made up of 2635 patients of more than 18 years of age and who received a kidney graft between January 1996 and September 2006. 215 patients returned to dialysis and 95 died with a functional kidney. The CrCl, calculated with the MDRD (Modification of Diet in Renal Disease), is dependent on the creatinine and takes into account recipient age and gender [91]. At 1-year, the mean CrCl is 50.9 ml/min (SD = 18.0).

7.3.1 Modeling assumptions

The last section provides a general framework for the three-class prognostic. However, depending on the analysed data, assumptions about the modeling must be made. We have used the Nelder-Mead algorithm to maximize the log likelihood function (7.1) and to compute the corresponding Hessian matrix (*optim* function in R). All the calculations of integrals are based on a simple trapezoidal rule.

Table 7.1: *Parameters of the semi-markovian model ($log\mathcal{V} = -1505.13$)*

Parameters	Estimations	Standard Deviations	p-values*
α	0.41	0.59	.
σ_1	2.31	0.69	.
ν_1	1.30	0.08	.
σ_2	18.35	9.33	.
β_1	-0.06	0.01	<0.0001
β_2	-0.02	0.01	0.0075

* Null hypothesis : the parameter is null (Wald test)

7.3.1.1 Distributions of the survival times

We have used the generalized Weibull distribution to fit the baseline survival functions : $S_{0i}(t) = exp(1 - (1 + (t/\sigma_i)^{\nu_i})^{1/\theta_i})$ with $\nu_i > 0$, $\sigma_i > 0$ and $\theta_i > 0$. This flexible 3-parameter distribution makes it possible to fit non-monotonic hazard functions : $\cup-$ or $\cap-$shaped [14]. It can be noticed that if $\theta_i = 1$, the survival function is Weibull-distributed. Moreover, if $\nu_i = 1$ the hazard function is constant (Exponential distribution).

A fully parametric approach is taken in the paper. This choice is motivated because we have already shown the adequacy of this 3-parameter Weibull distribution to the kidney transplant recipients survival [46] and because the equations (2) to (10) are easy to compute in comparison with a semiparametric or a nonparametric choice.

7.3.1.2 Distribution of the scores

Contrary to the distributions of the survival times, the distributions of the scores do not comply with any classic parametric law. We use a Gaussian kernel density estimator with 1000 points (*density* function in R).

7.3.2 The survival model

The estimated parameters of the semi-markovian model are described in Table (7.1). The corresponding log-likelihood equals -1505.13. The flexibility of the generalized Weibull distribution is useless compared to the retained distributions : Weibull for the times until a return to dialysis and Exponential for the time until a death (likelihood ratio statistic, null hypothesis $\theta_1 = \nu_2 = \theta_2 = 1$, $p > 0,05$). These results demonstrate the adequacy of the parametric approach.

The 1-year CrCl is significantly associated with both failures. The score z_1 equals -

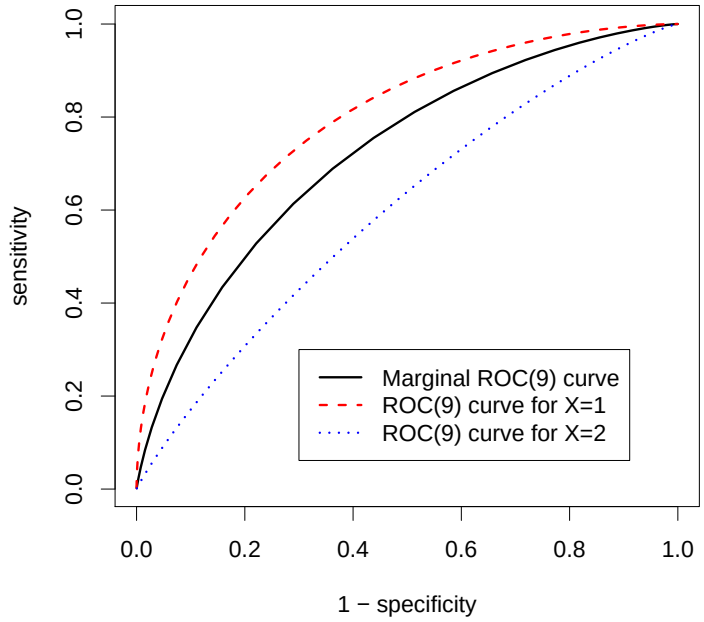

Figure 7.1: *ROC curves for a prognostic until the 10th anniversary of transplantation*

0.06×CrCl and the score z_2 equals -0.02×CrCl. High values of these scores are associated with high risk of the corresponding failure. More precisely, for patients returning to dialysis, an increase of 10 ml/min divides the risk by 1.8 ($p < 0,0001$). For the same difference, the division of the risk of failure is 1.2 for death ($p = 0,0075$).

7.3.3 Prognostic performances of the CrCl

7.3.3.1 Prognostic up to 9 years

Figure (7.1) describes the ROC curves for this prognostic period, i.e., $ROC_1(9)$, $ROC_2(9)$ and $ROC(9)$. Even if the CrCl is significantly associated with both failure times (p<0.05), it can be observed that the prognostic performance of this marker is clearly better for the return to dialysis ($ROC_1(9) = 0.81$, $IC_{95\%} = [0.75, 0.85]$) compared to death ($ROC_2(9) = 0.62$, $IC_{95\%} = [0.55, 0.69]$). Globally, the CrCl remains a good marker to predict one of the two failures ($ROC(9) = 0.75, IC_{95\%} = [0.71, 0.78]$).

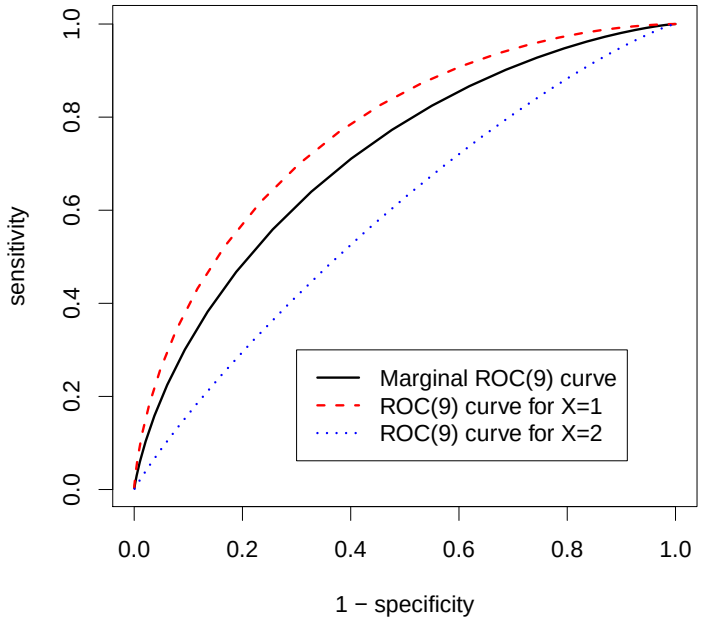

Figure 7.2: *ROC curves for a prognostic until the 6th anniversary of transplantation*

7.3.3.2 Prognostic up to $\tau = 5$ years

The prognostic capacities of CrCl are similar to the results obtained for $\tau = 9$ years (Figure 7.2). This similarity can be explained by the constant relationship between the CrCl and the hazard functions regardless of time (hazard proportionality). More precisely, the area under the ROC curve specific for return to dialysis equals 0.79 ($IC_{95\%} = [0.75, 0.83]$) and the area under the ROC curve specific for death with a functioning graft equals 0.62 ($IC_{95\%} = [0.55, 0.68]$). The marginal capacity of the CrCl to predict both failures is naturally in between ($ROC(5) = 0.74, IC_{95\%} = [0.70, 0.77]$).

7.4 Discussion

This paper proposed a method for a three-class and time-dependent prognostic analysis. This methodology was applied to the prognostic of kidney transplant recipient. We demonstrated that the CrCl constitutes an acceptable marker of return in dialysis but it cannot be used to predict death with functioning graft. These results are very different from those obtained by Kaplan, Schold and Meier-Kriesche [80]. The latter authors used traditional ROC curves unsuitable for analising time-dependent competing failures.

Saha and Heagerty [132] recently proposed another approach to evaluate the prognostic capacity of a marker in this context of competing events. They considered two measurements regarding the study of incident cases (probability that the failure occurs at a selected time) or cumulative cases (probability that the failure occurs before a selected time). Our proposed methodology is comparable regarding this second measurement. However, the main difference concerns the definitions of the sensitivities and specificities. In the paper of Saha and Heagerty, two sensitivities are estimated because two cases are defined ($X = 1$ or $X = 2$). In parallel, a single specificity is estimated because the control group represents patients without any failure. Both resulting ROC curves respectively represent the capacity of the marker to discriminate between (i) patients who will fail for the first cause ($X = 1$) and those who will have no failure, (ii) patients who will fail for the second cause ($X = 2$) and those who will have no failure. In our paper, we distinguish two definitions. In the cause-specific definition, two sensitivities are also defined for both cases ($X = 1$ or $X = 2$) but two control groups are respectively defined : patients without failure $X = 1$ and patients without failures $X = 2$. Both resulting ROC curves respectively represent the capacity of the marker to discriminate between (i) patients who will fail or not for the first cause ($X = 1$), (ii) patients who will fail or not for the second cause ($X = 2$). Additionally, we proposed a marginal definition where the cases are the patients who will fail (regardless of the cause) and the control group is the patients without any failure. The corresponding ROC curve represents the capacity of the marker to discriminate between patients who will fail or not (regardless of the cause). From our modeling, it is thus possible to compute the ROC curves defined by Saha and Heagerty [132] by plotting the cause-specific sensitivities against the marginal specificities.

Regarding our available data, we have made certain modeling choices (proportionality of hazards and generalized Weibull distribution) and we have tested if these assumptions hold true. Other approaches for modeling the survival process and the markers are also feasible. For example, even though the parametric approach seems adequate in our application, the baseline survival function could have been estimated by a non-parametric method [77] and the marker could have been introduced using the accelerated failure time approach. One could also consider using a semi-parametric model, which could also be applied to estimate time-dependent sensitivity and specificity [138]. However, the general framework remains independent from these modeling assumptions. A way of improvement would be to allow the marker to affect the probabilities of having events of type 1 versus 2, i.e. P_i $(i = 1, 2)$. This natural consideration is nevertheless associated with a problem of identifiability between the effects of the marker on the two parts of the semi-Markov process.

Acknowledgements

We thank the Roche Laboratory for its support of the DIVAT data bank and Dr J. Ashton-Chess for editing the manuscript. This work is a part of the French Transplantation Research Network (RTRS) supported by the "Fondation de Cooperation Scientifique CENTAURE".

Appendix of the papers by Foucher et al. (SIM, 2010)

Appendix A : Demonstrations of sensibilities and specificities

– The se of the score Z_i for the prediction of the failure i :

$$\begin{aligned} se_i(c_{i\tau}|\tau) &= P(z_i > c_{i\tau}|T \leq \tau, X = i) \\ &= P(z_i > c_{i\tau}, T \leq \tau|X = i)/P(T \leq \tau|X = i) \\ &= \int_{c_{i\tau}}^{\infty}(1 - S_i(\tau|z_i))g_i(z_i)dz_i \Big/ \int_{-\infty}^{\infty}(1 - S_i(\tau|z_i))g_i(z_i)dz_i \end{aligned}$$

– The sp of the score Z_i for the prediction of the absence of failure i (Equation 3) :

$$\begin{aligned} sp_i(c_{i\tau}|\tau) &= P(z_i \leq c_{i\tau}|T > \tau, X = i) \\ &= P(z_i \leq c_{i\tau}, T > \tau|X = i)/P(T > \tau|X = i) \\ &= \int_{-\infty}^{c_{i\tau}} S_i(\tau|z_i)g_i(z_i)dz_i \Big/ \int_{-\infty}^{\infty} S_i(\tau|z_i)g_i(z_i)dz_i \end{aligned}$$

– The marginal se of both scores for the prediction of at least one failure (Equation 4) :

$$\begin{aligned} se(c_\tau|\tau) &= P(\overline{A}|T \leq \tau) \\ &= P(\overline{A}, T \leq \tau)/P(T \leq \tau) \\ &= 1 - P(A, T \leq \tau)/P(T \leq \tau) \end{aligned}$$

where $c_\tau = (c_{1\tau}, c_{2\tau})$, $A = \{z_1 \leq c_{1\tau}, z_2 \leq c_{2\tau}\}$ and $\overline{A}$ is not A. Using the law of total probability, we have :

$$se(c_\tau|\tau) = 1 - \left\{\sum_{i=1}^{2} P_i P(A, T \leq \tau|X = i)\right\} \Big/ \left\{\sum_{i=1}^{2} P_i P(T \leq \tau|X = i)\right\}$$

However, since $z_1 = z_2/\gamma$, with $\gamma = \beta_2/\beta_1$, we obtain if $\gamma > 0$:

$$\begin{aligned} se(c_\tau|\tau) &= 1 - \left\{ \sum_{i=1}^{2} P_i P(z_i \leq \omega_i, T \leq \tau | X = i) \right\} \\ &\times \left\{ \sum_{i=1}^{2} P_i P(z_i > -\infty, T \leq \tau | X = i) \right\}^{-1} \\ &= 1 - \left\{ \sum_{i=1}^{2} P_i \int_{-\infty}^{\omega_i} (1 - S_i(\tau|z_i)) g_i(z_i) dz_i \right\} \\ &\times \left\{ \sum_{i=1}^{2} P_i \int_{-\infty}^{\infty} (1 - S_i(\tau|z_i)) g_i(z_i) dz_i \right\}^{-1} \end{aligned}$$

where $\omega_1 = min(c_{1\tau}, \gamma^{-1} c_{2\tau})$ et $\omega_2 = min(\gamma c_{1\tau}, c_{2\tau})$.

– The marginal sp of both scores for the prediction of the absence of failure (Equation 5) :

$$sp(c_\tau|\tau) = P(A|T > \tau) = P(A, T > \tau)/P(T > \tau)$$

Thus, based on the same notations as the marginal se, we obtain :

$$\begin{aligned} sp(c_\tau|\tau) &= \left\{ \sum_{i=1}^{2} P_i P(A, T > \tau | X = i) \right\} \\ &\times \left\{ \sum_{i=1}^{2} P_i P(T > \tau | X = i) \right\}^{-1} \\ &= \left\{ \sum_{i=1}^{2} P_i P(z_i \leq \omega_i, T > \tau | X = i) \right\} \\ &\times \left\{ \sum_{i=1}^{2} P_i P(z_i > -\infty, T > \tau | X = i) \right\}^{-1} \\ &= \left\{ \sum_{i=1}^{2} P_i \int_{-\infty}^{\omega_i} S_i(\tau|z_i) g_i(z_i) dz_i \right\} \\ &\times \left\{ \sum_{i=1}^{2} P_i \int_{-\infty}^{\infty} S_i(\tau|z_i) g_i(z_i) dz_i \right\}^{-1} \end{aligned}$$

Appendix B : Demonstrations of the positive and negative predictive values

– The ppv of the score Z_i for the prediction of the failure i (Equation 6) :

$$\begin{aligned} ppv_i(\tau|c_{i\tau}) &= P(T \leq \tau | z_i > c_{i\tau}, X = i) \\ &= P(T \leq \tau, z_i > c_{i\tau} | X = i)/P(z_i > c_{i\tau}) \\ &= \int_{c_{i\tau}}^{\infty} (1 - S_i(\tau|z_i)) g_i(z_i) dz_i \Big/ \int_{c_{i\tau}}^{\infty} g_i(z_i) dz_i \end{aligned}$$

– The npv of the score Z_i for the prediction of the absence of failure i (Equation 7) :

$$\begin{aligned} npv_i(\tau|c_{i\tau}) &= P(T > \tau | X = i, z_i \leq c_{i\tau}) \\ &= P(T > \tau, z_i \leq c_{i\tau} | X = i)/P(z_i \leq c_{i\tau}) \\ &= \int_{-\infty}^{c_{i\tau}} S_i(\tau|z_i) g_i(z_i) dz_i \Big/ \int_{-\infty}^{c_{i\tau}} g_i(z_i) dz_i \end{aligned}$$

– The marginal ppv of both scores for the prediction of at least one failure (Equation 8) :

$$\begin{aligned} ppv(\tau|c_\tau) &= P(T \leq \tau | \overline{A}) \\ &= P(T \leq \tau, \overline{A})/P(\overline{A}) \\ &= \{P(T \leq \tau) - P(T \leq \tau, A)\}/\{1 - P(A)\} \\ &= \left\{ \sum_{i=1}^{2} P_i P(T \leq \tau | X = i) - \sum_{i=1}^{2} P_i P(T \leq \tau, z_i \leq \omega_i | X = i) \right\} \\ &\times \left\{ 1 - P(z_1 \leq \omega_1) \right\}^{-1} \\ &= P(z_1 > \omega_1)^{-1} \sum_{i=1}^{2} P_i P(T \leq \tau, z_i > \omega_i | X = i) \\ &= \left\{ \sum_{i=1}^{2} P_i \int_{\omega_i}^{\infty} (1 - S_i(\tau|z_i)) g_i(z_i) dz_i \right\} \Big/ \left\{ \int_{\omega_1}^{\infty} g(z_1) dz_1 \right\} \end{aligned}$$

– The marginal npv of both scores for the prediction of the absence of failure (Equa-

tion 9) :

$$\begin{aligned} npv(\tau|c_\tau) &= P(T>\tau|A) \\ &= P(T>\tau,A)/P(A) \\ &= P(z_1\leq\omega_1)^{-1}\sum_{i=1}^{2}P_iP(T>\tau,z_i\leq\omega_i|X=i) \\ &= \left\{\sum_{i=1}^{2}P_i\int_{-\infty}^{\omega_i}S_i(\tau|z_i)g_i(z_i)dz_i\right\}\Big/\left\{\int_{-\infty}^{\omega_1}g(z_1)dz_1\right\} \end{aligned}$$

Appendix C : Special developments for $\gamma < 0$

The marginal se and sp are valid if the parameters of the semi-markovian model meet the condition $\gamma > 0$. However, if the signs of β_1 and β_2 are different, then :

$$\begin{aligned} se(c_\tau|\tau) &= 1-\left\{\sum_{i=1}^{2}P_i\int_{\kappa_i}^{c_{i\tau}}(1-S_i(\tau|z_i))g_i(z_i)dz_i\right\} \\ &\times \left\{\sum_{i=1}^{2}P_i\int_{-\infty}^{\infty}(1-S_i(\tau|z_i))g_i(z_i)dz_i\right\}^{-1} \\ sp(c_\tau|\tau) &= \left\{\sum_{i=1}^{2}P_i\int_{\kappa_i}^{c_{i\tau}}S_i(\tau|z_i)g_i(z_i)dz_i\right\} \\ &\times \left\{\sum_{i=1}^{2}P_i\int_{-\infty}^{\infty}S_i(\tau|z_i)g_i(z_i)dz_i\right\}^{-1} \end{aligned}$$

where $\kappa_1 = c_{2\tau}/\gamma$ and $\kappa_2 = \gamma c_{1\tau}$.

In parallel, the ppv and npv are rewritten as follows :

$$\begin{aligned} ppv(\tau|c_\tau) &= \left\{\sum_{i=1}^{2}P_i\int_{-\infty}^{\infty}(1-S_i(\tau|z_i))g_i(z_i)dz_i\right. \\ &- \left.\sum_{i=1}^{2}P_i\int_{\kappa_i}^{c_{i\tau}}(1-S_i(\tau|z_i))g_i(z_i)dz_i\right\} \\ &\times \left\{1-\int_{\kappa_1}^{c_{1\tau}}g(z_1)dz_1\right\}^{-1} \\ npv(\tau|c_\tau) &= \left\{\sum_{i=1}^{2}P_i\int_{\kappa_i}^{c_{i\tau}}S_i(\tau|z_i)g_i(z_i)dz_i\right\}\Big/\left\{\int_{\kappa_1}^{c_{1\tau}}g(z_1)dz_1\right\} \end{aligned}$$

- Chapitre 8 -

A flexible semi-Markov model for interval-censored data and goodness-of-fit testing

Foucher Y. Giral M., Soulillou J.P. and Daurès J.P.
Statistical Methods in Medical Research. 2010 Apr ;19(2) :127-45.

8.1 Introduction

In longitudinal analyses, multistate models are becoming increasingly popular to deal with the complex evolution of chronic diseases. The use of the Markov chain appears to be very useful for this purpose [87, 83, 8]. Another approach consists of modeling the probabilities of transitions according to the times spent in different states, the Markov chain being only associated with the sequence of states [113]. These semi-Markov models (SMM) are adapted to certain diseases, particularly to transplantation [24]. In this paper, we use this approach to study the evolution of kidney transplant recipients (a cohort followed up at Nantes University Hospital in France).

The first methodology issue concerns the use of a continuous marker in order to define the states of disease severity of a patient during his/her follow-up. This marker is available at certain visits, then the transitions between these transient states are interval-censored. This means that the transition times and the sequence of states are unknown during the elapsed time between two consecutive visits. This type of data is often encountered in longitudinal studies. The time of entry into the study and the time of the final event (for example the date of death) are exactly known ; contrary to the intermediate times of transitions, which are interval censored because of the irregularity of the observation process. In this paper, we thus chose to focus the modelling on multi-state data where only intermediate and transient states are interval censored.

Sternberg and Satten [142] proposed non-parametric estimators based on an EM algorithm and showed that a SMM can still be applied to interval-censored data with a unidirectional model without covariates. From a practical standpoint, parametric models are appealing since the parameters can be estimated at a $\sqrt{n}$ rate, leading to a smaller sample size to achieve a given accuracy. Furthermore the parameters are often interpretable and the covariates can easily be included. As noted by Lindsey [94], parametric regression models in the presence of heavily interval-censored data are robust. In this paper, we define a SMM based on the generalized Weibull distribution. Convolution products are used to adequately deal with interval-censored transitions, contrary to the recent paper by Foucher et al. [45], in which the interval-censoring concerns only the duration in a given state without considering the censoring of the sequence of consecutive states. Kang and Lagakos [79] also insist on the necessity to take into account all the information available in such a type of incomplete data, that is to say, both the times of transitions and the nature of the states visited are unknown.

Another extension consists in the introduction of covariates through the probability density functions of the times spent in each state and through the Markov chain associated with the sequence of states. The explanatory factors can then be associated either with the speeds or with the trajectories of the process, contrary to most studies where the covariates are only related to the durations [113, 24, 38, 70, 49].

The second issue in modeling multistate processes is the goodness-of-fit. Authors generally grant little attention to this point. Most available tests concern Markov models assuming that, for each subject, the number of visits is fixed and that no explanatory variables are measured [78]. The main assumption of the proposed SMM is the stationarity related to the time since the inclusion of patients. Castelli et al. [19] recently proposed a goodness-of-fit test of this stationary, but the authors did not deal with the interval censoring. Based on the recent works on the Markov process by Aguirre and Farewell [3], we herein define a Pearson-type goodness-of-fit statistic in order to determine whether this stationary assumption of the SMM is valid in the presence of interval censored data. However, their method only considers interval-censored times of transition between transient states. Like the modified goodness-of-fit test proposed by Titman and Sharples [147] for Markov or hidden Markov models, the statistic defined in the present paper takes into account the exact times of transition into absorbing states for the semi-Markov process.

We first present the data on which the model is based in Section 2. The modeling of the resulting multistate data is considered in Section 3. The test of the stationarity is presented in Section 4 and the methods and results are discussed in Section 5.

8.2 Data and multi-state structure

Such developments were motivated by the analysis of a French prospective study of kidney transplant recipients, extracted from the DIVAT (Données Informatisées et VAlidées en Transplantation) data bank where biological and clinical data of transplant patients have been recorded since 1990 at Nantes University Hospital. Data were computerized at each checkup visit and at each anniversary of the transplantation. See the paper by Giral et al. [53] for more details. Suppose that the sample consists of n subjects, denoted by h $(h = 1, 2, ..., n)$. The follow-up of each patient is a succession of visits at times since the transplantation $\{v_{h,0}, v_{h,1}, ..., v_{h,n_h}\}$. The mean and the median of the number of visits per patient are respectively 30 and 25.

In order to obtain a homogeneous sample, our data set concerns 839 patients older than 18 years and having received a kidney transplant between January 1996 and September 2006. The follow-up started 3 months after the date of transplantation (all patients had a visit at 3 months). The creatinine clearance (CL) is then regarded as stabilized. CL, calculated using the modification of diet in renal disease formula [91], is recorded at each visit and is used as a marker of the clinical aggravation towards the final events. We consider that all patients enter into an initial state where the risk of failure is low (state 1). For a subject, if her/his value of CL decreases by more than 30% of the maximum observed value, we consider that she/he transits into a state where the risk of failure is high (state 2). Absolute values of CL are not used to classify health-states, since preliminary analyses have shown that it is more informative to consider relative changes [41]. From state 2, no return into state 1 is considered, the medical assumption is of a nonreversible aggravation. Of course, the hypothesis of no backward transition between states 2 and 1 is a substantial assumption. If the state 2 is observed for a patient, the number of the following state will be greater than or equal to 2, whatever the CL level. In other words, we suggest that the recovery of the kidney after such a shock is not possible. However, the possible recovery should constitute a stronger assumption from a clinical point of view.

Two final events are recorded with their exact date of occurrence : the return to dialysis (state 3) and the death of the patient (state 4). In contrast to the transient states of disease severity, state 3 and state 4 are absorbing. The $3 \rightarrow 4$ transition is impossible since the follow-up of patients in the DIVAT data bank is stopped if the patient returns to dialysis. According to this structure, which is presented in Figure 8.1, the visit v_{h,n_h} relates to a final event or right-censoring with a functional transplant. Table (8.1) offers some additional descriptions of the dataset. For instance, 190 patients are right-censored in state 2 with a mean follow-up time since the transplantation of 4

Table 8.1: *Description of the sample according to the possible observed trajectories*

Observed states	Size	Percentages	mean[a]	median[a]
1	537	64,0 %	3.79	3.48
1 ; 2	190	22,7 %	4.05	3.63
1 ; 2 ; 3	61	7,3 %	3.63	3.61
1 ; 2 ; 4	18	2,1 %	2.83	2.51
1 ; 3	16	1,9 %	1.75	1.39
1 ; 4	17	2,0 %	2.74	1.18
total	839	100,0 %	3.77	3.44

[a] mean/median follow-up times in years

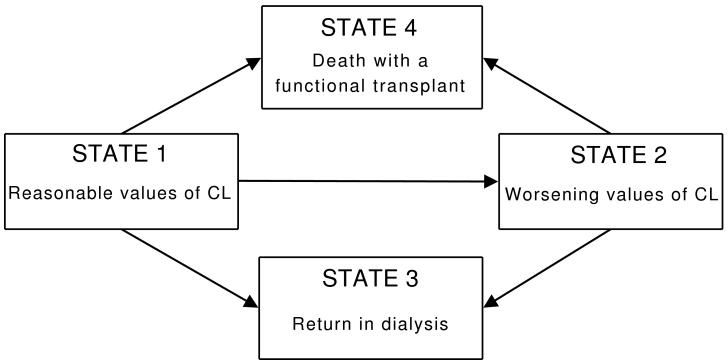

Figure 8.1: *Four-state model for the study of kidney transplant recipient evolution according to a worsening/failure structure.*

years.

Even though the description of this structure concerns the analysis of kidney transplant recipients, this type of model can also be adapted to other longitudinal data analyses. In many applications, the transient states of disease severity may be interval censored and the date of final failures exactly known.

Nine explanatory variables have been retained, the reference group is specified in the brackets : donor and recipient gender (women), cold ischemia time (less than 24 hours), donor and recipient age at the time of transplantation (less than 55 years), number of HLA-incompatibilities (less than 4), induction treatment (no Simulect), panel reactive antibody (0 %) and delayed graft function (less than 6 days). These thresholds were chosen by the clinicians according to the literature.

8.3 The semi-Markov model

8.3.1 The semi-Markov framework

Let $\{X_{h,r}, r = 1, ..., n_h\}$ be the observed sequence of states at each visit for the hth subject ($h = 1, ..., n$), where n_h is the number of visits for this subject. By definition, all patients begin in state 1, i.e. $X_{h,1} = 1$. If the end of the follow-up corresponds to return to dialysis, then $X_{h,n_h} = 3$. If the final event consists in the death of the patient, then $X_{h,n_h} = 4$. Otherwise, $\{X_{h,r}, r = 1, ..., n_h\} \in \{1, 2\}$. From $\{X_{h,r}, r = 0, ..., n_h\}$, the sequence of the distinct observed states can directly be deduced : $\{W_{h,r}, r = 0, ..., m_h\}$ where m_h is the number of observed transitions. By definition, $W_{h,0} = 1$. This sequence forms a Markov chain. Assuming the stationarity of the process with the time since the transplantation, the probabilities of jumping from state $W_{h,r} = i$ to state $W_{h,r+1} = j$, associated with this chain, can be written as :

$$P_{ij} = P(W_{h,r+1} = j | W_{h,r} = i) \tag{8.1}$$

with the constraint

$$\sum_j P_{ij} = 1. \tag{8.2}$$

If state i is not persistent, then $P_{ij} \geq 0$ for $i \neq j$ and $P_{ij} = 0$ for $i = j$. Otherwise, if state i is a final event, then $P_{ij} = 0$ for $i \neq j$ and $P_{ij} = 1$ for $i = j$. The semi-Markov property is assumed for the process, that is to say, time is reset to zero at each transition. The probability density function (PDF) of the duration in state $W_{h,r} = i$, before jumping to state $W_{h,r+1} = j$ (for $i \neq j$), is given by :

$$f_{ij}(d_{h,r}) = \lim_{\Delta d \to 0^+} P(d_{h,r} < D_{h,r} < d_{h,r} + \Delta d | W_{h,r+1} = j, W_{h,r} = i)/\Delta d \tag{8.3}$$

in which $D_{h,r}$ is the time spent in state $W_{h,r}$. As is usual in survival analysis, we deduce from f_{ij} the corresponding survival, hazard and cumulative hazard functions, S_{ij}, λ_{ij} and Λ_{ij} respectively. One can notice that, for a particular multi-state structure in which there is one initial state and a few absorbing states (without transient states), the proposed SMM model generalizes the traditional competing risk model (CRM). Indeed, for this type of CRM, the density function specific to the transition $i \to j$ is based on the following joined probability : $\lim_{\Delta d \to 0^+} P(d_{h,r} < D_{h,r} < d_{h,r} + \Delta d, W_{h,r+1} = j | W_{h,r} = i)/\Delta d$. This quantity is equal to $P_{ij} f_{ij}(d_{h,r})$. Our separate modelling of the trajectories (P_{ij}) and of the times of transitions (f_{ij}) provides a high degree of flexibility and is very convenient for result interpretations. For instance, P_{ij} directly represents the prediction of the proportion of transition $i \to j$ if all patients were followed-up until their final

failure. Moreover, and as we will see later, it is possible to incorporate covariates either through P_{ij} or through f_{ij}.

We based our modeling strategy on the generalized Weibull distribution with

$$\lambda_{ij}(d_{h,r}) = \theta_{ij}^{-1}(1 + (d_{h,r}/\sigma_{ij})^{\nu_{ij}})^{1/\theta_{ij}-1}(\nu_{ij}/\sigma_{ij})(d_{h,r}/\sigma_{ij})^{\nu_{ij}-1} \tag{8.4}$$

for all $d_{h,r} \geq 0$, $\nu_{ij} > 0$, $\sigma_{ij} > 0$ and $\theta_{ij} > 0$. This class of distribution has interesting properties [14]. Depending on the parameter values, the hazard function can be either constant, monotone (increasing or decreasing), $\cup-$shaped or $\cap-$shaped. For the non-monotonic functions, it is interesting to calculate the times corresponding to the minimum or the maximum hazard rate. If $0 < \theta_{ij} < \nu_{ij} < 1$, then the hazard function of the $i \to j$ transition decreases from ∞ to its minimum value at the time

$$c_{ij} = \sigma_{ij}\Big(\frac{\theta_{ij} - \nu_{ij}\theta_{ij}}{\nu_{ij} - \theta_{ij}}\Big)^{1/\nu_{ij}} \tag{8.5}$$

and then increases to ∞, it is then $\cup-$shaped. If $\theta_{ij} > \nu_{ij} > 1$, then the hazard rate increases from 0 to its maximum value at the time c_{ij} and then decreases to 0, it is then $\cap-$shaped. We can easily simplify this distribution using the likelihood ratio statistic (LRS). For example, if θ_{ij} is fixed at 1, the Weibull formulation is obtained. Moreover, if ν_{ij} equals 1, the hazard function is constant (exponential distribution).

8.3.2 Likelihood function

In order to compute the likelihood function, we need to identify the different observed trajectories according to Figure 8.1. Such trajectories are described in Figure 8.2. Demonstrations of the following developments are included in the Appendix.

Trajectory (i) - The individual h is observed in both states of disease severity before the occurrence of a final event k ($k = 3, 4$). The duration in the first state, $d_{h,0}$, is included in the interval $]d^0_{h,0}, d^1_{h,0}]$. The event k appears a time v_{h,n_h} after the transplantation. Let $C_{h,1}$ represent this individual likelihood contribution.

$$C_{h,1} = P_{12}P_{2k}\int_{d^0_{h,0}}^{d^1_{h,0}} f_{12}(u)f_{2k}(v_{h,n_h} - u)du \tag{8.6}$$

Trajectory (ii) - The individual h is not observed in state 2 before the occurrence of a final event k ($k = 3, 4$). Let $C_{h,2}$ be this contribution. Figure 8.1 shows that the patient may directly jump from state 1 to state k, but he/she may have passed through state 2. This observation illustrates the interval censoring : the durations and the sequence of

states.

$$C_{h,2} = P_{1k} f_{1k}(v_{h,n_h}) + P_{12} P_{2k} \int_{d^0_{h,0}}^{v_{h,n_h}} f_{12}(u) f_{2k}(v_{h,n_h} - u) du \tag{8.7}$$

Trajectory (iii) - The individual h is right-censored in state 1 at the time of her/his last visit v_{h,n_h}. Let $C_{h,3}$ denote this type of observation.

$$C_{h,3} = \sum_{j=2}^{4} P_{1j} S_{1j}(v_{h,n_h}) \tag{8.8}$$

Trajectory (iv) - The individual h is right-censored in state 2 at the time of his/her last visit v_{h,n_h}. The contribution of such trajectory is

$$C_{h,4} = P_{12} \int_{d^0_{h,0}}^{d^1_{h,0}} f_{12}(u) \Big\{ \sum_{j=3}^{4} P_{2j} S_{2j}(v_{h,n_h} - u) \Big\} du \tag{8.9}$$

Making the product of these individual contributions, we obtain the likelihood function of the observed sample. We used the quasi-Newtonian algorithm to maximize this function and to compute the Hessian matrix. This optimization was performed using the R software with the *optim* function.

The integral calculations are numerically approximated using the 10-points Gauss-Legendre quadrature. In order to evaluate the accuracy of the method, we also used a 20-points quadrature. One should notice that the previous integrals may include singularities :

a. expressions (8.6) and (8.9) present a singularity at $u = 0$, where $d^0_{h,0} = 0$ and $\nu_{12} < 1$. Tacking $s = (\frac{u}{\sigma_{12}})^{\nu_{12}}$ will remove this singularity.

b. expression (8.6) gives a singularity at $u = v_{h,n_h}$, where $d^1_{h,0} = v_{h,n_h}$ and $\nu_{2k} < 1$.

c. expression (8.7) gives a singularity at $u = v_{h,n_h}$, where $\nu_{2k} < 1$ and $d^0_{h,0} > 0$. Tacking $s = (\frac{v_{h,n_h} - u}{\sigma_{2k}})^{\nu_{2k}}$ will remove the singularities in cases b. and c.

d. expression (8.6) gives singularities at both $u = 0$ and $u = v_{h,n_h}$, where $\nu_{12} < 1$, $\nu_{2k} < 1$, $d^0_{h,0} = 0$ and $d^1_{h,0} = v_{h,n_h}$.

e. expression (8.7) gives singularities at both $u = 0$ and $u = v_{h,n_h}$, where $\nu_{12} < 1$, $\nu_{2k} < 1$ and $d^0_{h,0} = 0$. Cases d. and e. are more difficult to solve. One way of removing the singularities is to split the integral into two parts, the first part from 0 to $0.5 v_{h,n_h}$ and the second part from $0.5 v_{h,n_h}$ to v_{h,n_h}. Each part has only one singularity and previous substitutions can then be applied.

According to our database and the number of observations per subject, only 4 patients are concerned. These patients correspond to the trajectory (8.9) with $d^0_{h,0} = 0$. For these contributions, we have made the appropriate substitution. If many of the observed

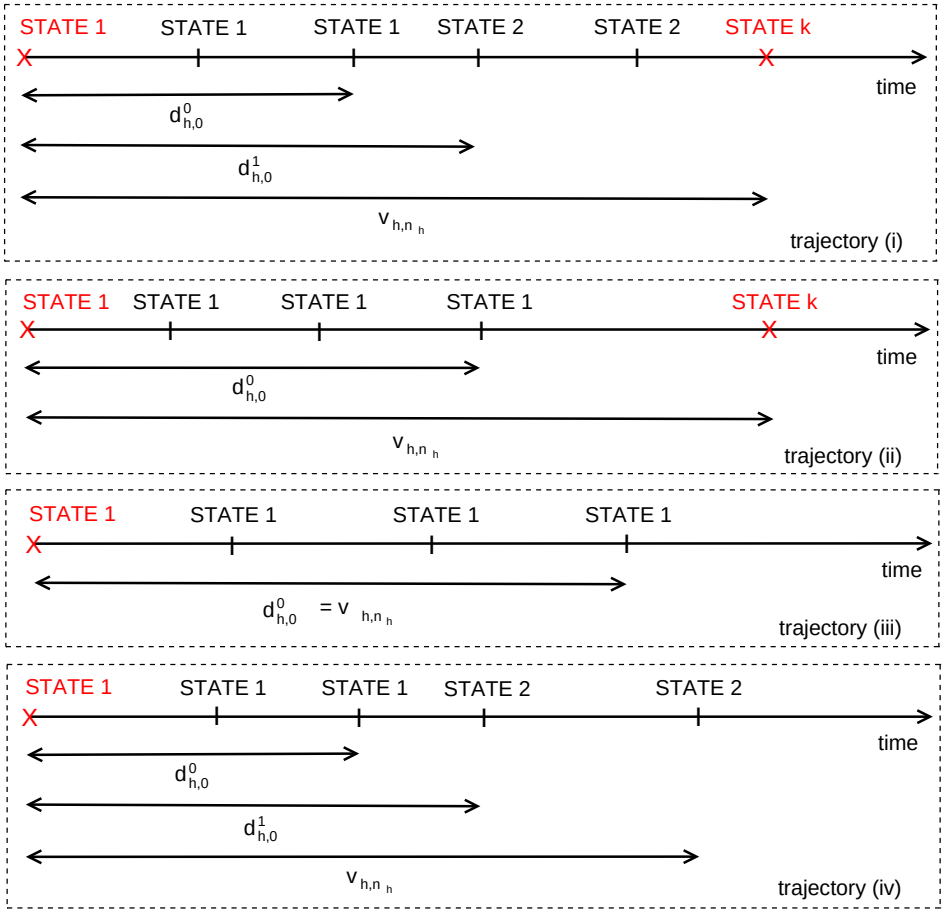

Figure 8.2: *Possible trajectories of a kidney transplant recipient.* | | | *Observed states of disease severity during visits ;* × × × *Events for which the time of entrance since the transplantation is exactly known.*

trajectories was associated with singularities, the Gauss-Legendre quadrature would not have been the best approach. The adaptive numerical integration (function *integrate* in R) would be preferable.

8.3.3 Incorporation of explanatory factors

In SMM, most authors include covariates through the PDF of the times spent in states, using the proportionality assumption. Many recent works in traditional survival analysis have shown that the assumption of proportionality does not hold in many cases and may lead to serious bias [107]. The same issue arises in multistate processes. We consider that the covariates are fixed in time and that the corresponding regression parameters depend on the durations. The hazard function of durations is defined by :

$$\lambda_{ij}(d_{h,r}, z_{h,ij}) = \lambda_{0,ij}(d_{h,r}) exp(\gamma_{ij}(d_{h,r}) z_{h,ij}) \tag{8.10}$$

in which $\lambda_{0,ij}$ is the baseline hazard function at the duration $d_{h,r}$ specific to the transition from state i to state j. We assume that this function corresponds to a generalized Weibull distribution. $\gamma_{ij}(d_{h,r})^T = (\gamma_{ij}^1(d_{h,r}), \gamma_{ij}^2(d_{h,r}), ..., \gamma_{ij}^{m_{ij}}(d_{h,r}))$ is the vector of time-dependent regression parameters, associated with the m_{ij} time fixed covariates $z_{h,ij} = (z_{h,ij}^1, z_{h,ij}^2, ..., z_{h,ij}^{m_{ij}})$ of the hth subject. Notice that the effect of the explanatory variables can change from one transition to another. In order to fit time varying effects, we adopt the method of Stablein, Carter and Nova [140] by introducing some time interaction terms. Although the linear term would eventually be sufficient to allow for a time-varying relationship, the quadratic term allows for the modeling of a surface that rises and falls, constituting a non-monotonic time dependence. This method of modelling time-dependent effects is particularly adapted to the following application since the related hazard functions seem to be exponentially distributed. Moreover, the inclusion in the exponential function of interactions with durations does not change the positivity of the hazard function, regardless of the value of regression parameters. One can also notice that the regression parameters are always interpretable as hazard ratios.

In order to compute the likelihood, we need the survival function associated with (8.10). No analytical solution is available. Using again the 10-points Gauss-Legendre quadrature, $exp(-\Lambda_{0,ij}(d_{h,r}, z_{h,ij}))$ is numerically approximated. Log-Minus-Log survival plots are used to determine if baseline hazard functions are proportional across groups of categorical variables.

Parallel to this approach, we propose to include the explanatory factors through the Markov chain (8.1). The idea is to take into account heterogeneity in the sequences of the process. Using regression models for categorical dependent variables with more than two response categories [98], the probabilities of jumping from the state $W_{h,r} = i$ to the state $W_{h,r+1} = j$, can be written as :

$$P_{ij}(\psi_{h,ik}, k \neq i) = exp(\beta_{ij}\psi_{h,ij}) / \sum_{k \neq i} exp(\beta_{ik}\psi_{h,ik}) \tag{8.11}$$

where $\beta_{ij}^T = (\beta_{ij}^0, \beta_{ij}^1, ..., \beta_{ij}^{u_{ij}})$ is the vector of regression parameters associated with the u_{ij} covariates $\psi_{h,ij} = (1, \psi_{ij}^1, ..., \psi_{ij}^{u_{ij}})$. The constraint (8.2) imposes a reference state. From state 1, a patient may transit to states 2, 3 or 4. By convention, we fix $\beta_{14} = 0$. Identically, $\beta_{24} = 0$.

8.3.4 Results

Explanatory factors and PDF parameters are first tested in univariate ($p \leq 0.20$) and the assumption of proportionality is graphically examined. Among the 90 possible parameters for the complete model without interaction, 42 are retained. We have plotted

Table 8.2: *Regression parameters of the semi-Markov model*

Transition	Covariate	Coef.	SE	p-value
Regression parameters associated with the trajectories, $P_{ij}()$.				
$1 \to 2$	Intercept	1.87	0.68	0.0064
$1 \to 2$	Delayed graft function	0.73	0.34	0.0302
$1 \to 2$	Donor age	-2.03	0.69	0.0035
$1 \to 3$	Intercept	-2.81	0.53	0.0001
$2 \to 3$	Intercept	1.13	0.34	0.0008
$2 \to 3$	Incompatibilities A+B+DR	0.93	0.47	0.0473
Regression parameters associated with the durations, $f_{ij}()$.				
$1 \to 2$	Induction treatment	0.36	0.14	0.0094
$1 \to 2$	Recipient gender	-0.26	0.13	0.0542
$1 \to 2$	Donor age	0.96	0.23	0.0001
$1 \to 3$	Cold ischemia time	5.02	1.20	0.0001
$2 \to 3$	Incompatibilities A+B+DR	0.88	0.28	0.0019
$2 \to 3$	Panel reactive antibody	1.09	0.35	0.0016
$2 \to 3$	Panel reactive antibody $\times$ d [a]	-0.48	0.22	0.0308
$2 \to 4$	Delayed graft function	2.03	0.60	0.0008
$2 \to 4$	Recipient gender	1.54	0.64	0.0165
$2 \to 4$	Recipient gender $\times$ d [a]	-4.30	1.19	0.0003
$2 \to 4$	Recipient gender $\times$ d^2 [a]	1.30	0.32	0.0001

[a] Time interaction with the duration d in the state

the logarithm of the estimated cumulative hazard function against the duration, for each modality of each covariate and for all transitions. If the resulting curves are nonparallel, then we model the effect of this covariate on this transition by adding interactions with the duration. After the descending procedure ($p \leq 0.05$), 11 factors are finally retained with a log likelihood equals -1550.71, regardless of the number of points in the Gauss-Legendre quadrature. The estimations of regression parameters are given in Table 8.2. Notice that in the case of convergence problems associated with the 42-parameters model, it would have been more adequate to adopt an ascending method to construct the multivariate model. In our database, the $1 \to 2$ transitions are interval censored into a short time period with a median equal to 2.5 months. Moreover, the proportion of observations for which the trajectory of the subject is unknown (contributions 8.7) is relatively small, representing only 3.9% of the total number of contributions. Problems of fitting may have been greater if less observations were made per patient, i.e. higher interval censored periods and a higher proportion of unknown trajectories.

Concerning the trajectories, the $1 \to 2$ transition is more likely in patients with

Table 8.3: *PDF parameters related to the durations in states*

	σ_{ij}		ν_{ij}		θ_{ij}	
Transition	Estim.	SE	Estim.	SE	Estim.	SE
$1 \to 2$	68.80	76.53	0.77	0.05	0.25	0.20
$1 \to 3$	42.84	47.34	1	.	1	.
$1 \to 4$	109.83	73.00	1	.	1	.
$2 \to 3$	12.66	3.20	1	.	1	.
$2 \to 4$	6.54	4.13	1	.	1	.

a delayed graft function greater than or equal to 6 days compared to the others. In contrast, receiving a kidney from a donor greater than 55 years of age is a risk factor for death, given that the initial state is occupied. Finally, kidney transplant recipients with more than 4 HLA-incompatibilities (A+B+DR) constitute a risk factor for returning to dialysis, given that the second state is occupied.

The effects of covariates on the durations in the different states can also be examined. Given that state 2 follows, the time spent in the initial state is shorter for men treated by Simulect or receiving a transplant from a donor greater than 55 years of age. Given that the subject returned to dialysis from the initial state, this transition seems to be accelerated when the cold ischemia time is greater than 24 hours. For subjects in the worsened state who returned to dialysis, the transition is faster for transplantations with high HLA-incompatibilities. Again for this transition, individuals with panel reactive antibody (greater than 0 %) seem to transit quicker than the others. However, this effect decreases linearly according to time spent in the worsened state. Finally, regarding patient death, the duration in state 2 appears to be shorter for men with a long delayed graft function. The regression coefficient associated with gender varies across the time spent in the worsened state.

Concerning the PDF parameters, presented in Table 8.3, the duration hazard function of the $1 \to 2$ transition is $\cup-$shaped with a minimum about 4 years post transplantation. Using the LRS, the other transitions seem to be exponentially distributed, i.e. the baseline hazard functions are constant over the durations. If the $1 \to 2$ transition had been exponentially distributed, the modeling of a Markov versus a semi-Markov process would have been more suitable. However, the semi-Markov property we describe here is interesting for several reasons. Firstly, even if the $2 \to 3$ and $2 \to 4$ transitions times are exponentially distributed, the resulting distribution of duration in state 2 is a mixture of exponential distributions. Secondly, again concerning this duration in state 2, we have identified significant interactions between panel reactive antibody, recipient gender and durations. Thus, the specific hazard functions of the $2 \to 3$ and $2 \to 4$ are not constant according to durations for patients with positive panel reactive antibody and for men,

respectively. Finally, in order to compute the probability of entering into an absorbing state according to chronological time since the transplantation, convolution products similar to expressions (8.6) and (8.7) use the density f_{12}, which is not constant over time regardless of the covariate values. Thus, the results obtained are globally different from those obtained using a Markov model.

The values of all the parameters are presented in Table 8.4 for the integral calculation made using 20-points Gauss Legendre quadrature. One can see that the supplementary number of points is not associated with a significant difference in terms of parameter estimation, compared to the 10-points approximation.

8.4 The goodness-of-fit test

8.4.1 Definition of the statistic

In SMM, the estimated transition probabilities depend on the sequence of observed states, on the time spent in states and on the covariate pattern. In section 3, the LRS is used to test the significance of parameters associated with duration distribution and of regression coefficients. Explanatory variables are also graphically examined. One of the assumptions of the modeling is the stationarity (8.1) regarding the chronological time, i.e. the time since the transplantation. The null hypothesis (H_0) assumes that the semi-Markov is stationary. The alternative hypothesis (H_1) assumes that the model is not stationary. To test if the stationarity is valid, the behavior of the model across chronological time should be examined. We propose a Pearson-type goodness-of-fit statistic, in which observations are grouped according to K categories of transitions and to L intervals of chronological times with boundaries $\{t_0, ..., t_l ..., t_L\}$.

In contrast to the $1 \rightarrow 2$ transitions, the times of occurrence of the final events k ($k = 3, 4$) are exactly known. This property of the data makes the classification of these transitions according to chronological time easier. In other words, classification of the $1 \rightarrow 2$ transitions according to chronological time is not possible, since all the durations of these transitions are interval-censored. We thus consider two types of observations ($K = 2$) : the jump to state 3 ($e \rightarrow 3$, with $e = 1, 2$) and the jump to state 4 ($e \rightarrow 4$, with $e = 1, 2$).

Concerning the classification of transitions according to chronological time, we define $L = 5$ intervals limited using the quantiles of the times of occurrence of a final event. These choices lead to balanced number of observed transitions.

Let $e_{l,k}$ and $o_{l,k}$ be respectively the expected and observed number of transitions in

cell (l, k). The Pearson type goodness-of-fit statistic, to examine the adequacy of the stationary SMM, is

$$G = \sum_{l=1}^{L} \sum_{k=3}^{4} (o_{l,k} - e_{l,k})^2 / e_{l,k} \tag{8.12}$$

This goodness-of-fit test bears some resemblances to the test for lifetables proposed by Lawless [89], which have been applied to testing the fit of the survival curve for a multistate model with a single absorbing state [114]. In these works, the statistic is based on the number of entries into the absorbing state in time periods conditional on an individual being in a transient state at the start of that period, meaning an individual contributes a series of binary observations (whether they survived or were absorbed in each period). In contrast the statistic in this paper categorises observations according to which absorbing state they entered and the time period, meaning an individual contributes a single multinomial observation.

Respecting the expression (8.7) and taking into account the right-censoring, the expected number of transitions into the state k $(k = 3, 4)$ between t_{l-1} and t_l $(l = 1, ..., L)$ equals

$$\begin{aligned} e_{l,k} &= \sum_{R(t_{l-1})} \Big\{ P_{1k}(\psi_{h,12}, \psi_{h,13}) \int_{t_{l-1}}^{min(v_{h,n_h}, t_l)} f_{1k}(x, z_{h,1k}) dx \\ &+ P_{12}(\psi_{h,12}, \psi_{h,13}) P_{2k}(\psi_{h,23}) \int_{t_{l-1}}^{min(v_{h,n_h}, t_l)} \int_0^x f_{12}(u, z_{h,12}) \\ &\times f_{2k}(x-u, z_{h,2k}) du dx \Big\} \end{aligned} \tag{8.13}$$

where v_{h,n_h} is the possible time of right-censoring of the subject h and $R(t_{l-1})$ represents the sum over all the individuals h which are not censored at the time t_{l-1} since the transplantation. The number of final events is obtained by counting their occurrence in an interval :

$$o_{l,k} = \sum_{R(t_{l-1})} I_{\{v_{h,n_h} \leq t_l \text{ and } W_{h,n_h} = k\}}$$

where $I_{\{a\}} = 1$ if the condition a is respected and 0 otherwise.

8.4.2 Distribution of the statistic

As written in section 4.1., the $1 \to 2$ transitions are not included in the statistic (8.12), since only the times of entering in absorbing states are exactly known. Thus, the proposed statistic does not include the right-censored contributions. For this reason, and regardless of the asymptotic circumstances, the approximation using a χ^2 distribution is not adequate. One method for estimating the distribution of the statistic (8.12) is

to generate B independent bootstrap samples from the parameters of the SMM under H_0 and to calculate the goodness-of-fit statistic for each sample. This semi-parametric bootstrap procedure can be divided into the following steps :

(i) - Generation of B bootstrap samples, constituted of n subjects. Each individual h^* ($h^* = 1, ..., n$) is observed during visits at the times since the transplantation $\{v^*_{h^*,0}, v^*_{h^*,1}, ..., v^*_{h^*,n_{h^*}}\}$. This vector corresponds to the observed times of visit in the initial sample. Since we do not simulate these times, the bootstrap is named semi-parametric. All patients begin in the state 1, $W^*_{h^*,0} = 1$.

(ii) - Simulation of the trajectory of each patient h^* from the estimated parameters of the SMM under H_0. From the estimated regression coefficients associated with the Markov chain, $\{\hat{\beta}_{1j}, j = 2, 3\}$, the second bootstrap state, $W^*_{h^*,1}$, is defined using the multinomial distribution with parameters $\{P_{1j}(\psi_{h^*,12}, \psi_{h^*,13}), j = 2, 3, 4\}$. If $W^*_{h^*,1} = 2$, then the value of $W^*_{h^*,2}$ is obtained using a binomial distribution with probabilities $\{P_{2j}(\psi_{h^*,23}), j = 3, 4\}$. Given these bootstrap sequence of states, the duration in state 1 for the subject h, $D^*_{h^*,0}$, follows the PDF $f_{1W^*_{h^*,1}}$, which takes into account the covariates $z_{h,1W^*_{h^*,1}}$. Thus, the inverse of the corresponding distribution function can be used to obtain a simulated duration from a random uniform variable $U(0, 1)$. Identically, if $W^*_{h^*,1} = 2$, then the duration in this can be simulated from $f_{2W^*_{h^*,2}}$.

(iii) - Estimation of the SMM for each bootstrap sample. The contributions are obtained as follows.

- If $d^*_{h^*,0} > v^*_{h^*,n_{h^*}}$, then this subject h^* is censored in state1 and her/his contribution is equivalent to (8.8), replacing v_{h,n_h} by $v^*_{h^*,n_{h^*}}$.
- Otherwise if $v^*_{h^*,r-1} < d^*_{h^*,0} \leq v^*_{h^*,r}$ ($r = 1, ..., n_{h^*}$) and $W^*_{h^*,1} = k$ ($k = 3, 4$), the contribution corresponds to (8.7), where v_{h,n_h} becomes $d^*_{h^*,0}$ and $d^0_{h,0}$ becomes $v^*_{h^*,r-1}$.
- Otherwise if $v^*_{h^*,r-1} < d^*_{h^*,0} \leq v^*_{h^*,r}$ ($r = 1, ..., n_{h^*}$), $W^*_{h^*,1} = 2$ and $d^*_{h^*,0} + d^*_{h^*,1} > v^*_{h^*,n_{h^*}}$, then the contribution equals (8.9), in which the values $\{d^0_{h,0}, d^1_{h,0}, v_{h,n_h}\}$ become $\{v^*_{h^*,r-1}, v^*_{h^*,r}, v^*_{h^*,n_{h^*}}\}$.
- Otherwise if $v^*_{h^*,r-1} < d^*_{h^*,0} \leq v^*_{h^*,r}$ ($r = 1, ..., n_{h^*}$), $W^*_{h^*,1} = 2$ and $d^*_{h^*,0} + d^*_{h^*,1} \leq v^*_{h^*,n_{h^*}}$, then the contribution is equivalent to (8.6), in which the values $\{d^0_{h,0}, d^1_{h,0}, v_{h,n_h}\}$ become $\{v^*_{h^*,r-1}, v^*_{h^*,r}, d^*_{h^*,0} + d^*_{h^*,1}\}$.

(iv) - Calculation of the distribution of the bootstrap statistics. For each bootstrap sample, the statistics G^*_b ($b = 1, ..., B$) is calculated, using the expressions (8.12) and (8.13). The p-value, i.e. the error probability of rejecting H_0 if this hypothesis is valid, equals $B^{-1} \sum_{b=1}^{B} I_{\{G^*_b \geq G\}}$, where $I_{\{G^*_b \geq G\}}$ equals 1 if the bootstrap statistics G^*_b is greater

than or equal to G, and 0 otherwise.

8.4.3 Application to data

The expected and observed counts of transitions into a final state are shown in the Contingency Table 8.5. The first two rows correspond to the observed and expected returns to dialysis and death between 0.011 and 0.689 years post transplantation. For these two types of events, the respective cells contribute 5.19% and 10.12% of the value of the goodness-of-fit statistic, with $G = 14.12$. The two cells in bold contribute 46.52% and correspond to the returns to dialysis. For reasons of computing time, 400 bootstrap samples were carried out. The quantiles of the cumulative probability function of the bootstrap statistic in $\{0.75, 0.50, 0.25, 0.10, 0.05, 0.01\}$ are respectively $\{9.79, 12.77, 15.26, 18.27, 20.95, 27.37\}$. 159 bootstrap statistics are greater than or equal to G, corresponding to a p-value equal to 0.3975. Thus, the fitted stationary SMM seems to be adequate to the kidney transplant data.

8.5 Conclusions

We have considered a flexible multistate model in order to analyze inherently complex longitudinal data. This model has been applied to the follow-up of kidney transplant recipients. The generalized Weibull PDF appears to be suitable. The use of a multinomial and multivariate logistic approach, in order to introduce explanatory factors in the Markov chain, also lead to a more precise modeling.

This parametric regression allows for the analysis of interval-censored data. More precisely, as is often the case in longitudinal data, the date of the final and absorbing events are exactly known (the death or the return to dialysis in our application) and only the transition times of the intermediate and transient states are interval censored. Between two consecutive visits, the information regarding a patient is unavailable. Using convolution products, the proposed estimation procedure deals with all the possible trajectories of a patient during this interval and can be easily adapted to other applications.

This model is fitted using standard likelihood estimation. The LRS can then be applied to the parameters selection strategy. This relates to the parsimony of the distribution functions and the selection of the significant covariates.

In order to determine whether the stationary assumption of the SMM is valid, we proposed the Pearson-type goodness-of-fit test. Because the theoretical distribution of

Table 8.4: *Comparison of the parameter estimations according to the number of points in the Gauss-Legendre quadrature*

Transition	Parameters	Estim. (10 points)	Estim. (20 points)
	Parameters of the baseline hazard functions		
$1 \to 2$	σ_{12}	68.80	68.75
$1 \to 2$	ν_{12}	0.77	0.05
$1 \to 2$	θ_{12}	0.25	0.25
$1 \to 3$	σ_{13}	42.84	42.76
$1 \to 4$	σ_{14}	109.83	109.87
$2 \to 3$	σ_{23}	12.66	12.66
$2 \to 4$	σ_{24}	6.54	6.83
	Regression parameters associated with the trajectories, $P_{ij}()$.		
$1 \to 2$	Intercept	1.87	1.87
$1 \to 2$	Delayed graft function	0.73	0.73
$1 \to 2$	Donor age	-2.03	-2.03
$1 \to 3$	Intercept	-2.81	-2.81
$2 \to 3$	Intercept	1.13	1.13
$2 \to 3$	Incompatibilities A+B+DR	0.93	0.93
	Regression parameters associated with the durations, $f_{ij}()$.		
$1 \to 2$	Induction treatment	0.36	0.36
$1 \to 2$	Recipient gender	-0.26	-0.26
$1 \to 2$	Donor age	0.96	0.96
$1 \to 3$	Cold ischemia time	5.02	5.02
$2 \to 3$	Incompatibilities A+B+DR	0.88	0.88
$2 \to 3$	Panel reactive antibody	1.09	1.09
$2 \to 3$	Panel reactive antibody $\times$ d [a]	-0.48	-0.47
$2 \to 4$	Delayed graft function	2.03	2.06
$2 \to 4$	Recipient gender	1.54	1.56
$2 \to 4$	Recipient gender $\times$ d [a]	-4.30	-4.30
$2 \to 4$	Recipient gender $\times$ d^2 [a]	1.30	1.30

[a] Time interaction with the duration d in the state

this statistic is intractable, a bootstrap algorithm was proposed. This test does not provide evidence of lack of fit. To sum up, this paper offers three extensions relating to the existing works in SMM : the approach for addressing covariates, the modeling of interval-censoring and the testing of stationarity.

A way to improve this statistic should take into account the $1 \rightarrow 2$ transition directly in the contingency table. In our application, all the transitions form state 1 into state 2 are interval-censored with different intervals for each subject. Thus, the classification of the number of $1 \rightarrow 2$ transitions in some intervals of time since the transplantation is difficult. However, even if the statistic is based on the number of final failures, the $1 \rightarrow 2$ transition enters into the computation of the expected number of final events (see equation 8.13).

As we mentioned previously, the date of entry into an absorbing state, v_{h,n_h}, is exactly known. However, the estimation of the model can be adapted if this date is interval censored. For example, we can reconsider the individual likelihood contribution of a subject h respecting the trajectory *(i)*, but where the date of entry into the state $W_{h,2} = k$ occurs between v^0_{h,n_h} and v^1_{h,n_h}. In this case, the equation (8.6) can be extend as follows :

$$P_{12}P_{2k}\int_{d^0_{h,0}}^{d^1_{h,0}}\int_{v^0_{h,n_h}}^{v^1_{h,n_h}} f_{12}(u)f_{2k}(w-u)dwdu$$

One can see that this expression is equal to (8.6) when the difference between v^0_{h,n_h} and v^1_{h,n_h} tends to be null. The double integral calculations are always feasible using the Gauss-Legendre quadrature. For a higher dimension of integrals, for instance if three consecutive transitions were interval censored for a given subject (2 transient and 1 absorbing states), this numerical approximation is no longer feasible since the computation time increases too much and since removing singularities becomes harder.

The main limit is the assumption of a unidirectional model. Indeed, if returns are included (i.e. the transition from State 2 to State 1). The recent paper of Kang and Lagakos [79] considers the case where backward transitions are permitted. This excellent work constitutes a way of extending our proposed SMM, since the authors do not include covariate effects in their modelling.

Acknowledgments

We thank the Roche Laboratory for its support of the DIVAT data bank and Dr J. Ashton-Chess for editing the manuscript.

Table 8.5: *Contingency table for the observed and expected counts of final events.*

Chronological time (in years)		Transition $e \to 3$	Transition $e \to 4$	Percentage $e \to 3$	Percentage $e \to 4$
$]0.011; 0.689]$	Observed	12	8	5.19%	10.12%
	Expected	9.38	5.25		
$]0.689; 2.168]$	Observed	13	8	**21.21%**	2.89%
	Expected	20.91	10.02		
$]2.168; 3.826]$	Observed	16	5	15.73%	11.73%
	Expected	23.17	8.81		
$]3.826; 5.213]$	Observed	17	4	2.14%	4.45%
	Expected	14.87	6.18		
$]5.213; 9.158]$	Observed	14	7	**25.31%**	0.24%
	Expected	23.08	7.51		

Appendix of the paper by Foucher et al. (SMMR, 2010)

Trajectory (i) - The contribution is the joint probability that the subject h jumped into state 2 after a duration in state 1 between $d^0_{h,0}$ and $d^1_{h,0}$ and that she/he returned to dialysis ($k = 3$) or died ($k = 4$) at time v_{h,n_h} post transplantation.

$$
\begin{aligned}
C_{h,1} &= \lim_{\Delta d \to 0^+} \Big\{ P(v_{h,n_h} - d_{h,0} < d_{h,1} < v_{h,n_h} - d_{h,0} + \Delta d, W_{h,2} = k, \\
&\quad d^0_{h,0} < d_{h,0} < d^1_{h,0}, W_{h,1} = 2 | W_{h,0} = 1) / \Delta d \Big\} \\
&= \lim_{\Delta d \to 0^+} \Big\{ P(v_{h,n_h} - d_{h,0} < d_{h,1} < v_{h,n_h} - d_{h,0} + \Delta d, W_{h,2} = k | \\
&\quad d^0_{h,0} < d_{h,0} < d^1_{h,0}, W_{h,1} = 2, W_{h,0} = 1) \\
&\times \; P(d^0_{h,0} < d_{h,0} < d^1_{h,0}, W_{h,1} = 2 | W_{h,0} = 1) / \Delta d \Big\} \\
&= \lim_{\Delta d \to 0^+} \Big\{ P(W_{h,2} = k | W_{h,1} = 2) P(v_{h,n_h} - d_{h,0} < d_{h,1} < v_{h,n_h} - d_{h,0} \\
&+ \; \Delta d | W_{h,2} = k, W_{h,1} = 2) / \Delta d \Big\} \\
&\times \; P(W_{h,1} = 2 | W_{h,0} = 1) P(d^0_{0,h} < d_{h,0} < d^1_{h,0} | W_{h,1} = 2, W_{h,0} = 1) \\
&= P_{12} P_{2k} \int_{d^0_{h,0}}^{d^1_{h,0}} f_{12}(u) f_{2k}(v_{h,n_h} - u) du
\end{aligned}
$$

Trajectory (ii) - Because the state 2 is not observed, we must take into account that the individual transited directly between state 1 and state k at time v_{h,n_h} after the transplantation or that her/he had two consecutive transitions $1 \to 2 \to k$ in the interval

$]d^0_{h,0}, v_{h,n_h}]$. Respecting the last development, we obtain :

$$\begin{aligned} C_{h,2} &= \lim_{\Delta d \to 0^+} \Big\{ P(v_{h,n_h} < d_{h,0} < v_{h,n_h} + \Delta d, W_{h,1} = k | W_{h,0} = 1)/\Delta d \Big\} \\ &+ \lim_{\Delta d \to 0^+} \Big\{ P(v_{h,n_h} - d_{h,0} < d_{h,1} < v_{h,n_h} - d_{h,0} + \Delta d, W_{h,2} = k, \\ &\qquad d^0_{h,0} < d_{h,0} < v_{h,n_h}, W_{h,1} = 2 | W_{h,0} = 1)/\Delta d \Big\} \\ &= P_{1k} f_{1k}(v_{h,n_h}) + P_{12} P_{2k} \int_{d^0_{h,0}}^{v_{h,n_h}} f_{12}(u) f_{2k}(v_{h,n_h} - u) du \end{aligned}$$

Trajectory (iii) - The individual h is right-censored in state 1 after a time v_{h,n_h} post transplantation.

$$\begin{aligned} C_{h,3} &= P(d_{h,0} > v_{h,n_h} | W_{h,0} = 1) \\ &= \sum_{j=2}^{4} P(W_{h,1} = j | W_{h,0} = 1) P(d_{h,0} > v_{h,n_h} | W_{h,1} = j, W_{h,0} = 1) \\ &= \sum_{j=2}^{4} P_{1j} S_{1j}(v_{h,n_h}) \end{aligned}$$

Trajectory (iv) - The individual h is right-censored in state 2 after a time v_{h,n_h} post transplantation and her/he jumped into state 2 after a duration in state 1 between $d^0_{h,0}$ and $d^1_{h,0}$.

$$\begin{aligned} C_{h,4} &= P(d_{h,1} > v_{h,n_h} - d_{h,0}, W_{h,1} = 2, d^0_{h,0} < d_{h,0} < d^1_{h,0} | W_{h,0} = 1) \\ &= P(d_{h,1} > v_{h,n_h} - d_{h,0} | W_{h,1} = 2, d^0_{h,0} < d_{h,0} < d^1_{h,0}, W_{h,0} = 1) \\ &\times P(d^0_{h,0} < d_{h,0} < d^1_{h,0}, W_{h,1} = 2 | W_{h,0} = 1) \\ &= P(W_{h,1} = 2 | W_{h,0} = 1) P(d^0_{h,0} < d_{h,0} < d^1_{h,0} | W_{h,1} = 2, W_{h,0} = 1) \\ &\times \sum_{j=3}^{4} P(W_{h,2} = j | W_{h,1} = 2) P(d_{h,1} > v_{h,n_h} - d_{h,0} | W_{h,2} = j, W_{h,1} = 2) \\ &= P_{12} \int_{d^0_{h,0}}^{d^1_{h,0}} f_{12}(u) \Big\{ \sum_{j=3}^{4} P_{2j} S_{2j}(v_{h,n_h} - u) \Big\} du \end{aligned}$$

Troisième partie

Les projets de recherche

- Chapitre 9 -

The prognostic in renal transplantation : an overview of the projects

Determining early surrogate markers of long-term graft outcome is important for optimal medical management. Non-invasive (blood or urine) biomarkers have been proposed for the prediction of post-transplant clinical event. However, these biomarkers still need further validations in large patient cohorts and the evaluation of their prognostic capacity is often based on a simple correlation between the marker and the time-to-event [42].

Currently, the 6- and 12-month post-transplant serum creatinine level is considered to be the simplest marker that is significantly correlated with graft survival. Again, this correlation have to be interpreted with caution : although 6–12-month serum creatinine level correlates with graft loss, this marker has been shown to be poorly predictive.

According to this need of useful prognostic indicators, three risk scores of graft failure have been published from different groups in the last 2 years [82, 101, 43], including our KTFS. However, all these articles involve several avenues for improvements, particularly regarding the methodology. We obtained two grants to support these developments : from the French National Agency of Research (ANR-11-JSV1-0008-01) and from the French Ministry of Health (PHRC, PROG/11/85, 2011).

According to these grants, with the additional support from the CENTAURE foundation, we thus have organized a working group around this general objective in renal transplantation. Magali Giral (Professor in Nephrology) is in charge of the clinical dimension of the group. She also coordinates the DIVAT network. She is the thesis director of all the PhD students. Etienne Dantan (Assistant Professor in Biostatistics) was recruited two years ago in the SPHERE laboratory. Philippe Tessier (Assistant Professor in Health Economy) is specialized in medical decision making. Marine Lorent, Katy

Trébern-Launay and Florence Gillaizeau are PhD students in this group.

The following paragraphs present the main projects as a summary. Some of them are detailed in the next chapters.

9.1 The DGFS : a very simple score to predict delayed graft function

Within the past decade, even though the frequency of acute rejection episodes dramatically decreased under modern immunosuppressive regimen, the incidence and severity of delayed graft function (DGF) remained stable. The DGF is defined as the need for dialysis within the first seven days after the renal transplantation. DGF is known to impair 1-year post transplantation renal function and to decrease long-term graft survival. Consequences of DGF are also economic with more prolonged hospitalization, increase cost of patient management, need for dialysis, diagnostic radiology, biopsies and closer immunosuppressive drug monitoring [158, 20].

Therefore, predicting patient DGF appears important in kidney transplantation management in order to propose at least new preventive drugs in patients with high risk of DGF or optimally to decrease the overall DGF prevalence by a better graft allocation. Several DGF scoring system have actually been proposed within the last few years, especially by Irish et al. [71]. In this study, the authors refined their first model by using additional donor and recipient characteristics known at the time of transplantation. By using the United State Renal Data System (USRDS) registry, authors proposed a score calculated at the time of transplantation. The predictive capacity of the model was validated by using an independent data set associated with an area under the ROC curve at 0.70 (no confidence interval available). As usual, regardless the high quality of the methodology, this score is associated with several limitations :

- It was developed from north-American recipients, while the immunosuppressive therapy (dosage, induction, etc.), the definition of adverse events and the patients' profiles differ substantially with European patients. It can be therefore inadequate to use this score in Europe.
- Patients were included between 2003 and 2006, which may not be representative of recent transplantations.
- It is based on high number of variables (eighteen), without decision threshold aiming to classify patients according to their risk of DGF.

For all of these reasons, we thus proposed to develop a complementary DGF risk score, called DGFS, from the recipients transplanted between 2007 and 2012 in the DIVAT network, receiving Tacrolimus and Mycophenolate mofetil as maintenance therapy.

Adult kidney transplant recipients in the DIVAT data bank were included in the study. Patients who received a kidney from non heart-beating donors, pre-emptive transplantation, use of pulsatile perfusion machine for graft's preservation, under peritoneal dialysis or who displayed less than 7 days of patient-and-graft survival were excluded from the study. Patients with at least one missing values on the following variables considered as major risk factors of DGF in literature : cold ischemia time, recipient's gender, HLA incompatibilities and PRA. Patients with missing value on DGF were also removed from the analysis. 1 844 adult renal transplant recipients were included in the study, among whom 468 (25.4%) had a DGF.

These patients were randomly assigned into two samples in a proportion of two-thirds for learning (N=1238) and one-third for validation (N=606). The construction of the scoring system was done based on the learning sample. A logistic regression model was performed to study the probability of DGF. Univariate analyses were performed for a first selection of explicative variables (p-value < 0.20). The covariates were further analysed in a multivariate model and retained if their p-value was lower than 0.05 in a backward selection. Finally, the last significant variables were removed if the corresponding decrease of the area under ROC curve was lower than 1%. The scoring system was the linear predictor of this final model. For a better interpretation, we normalized this score by subtracting the mean and dividing by the standard deviation in order to obtain the DGFS.

The previous model was validated with the independent data set and the corresponding ROC curve was estimated for validation. 95% confidence intervals were non-parametrically obtained by bootstrap resampling (1000 iterations). The calibration of the model was also evaluated, i.e. the concordance between the observed probabilities of DGF and the expected ones. As usual, 10 intervals of the DGFS were used for the corresponding plot and Hosmer-Lemeshow statistic.

For this project, we also have performed an alternative strategy by adapting the bootstrap 0.632+ algorithm for the estimation of the area under the ROC curve with complete data (no censoring or truncation, see chapter 4 for more details about this approach). We used a logistic regression with *lasso* penalization. Regarding the sample size and the number of explicative variables, no over-fitting was expected. We validated this expectation : the traditional approach (division of the sample) offered similar results than the bootstrap 0.632+. We thus chose the traditional approach in order to present a more simple methodology for a clinical publication of the results. All the analyses were performed by using R software version 2.15 [124]. In order to compute the DGFS and to interpret the resulting value, we proposed an application available on smartphones, tablets or computers at `www.divat.fr/en/online-calculators/dgfs`.

Only five covariates were finally retained in the scoring system associated with the risk of DGF. The area under the ROC curve was 0.73, closed to the value at 0.70 found by Irish et al. These five variables are easily available at the time of transplantation. The corresponding article of this work is in process and we study the possibility of a patent with the support of Nantes University Hospital. The first part of the statistical work was performed by Florent LeBorgne (Master 2 in Biostatistics) and the writing of the paper is under the responsibility of Marion Chapal (resident).

9.2 A multistate model to investigate the relationship between biomarkers and kidney transplant recipients' outcomes (PhD student : F. Gillaizeau).

The angiotensin II type 1 receptor (AT1R) is an emerging target of functional non-HLA antibodies. We examined the potential of pre-sensitization against AT1R as a risk factor for long-term graft survival and acute rejection episodes (ARE).

The study included 599 kidney transplanted patients between 1998 and 2007. Serum samples were analysed in a blinded fashion for anti-AT1R antibodies (AT1R-Abs) using a quantitative solid-phase assay. A threshold of AT1R-Abs levels was statistically determined at 10 units based on the time to graft failure (death censored) by using the methodology proposed by Hothorn and Zeileis [69]. Two extended Cox models with time-dependent regression coefficients were used to determine the risk factors for time to graft failure (death censored) and time to first ARE. AT1R-Abs $>$ 10 units were detected in 283 patients (47.2%) before transplantation. Patients who had a level of AT1R-Abs $>$ 10 units had a 2.6 higher risk of graft failure from 3 years post transplantation onwards (p=0.0003) and a 1.9 fold higher risk of experiencing ARE within the first 4 months of transplantation (p=0.0373). Thus, these traditional analyses illustrated that pre-transplant AT1R-Abs constitutes an independent risk factor for long-term graft loss in association with a higher risk of early ARE. The corresponding paper is in revision in the American Journal of Transplantation.

Nevertheless, a question persists : is the higher risk of graft failure for patients with AT1R-Abs $>$ 10 units due to the higher frequency of ARE ? We developed a multistate model to assess the relationship between AT1R-Abs and the transplantation outcome. Outcomes were the transition probabilities between 4 states : graft without any acute ARE, graft with at least one ARE, return to dialysis and patient death (figure 9.1). We used a parametric Semi-Markov model with generalized Weibull distribution of the waiting times [49]. This model has few advantages : ARE represents an outcome but

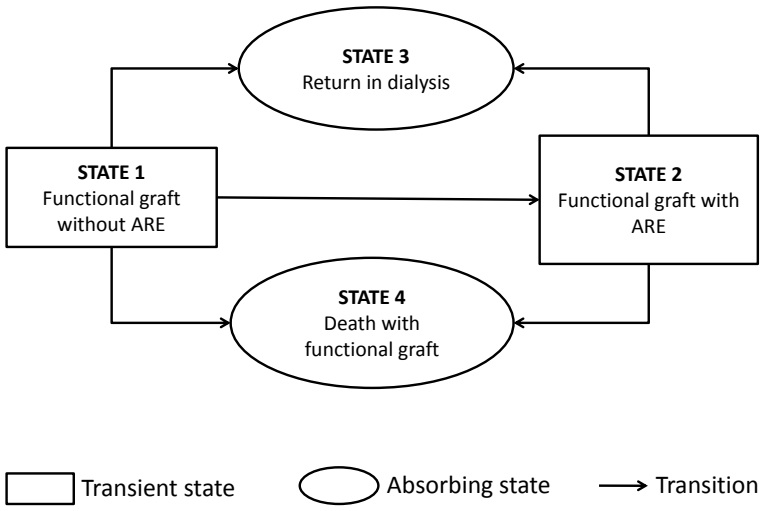

Figure 9.1: *Multistate model to investigate the relationship between AT1R and kidney transplant recipients outcomes*

also a risk factor of return in dialysis, which is in competition with death (avoiding the assumption of non-informative censoring). Here are the preliminary results :

1. It confirms that patients with AT1R-Abs > 10 units had a significant increase risk of ARE.
2. Among patients with ARE, it seems that these acute events are delayed for patients with AT1R-Abs > 10 units. It can suggest that immunization related to AT1R-Abs is associated with non-usual delayed ARE (the majority of ARE occurs usually in the first post transplantation year). Nevertheless, after investigations, these delayed rejection seem to be related to non-compliance.
3. Among patients with ARE, patients with AT1R-Abs > 10 units seem to be more susceptible for returning in dialysis compared with patients with AT1R-Abs ≤ 10 unit.
4. After 3 years post transplantation and among patients returning in dialysis, AT1R-Abs > 10 units at transplantation seems associated with earlier failures. It may suggest a higher propensity of subclinical ARE for patients with AT1R-Abs > 10 units or chronic rejection process.

All these results confirm that pre-graft AT1R-Abs is a risk factor of return in dialysis. A part of this association is explained by the correlation between pre-graft AT1R-Abs and the higher frequency of ARE. Nevertheless, this multistate model also demonstrated an independent role of the pre-graft AT1R-Abs level which may be associated with an embedded immunization.

This work will constitute the basis to communicate the interest of such multistate models in biomarkers' studies. The results are more informative compared to traditional approaches. In clinical epidemiology, such modelling is now accepted, but efforts are still needed in translational research. We just begin a collaborative study with the Pr. Siamak Bahram, the project leader of Transplantex. Transplantex has been selected as a 'Laboratory of Excellence' within the framework of the French government's 'Investissements d'Avenir'. The objective is to identify novel histocompatibility targets and transplantation biomarkers. DIVAT and the associated biocollection constitute the materials of this project. We think that the developments performed in the AT1R study will also be relevant to this new project. Florence Gillaizeau, PhD student, is in charge of these projects involving multistate models.

9.3 The concept of net survival

The current changes in the demographic profile of patients with End Stage Renal Disease (ESRD) awaiting a transplantation, such as increasing age and presence of serious comorbidities, would be expected to increase mortality. Instead, the post-transplant mortality rate is still stable. One major hypothesis to explain this decrease in overall mortality rate (including all causes of death) is a specific reduction in mortality related to the transplantation status. This recipient-related mortality may be due to the addition of transplant-inherent conditions to pre-existing classical risk factors in ESRD patients. Understanding the factors associated with recipient-related mortality would benefit the medical management of ESRD.

However, most studies in kidney transplantation are based on overall mortality, which obviously limits their interpretation. For example, recipient age is presented as the greatest risk factor for mortality, but this could mainly be due to ageing, which is also observed in the general population. The analysis of this related mortality is challenging due to the difficulty in establishing the exact cause of death in transplant recipients. For example, it is impossible to ascertain whether or not a suicide or a cancer is related to the disease and treatments. Nevertheless, the relative survival method estimates the specific mortality excess for a disease in comparison with the non-specific mortality from a matched general population. This methodology has been widely applied to cancer registries and has improved the understanding of cancer-related mortality.

Two cohorts were included for this study : the DIVAT database with 3641 patients and the UNOS database with 8291 patients (USA, `www.unos.org`). The excess mortality related to kidney transplant recipients was assessed by subtracting the expected mortality of the general population from the overall mortality observed in both cohorts.

Expected mortality for both cohorts were determined from the lifetime tables of each country proposed by the human mortality database (HMD, `www.mortality.org`). For each recipient, this subtraction was performed according to gender, age and year of transplantation. Risk factors were evaluated by the corresponding hazard ratios using the additive model as proposed by Estève [39]. The covariate selection procedure was similar to the overall mortality analysis. The proportionality of hazards was also evaluated [141]. The relative survival model was estimated using the `relsurv` R-package using the expectation-maximization algorithm for parameters estimation.

As expected, the demographic parameters were different between both cohorts. In the U.S. cohort, the recipients were younger, heavier and more diabetic at transplantation. The U.S. patients also received better matched kidneys, from younger donors and with a shorter cold ischemia time. The overall 10-year cumulative mortality was also different, 13% for the DIVAT cohort and 27% for the UNOS cohort.

By removing the expected mortality, we demonstrated that recipient age, which was reported as being a key for predicting overall mortality, was no longer a significant risk factor for excess mortality in the DIVAT-based analysis. Recipient age remained a significant risk factor in the UNOS-based analysis (United Network for Organ Sharing), but the corresponding hazard ratios were lower for the analysis of the mortality related to transplantation status compared to the analysis of overall mortality. This difference may be due to the smaller sample size of the DIVAT cohort (less statistical power) or the differences in comorbidities between both cohorts with more diabetic and overweight recipients in the U.S. cohort that are not taken into account in the expected mortality analysis. Therefore, we cannot exclude the fact that the mortality observed in older recipients is higher than that observed in the general population.

The other factor taken into account in the lifetime tables was recipient gender. In terms of overall mortality, gender was not significantly correlated in the UNOS cohort ($p=0.2318$), whereas it was an independent risk factor in the DIVAT cohort ($p=0.0005$). Nevertheless, gender was not found to be a significant risk factor for recipient-related mortality in either cohort. This difference was due to the higher mortality rates of men compared to women in France. Therefore, the relative survival approach held the advantage of being able to obtain more comparable results for the international medical community by removing the background mortality of each country.

Our study highlights the fact that the relationships between variables such as recipient age and gender, and the excess recipient-related mortality were overestimated when analysing the overall mortality with the usual Cox model. The project was initiated during the post-doctoral position of the Dr. Ahmed Akl. The main limitation is the choice of the general population as reference population. Another relevant refe-

rence population is patients on waiting list. This model will then be used as the expected mortality in the analysis of transplant-related mortality. Few methodological issues need first to be solved (see following sections from 9.4 to 9.7).

9.4 An original additive model

Suppose a sample of N kidney transplant recipients. Let d_i be the post transplantation time until death with functioning graft of the ith recipient ($i = 1, ..., N$). Moreover, let t_i be the time between the registration on waiting list and the transplantation. The observed mortality hazard for this recipient, $\lambda_O(d_i|z_i)$, can be written as follow :

$$\lambda_O(d_i|z_i) = \lambda_P^*(t_i + d_i|z_{P,i}) + \lambda_E(d_i|z_{E,i}) \tag{9.1}$$

where $\lambda_P^*(t_i + d_i|z_{P,i})$ is the expected hazard function on waiting list at post transplantation time d_i if the recipient was not transplanted. $\lambda_E(d_i|z_{E,i})$ is the excess of hazard related to the transplantation at time d_i post-transplantation. $z_{P,i} \in z_i$ and $z_{E,i} \in z_i$ represents the respective explicative variables of both hazards. The problem in renal transplantation is that the observed hazard may be lower than the expected hazard, in particular after few months post transplantation [156]. Thus, we are proposing the following additive model into two parts :

$$\lambda_O(d_i|z_i) = \lambda_P^*(t_i + d_i|z_{P,i}) + \lambda_{E1}(d_i|z_{E1,i})I(d_i \leq \tau) - \lambda_{E2}(d_i|z_{E2,i})I(d_i > \tau) \tag{9.2}$$

where $\lambda_{E1}()$ and $\lambda_{E2}()$ represent respectively the excess and the default of mortality of transplanted patients compared to patients under dialysis. By assuming the proportionality of hazards, z_{E1} and z_{E2} will constitute the scoring systems for 1) stratifying the patients according to their risk of mortality due to transplantation before time τ and 2) stratifying the patients according to the benefit of the transplantation in term of life expectancy after this acute period. Therefore, regarding to the equation (9.2), the first step is to estimate $\lambda_P^*(t_i + d_i|z_{P,i})$, i.e. the mortality and the risk factors of mortality of patients on waiting list.

9.5 Modeling the mortality on waiting list (PhD student : K. Trébern-Launay)

We estimate this mortality by using the REIN registry. The baseline is the time of the inscription on waiting list for transplantation. All the explicative variables are collected

at this baseline. Suppose a sample of N kidney transplant recipients ($j = 1, ..., N$). The estimation of the distribution of the time between the registration on waiting list and the death have to be taken into account. We have to complete this description regarding the informative competition between transplantation and death on waiting list : the healthy patients have more chance to be transplanted. Let $K = 1$ if the the time t_j corresponds to a transplantation and $K = 2$ for a death. The hazard function of having the event k at time t_j is :

$$\lambda_k(t_j|z_{P,j}) = \lim_{\Delta t \to 0} P(t_j < T < t_j + \Delta t, K = k | T > t_j, z_{P,j}) \Big/ \Delta t \tag{9.3}$$

Respecting our last developments with semi-Markov assumption [45, 46, 47, 49], one can divide the equation (9.3) into two parts :

$$\lambda_k(t_j|z_{P,j}) = P_k(z_{P1,j})\lambda'_k(t_j|z_{P2,j}) \tag{9.4}$$

where $P_k(z_{P1,j})$ is the embedded Markov chain depending of a subset of explicative variables $z_{P1,j} \in z_{P,j}$:

$$P_k(z_{P1,j}) = P(K = k|z_{P1,j}) \quad \text{with} \quad \sum_{k=1}^{2} P_k(z_{P1,j}) = 1 \tag{9.5}$$

and $\lambda'_k(t_j|z_{P2,j})$ is the hazard function of the waiting time before the event, given that this event is k and according to a subset of explicative variables $z_{P2k,j} \in z_{P,j}$:

$$\lambda'_k(t_j|z_{P2,j}) = \lim_{\Delta t \to 0} P(t_j < T < t_j + \Delta t | K = k, T > t, z_{P2k,j}) \Big/ \Delta t \tag{9.6}$$

$z_{P1,j}$ can be included assuming a logistic regression and $z_{P2,j}$ can be modeled assuming the proportionality of hazards. These parameters can be estimated by likelihood maximization (see equation 7.1). Preliminary results based on four different French regions (Pays de Loire, Midi-Pyrénées, Languedoc-Roussillon, Limousin) illustrates the adequacy of this model (generalized Weibull distribution of the waiting times). We are waiting the data from Lorraine in order to finalize the results.

The additional interest of this model is also to allow the construction of life tables as those available for traditional relative survival analyses (Human Mortality Database, HMD, for instance). The quantity of interest in such table is $q(m|z_P)$, the probability of dying at month m (before the $m + 1$th month) post registration, for patients with z_P.

If the time unit is in days, then :

$$\begin{aligned} q(m|z_P) &= P(m < T < m+1, K = 2|T > m, z_P) \\ &= P(t < T < t+30, K = 2|z_P)/P(T > t|z_P) \\ &= P(K = 2|z_{P1})P(t < T < t+30|K = 2, z_{P2k})/P(T > t|z_P) \end{aligned}$$

Thus,

$$q(m|z_P) = P_2(z_{P1})\frac{S_2'(t|z_{P2k}) - S_2'(t+30|z_{P2k})}{\sum_{k=1}^{2} P_k(z_{P1})S_k'(t|z_{P2k})}$$

where $S_k'(t|x) = \exp(\int_0^t \lambda'(u|x)du)$.

9.6 The accuracy of a marker to predict disease related mortality (PhD student : M. Lorent)

Developing prognostic markers of mortality for patients with chronic disease is important for identifying subjects at high risk of death and for optimizing medical management. The usual approach to evaluate the accuracy of such marker is the time-dependent ROC curve, which is well-adapted for censored data. Nevertheless, as we explained previously, an important part of the mortality may not be due to the chronic disease and it is often impossible to individually determine whether or not the deaths are related to the disease itself. The solution is to distinguish between the expected mortality of a reference population and the excess mortality related to the disease, by using an additive relative survival model.

We thus are developing a new estimator of time-dependent ROC curves that includes this concept of net survival. The objective is to evaluate the capacity of a marker to predict disease-specific mortality. Simulations are performed in order to validate this estimator. We also illustrate this method with two different applications : 1) predicting mortality related to primary biliary cirrhosis of the liver (a well known available data set often used by researchers in Biostatistics) and 2) predicting mortality related to kidney transplantation in end-stage renal disease patients. For each application, a scoring system already established is evaluated.

This project constitutes the first part of the PhD position of Marine Lorent. This methodology is fully explained in the chapter 10 and only concerns the traditional context with life tables from the general population. The integration of the mortality estimated from patients on waiting list will be thereafter included. But an important issue is that this expected mortality estimated from REIN registry (section 9.5) cannot be considered as fixed parameters, in opposition when life tables are available.

9.7 How to deal with the absence of life tables (PhD student : K. Trébern-Launay)

Relative survival models are traditionally used for the evaluation of mortality related to chronic diseases, taking into account the expected mortality of the general population (life table by gender, calendar year and age). Therefore, a problem appears when the expected hazard was not obtained from life tables, i.e. it cannot be assumed as fixed parameters. This situation can appear when the endpoint is not the time-to-death or when the reference population is not the general population. We described this issue previously, but we also have this problem in the analysis of the graft failure (return-to-dialysis or patient death) between second and first kidney transplant recipients. In this project, we are proposing an adaptation of the multiplicative-regression model for relative survival to study the heterogeneity of risk factors between two groups of patients. This project constitutes the second part of the PhD position of Katy Trébern-Launay.

As a solution, estimation of the parameters and the corresponding standard deviations are based on partial likelihood maximization and Monte-Carlo simulations and bootstrap re-sampling (details in the chapter 11). The results are validated by using an adaptation of stratified Cox model. While the traditional Cox model does not provide original results based on the renal transplantation literature, the proposed relative model reveals new perspectives that are useful for clinicians. This methodology may be of interest in other medical fields when the principal objective is the comparison of risk factors between two populations. This methodology will also be useful for the previous additive model in which the time-to-death distribution is estimated by using a sample from the REIN network.

- Chapitre 10 -

Net time-dependent ROC curves : a solution for evaluating the accuracy of a marker to predict disease-related mortality

This work constitutes the first part of the thesis of Marine Lorent. Her developments have been submitted in *statistics in medicine*.

10.1 Introduction

Determining early markers of patient survival is essential for optimal healthcare management. For instance, in kidney transplantation, some papers have focussed on the prediction of recipient survival after renal transplantation, such as the paper by Hernández and others [66, 67]. A new therapeutic era has increased long-term graft survival, leading to a rise in the proportion of deaths unrelated to the kidney transplant status itself. Moreover, the optimizing medical management of transplant recipients should decrease mortality specifically related to this status. In addition, the relationship between new markers or therapies and the long-term mortality related to Kidney Transplant Recipient status (KTR) are currently impossible to demonstrate with cause-specific approaches, given the impossibility to identify deaths specifically related to the transplantation status. For example, the occurrence of a cancer that leads to recipient death may or may not be the consequence of immunosuppressive drug exposure [18].

One solution is to distinguish the expected mortality of the general population and the excess mortality associated to the disease. The development of these relative survival models has been a subject of long-standing interest in the analysis of data from cancer registries [33, 17, 60, 29]. The main objective of such models is to estimate the net sur-

vival, i.e. the survival when the only possible deaths are related to the disease. Recently, Pohar et al. proposed a comprehensive methodology for such analyse [118, 119, 120, 115]. The methods for evaluating the excess mortality related to the disease are thus becoming well established in the statistical literature. Nevertheless, little attention has been paid to the development of methods for evaluating the accuracy of a marker to predict the deaths related to the disease, while it appears to be useful for medical decision making and complementary with the prediction of the all-cause mortality.

In diagnostic medicine, the standard criteria used to appraise the predictive capacity of a marker are its sensitivity and the specificity. Thanks to the initial work of Heagerty and others [63], these criteria can be estimated for a long-term prognostic marker with incomplete data. This time-dependent ROC theory has been extended to the competing risk framework by Saha and Heagerty [132] and Foucher and others [43]. A particular competing risk model can be used when a distinction is feasible between deaths related to the disease and those that are not. Again, because it is often impossible to perform such differentiation in chronic disease analyse, such competing risk approaches do not appear to have been adapted.

The aim of this paper is to propose an estimator of time-dependent ROC curves that includes the concept of net survival. We describe this approach in Section 2. We illustrate its utility in Section 3, via simulation studies, and in Sections 4 and 5 via two different applications : patients with primary biliary cirrhosis of the liver and recipients of a kidney transplant.

10.2 Methods

10.2.1 Estimation of the cause-specific hazard

Let X be the random variable representing the prognostic marker and x_j the associated observation for the subject j ($j = 1, ..., n$). n is the sample size. Let T_j be the time of the death of the subject j, with $T_j = \min(T_{Ej}, T_{Pj})$. T_{Ej} corresponds to the time to death related to the disease and T_{Pj} corresponds to the time to expected death as in the general population. We define C_j as the time of the last follow-up point (right censoring). The cumulative hazard function of T_{Pj} at time t, $\Lambda_{Pj}(t)$, is obtained from life tables, respecting age, calendar year and gender of the subject j. $S_{Pj}(t) = exp(-\Lambda_{Pj}(t))$ is the corresponding survival function. In order to estimate the distribution of T_E, Pohar and others proposed to weight the at-risk process and the counting process by the expected survival probability of each subject [115]. More precisely, the number of at-risk subjects at time t equals $Y(t) = \sum_{j=1}^{n} Y_j(t)$ with $Y_j(t) = I(T_j > t, C_j > t)/S_{Pj}(t)$.

The function $I(a)$ equals 1 if a is true and 0 otherwise. Moreover, the number of deaths (regardless of the cause) observed before t equals $N(t) = \sum_{j=1}^{n} N_j(t)$ where $N_j(t) = I(T_j \leq t, C_j \geq T_j)/S_{Pj}(t)$. A consistent estimator of the cumulative excess hazard of T_E is defined by

$$\hat{\Lambda}_E(t) = \int_0^t \frac{dN(u)}{Y(u)} - \int_0^t \frac{\sum_{j=1}^{n} Y_j(u) d\Lambda_{Pj}(u)}{Y(u)} \tag{10.1}$$

10.2.2 Definition of the net time-dependent ROC curve

The aim is to evaluate the capacity of X to predict the deaths related to the disease up to time τ. By convention, we will assume an increasing disease-related mortality with the value of X. By defining a binary test at the cut-off c, the net sensitivity represents the probability of observing $X > c$ given that the disease-related death occurs before time τ. The net specificity represents the probability of observing $X \leq c$ given that the disease-related death occurs after time τ, i.e. $se_\tau(c) = Pr(X > c | T_E \leq \tau)$ and $sp_\tau(c) = Pr(X \leq c | T_E > \tau)$. By adapting the approach of Heagerty and others [63], these two probabilities can be developed as follows :

$$se_\tau(c) = \{(1 - G_X(c)) - S_{X,E}(c, \tau)\}/\{1 - S_{X,E}(-\infty, \tau)\} \tag{10.2}$$

$$sp_\tau(c) = 1 - \{S_{X,E}(c, \tau)/S_{X,E}(-\infty, \tau)\} \tag{10.3}$$

where $G_X(a) = Pr(X < a)$ is the distribution function of X and $S_{X,E}(a, b) = Pr(X > a, T_E > b)$ is the bivariate survival function of X and T_E. $G_X()$ can be estimated by using the empirical distribution function

$$\hat{G}_X(c) = n^{-1} \sum_{j=1}^{n} I(x_j < c) \tag{10.4}$$

and $S_{X,E}()$ by using the Akritas estimator [4] :

$$\hat{S}_{X,E}(c, \tau) = n^{-1} \sum_{j=1}^{n} exp(-\hat{\Lambda}_E(\tau | X = x_j)) I(x_j > c) \tag{10.5}$$

The estimation of the conditional cumulative excess hazard can be obtained similarly to (10.1) with the Akritas estimator. Let $Y_{jl}^{\pi}(t) = I(T_l > t, C_l > t, |\hat{G}_X(x_j) - \hat{G}_X(x_l)| < \pi)/S_{Pl}(t)$ be the at-risk process for the subject l eligible in the neighbors of the subject j. The number of at-risk individuals at time t equals $Y_{j.}^{\pi}(t) = \sum_{l=1}^{n} Y_{jl}^{\pi}(t)$. 2π represents the proportion of included neighbors. Moreover, the number of deaths before t equals $N_{j.}^{\pi}(t) = \sum_{l=1}^{n} N_{jl}^{\pi}(t)$ with $N_{jl}^{\pi}(t) = I(T_l \leq t, C_l \geq T_j, |\hat{G}_X(x_j) - \hat{G}_X(x_l)| < \pi)/S_{Pj}(t)$.

The conditional estimator is therefore :

$$\hat{\Lambda}_E(t|X = x_j) = \int_0^t \frac{dN^{\pi}_{j.}(u)}{Y^{\pi}_{j.}(u)} - \int_0^t \frac{\sum_{l=1}^{n} Y^{\pi}_{jl}(u) d\Lambda_{Pj}(u)}{Y^{\pi}_{j.}(u)} \tag{10.6}$$

The capacities of X to predict the disease-related deaths can be summarized by the net time-dependent ROC curve at time τ, which represents $\hat{se}_\tau(c)$ plotted against $1 - \hat{sp}_\tau(c)$ for all the thresholds c. The area under the curve (AUC) is computed by the trapezoidal rule. The 95% confidence interval of the AUC is obtained from 1000 bootstrap replications and by using the 2.5th and 97.5th percentiles of the empirical distribution. The complete methodology has been implemented in an R package `ROCt` available at http ://www.divat.fr/en/softwares/roct and is additionally available from the authors upon request. For the following applications, lifetime tables were determined by using the human mortality database (HMD, www.mortality.org).

10.3 Simulation studies

10.3.1 Methods

The expected ages of death in the general population were simulated using a Weibull PH model (shape and scale parameters equal to 1.60 and 13.75 respectively) according to sex and year of birth (regression coefficients equal to 0.16 and -0.02 respectively). The life tables used in the estimation of net time-dependent ROC curves were constructed based on the same model. Regarding the following application in transplantation, the patient characteristics were simulated as follows : binomial distribution for sex, 18-years and 70-years interval-truncated normal distribution for age at baseline and uniform distribution for calendar year at baseline. Three different scenarios were considered with various sample sizes (N=100, 250, 500 and 1000) and censoring rates (0.3, 0.5 and 0.7). The censoring times were simulated independently respecting Exponential distributions. 250 samples were simulated for each combination.

- *First scenario : all the observed mortality is in excess.* The times to death related to the disease were simulated using a Weibull PH model (shape and scale parameters equal to 1.50 and 4.48 respectively) according to one explanatory variable Z (regression coefficient equal to 1.20). Z was simulated using standard normal distribution. Related times-to-death were voluntarily quite short such that almost all patients died from their disease. The objective was to evaluate the capacity of Z to predict the excess mortality. In this scenario the area under the traditional

time-dependent ROC curve (all-cause AUCt) is the value to reach by the area under the net time-dependent ROC curve (net AUCt).

– *Second scenario : deaths may be expected or related to the disease.* The times to death related to the disease were simulated using a Weibull PH model (shape and scale parameters equal to 1.35 and 29.96 respectively) according to one explanatory variable Z (regression coefficient equal to 1.50). We were able to exactly identify the cause of death based on the minimum between the expected and the related times-to-death. Given the parameters values of the Weibull models, around 50% of the deaths were related to the disease. By censoring the expected deaths, the corresponding cause-specific AUCt represents the true prognostic capacity, i.e. the value to reach by the net AUCt used regardless the cause of the deaths which is unknown in practice. We performed two analyses within this scenario :
  (a) The evaluation of the capacity of Z to predict the excess mortality. Because Z is specific to the related time-to-event, the all-cause AUCt should be lower than the cause-specific AUCt. The worst result would be to obtain the net AUCt similar to the all-cause AUCt. The value to reach by the net AUCt is always the one obtained by the cause-specific AUCt.
  (b) The evaluation of the capacity of the linear predictor of the cause-specific Cox model based on age, sex and calendar year. Because these three factors are independent of related time-to-event, the cause-specific AUCt and the net AUCt should be close to 0.5 (i.e. non informative predictor of the related deaths)

– *Third scenario : deaths may be expected or related to the disease, but the excess mortality also depends on sex and year of birth.* The times to death related to the disease were simulated using a Weibull PH model (shape and scale parameters equal to 1.50 and 8.17 respectively), which depends on year of birth (regression coefficient equal to -0.01), sex (0.12) and Z (1.50). Z was always normally distributed. Given the parameter values of the Weibull models, around 40% of the deaths were related to the disease. Both of the previous analyses (a) and (b) will also be performed.

10.3.2 Results

In the first scenario, in which the observed mortality corresponded exclusively to the excess mortality, the means of the all-cause AUCt (value to reach) and net AUCt estimations were similar (Table 10.1). The standard deviations decreased for high sample sizes and for low censoring rates, as for every scenario.

Table 10.1: *Results of the first scenario in which the observed mortality corresponded exclusively to the excess mortality. Mean and standard deviation (within brackets) of the all-cause AUCt and the net AUCt at 10 years, calculated from the 250 simulated samples for each combination of sample size and censoring rate.*

Cens. rate	Effective	All-cause AUCt	Net AUCt
≈ 0.30	100	0.879 (0.048)	0.894 (0.048)
	250	0.893 (0.028)	0.908 (0.027)
	500	0.895 (0.020)	0.911 (0.020)
	1000	0.896 (0.015)	0.912 (0.015)
≈ 0.50	100	0.850 (0.091)	0.859 (0.085)
	250	0.856 (0.056)	0.865 (0.053)
	500	0.856 (0.035)	0.866 (0.033)
	1000	0.855 (0.028)	0.865 (0.026)
≈ 0.70	100	0.806 (0.102)	0.811 (0.098)
	250	0.810 (0.070)	0.815 (0.068)
	500	0.809 (0.049)	0.814 (0.047)
	1000	0.810 (0.035)	0.815 (0.033)

In the first analysis (a) of the second scenario (Table 10.2), the net AUCt was higher than the all-cause AUCt but lower than the cause-specific AUCt. Even if this observation is true regardless of the sample size and the censoring rate, one can notice that the distance between the net AUCt and the cause-specific AUCt (value to reach) decreased for high sample sizes and low censoring rates.

In the second analysis (b) of the second scenario (Table 10.2), conversely, the linear predictor was only related to the expected time-to-event. The net AUCt were close to 0.5 for a sample size larger than 100. However, the results were roughly equivalent whatever the censoring rate.

The results of the third scenario are presented in Table 10.3. Even though the excess mortality also depends on the risk factors of the expected mortality, the results were similar to those of the second scenario.

10.4 First application : Primary Biliary Cirrhosis data

This well-known data set includes 312 patients with hepatic failure affected by primary biliary cirrhosis (PBC) from the Mayo Clinic trial conducted between 1974 and 1984. We used the score proposed by Heagerty and Zheng [64] that takes into account 5 variables (bilirubin, albumin, prothrombin time, edema and age) : X = 0.877 * log bilirubin (in mg.l^{-1}) + 3.013 * log prothrombin time (in min) + 0.785 * (1 if edema and 0 otherwise) - 0.944 * albumin (in g.l^{-1}) + 0.033 * Age (in years)

Table 10.2: *Results of analysis (a) and (b) of the second scenario in which deaths may be expected or related to the disease. Mean and standard deviation (within brackets) of the all-cause AUCt, the cause-specific AUCt and the net AUCt at 10 years, calculated from the 250 simulated samples for each combination of sample size and censoring rate.*

		AUCt in analysis (a)			AUCt in analysis (b)		
Cens. rate	Effective	All-cause	Cause-specific	Net	All-cause	Cause-specific	Net
≈ 0.30	100	0.724 (0.054)	0.959 (0.019)	0.866 (0.083)	0.610 (0.056)	0.520 (0.069)	0.586 (0.103)
	250	0.726 (0.033)	0.964 (0.011)	0.884 (0.059)	0.588 (0.041)	0.482 (0.049)	0.516 (0.076)
	500	0.726 (0.026)	0.966 (0.008)	0.886 (0.047)	0.584 (0.027)	0.478 (0.030)	0.497 (0.046)
	1000	0.726 (0.017)	0.966 (0.005)	0.886 (0.035)	0.582 (0.020)	0.472 (0.022)	0.490 (0.036)
≈ 0.50	100	0.709 (0.056)	0.942 (0.022)	0.842 (0.083)	0.588 (0.058)	0.527 (0.069)	0.575 (0.100)
	250	0.720 (0.031)	0.953 (0.013)	0.862 (0.055)	0.570 (0.037)	0.488 (0.045)	0.514 (0.064)
	500	0.718 (0.024)	0.953 (0.010)	0.870 (0.039)	0.566 (0.027)	0.483 (0.031)	0.497 (0.046)
	1000	0.720 (0.017)	0.953 (0.006)	0.872 (0.032)	0.564 (0.021)	0.477 (0.023)	0.487 (0.038)
≈ 0.70	100	0.709 (0.060)	0.933 (0.026)	0.827 (0.083)	0.589 (0.057)	0.540 (0.080)	0.586 (0.100)
	250	0.716 (0.037)	0.943 (0.015)	0.840 (0.057)	0.565 (0.039)	0.499 (0.051)	0.525 (0.069)
	500	0.713 (0.028)	0.943 (0.011)	0.839 (0.046)	0.559 (0.029)	0.491 (0.034)	0.506 (0.047)
	1000	0.714 (0.019)	0.944 (0.008)	0.842 (0.032)	0.556 (0.022)	0.483 (0.025)	0.493 (0.036)

Table 10.3: *Results of analysis (a) and (b) of the third scenario in which deaths may be expected or related to the disease. Mean and standard deviation (within brackets) of the all-cause AUCt, the cause-specific AUCt and the net AUCt at 10 years, calculated from the 250 simulated samples for each combination of sample size and censoring rate.*

		AUCt in analysis (a)			AUCt in analysis (b)		
Cens. rate	Effective	All-cause	Cause-specific	Net	All-cause	Cause-specific	Net
≈ 0.30	100	0.769 (0.049)	0.955 (0.015)	0.891 (0.089)	0.624 (0.055)	0.554 (0.072)	0.597 (0.084)
	250	0.774 (0.035)	0.963 (0.008)	0.906 (0.061)	0.608 (0.040)	0.531 (0.043)	0.549 (0.060)
	500	0.773 (0.024)	0.963 (0.006)	0.912 (0.049)	0.601 (0.027)	0.518 (0.030)	0.528 (0.041)
	1000	0.772 (0.017)	0.964 (0.004)	0.910 (0.038)	0.598 (0.021)	0.517 (0.024)	0.518 (0.032)
≈ 0.50	100	0.756 (0.056)	0.945 (0.017)	0.872 (0.094)	0.594 (0.052)	0.547 (0.067)	0.575 (0.085)
	250	0.766 (0.034)	0.954 (0.010)	0.886 (0.067)	0.589 (0.037)	0.532 (0.040)	0.544 (0.059)
	500	0.764 (0.024)	0.953 (0.007)	0.888 (0.051)	0.584 (0.026)	0.519 (0.030)	0.527 (0.042)
	1000	0.765 (0.018)	0.955 (0.005)	0.889 (0.037)	0.580 (0.018)	0.519 (0.021)	0.517 (0.031)
≈ 0.70	100	0.747 (0.063)	0.940 (0.020)	0.839 (0.105)	0.594 (0.061)	0.559 (0.073)	0.589 (0.085)
	250	0.754 (0.043)	0.941 (0.014)	0.850 (0.073)	0.591 (0.039)	0.543 (0.048)	0.564 (0.059)
	500	0.752 (0.032)	0.944 (0.009)	0.846 (0.057)	0.575 (0.031)	0.526 (0.037)	0.534 (0.048)
	1000	0.750 (0.019)	0.943 (0.006)	0.843 (0.034)	0.573 (0.023)	0.523 (0.024)	0.526 (0.034)

The objective of this score is to predict the death (regardless of the cause) of patients on the waiting list for liver transplantation. The mean age at baseline was 50.0 years (± 10.6). 88.5% of the patients were male. The median follow-up was 5.0 years (interquartile range : from 3.3 to 7.4]). Among the patients included, 125 deaths were observed.

Figure 10.1A presents the all-cause and the net estimations of the cumulative incidence functions for deaths. At 8 years, the all-cause Cumulative Incidence Function (CIF) was 42.7% (95% CI = [35.6 ;49.0]) whereas the net CIF was 38.9% (95% CI = [31.3 ;45.7]). Because the all-cause and the net CIF were similar, most deaths seemed therefore to be related to the disease. Indeed, less than 10% of deaths up to 8 years were unrelated to the PBC. The areas under the all-cause and the net ROC curves for a prognosis up to 8 years were 0.83 for both (all-cause 95% CI = [0.77 ;0.89] and net 95% CI = [0.77 ;0.87]). Thus, the capacity of the score to predict the all-cause and the PBC-related death was similar regardless the prognostic time (figure 10.1B). One can notice that both curves logically decrease over the prognostic time, but with local increases probably due to the variance of the estimator.

10.5 Second application : Kidney Transplant Recipient data

Hernández and others have developed a score to predict the long-term survival of kidney transplant recipients (KTR) beyond the first year post transplantation [67]. This score takes into account 8 variables : recipient age at the time of transplantation, pretransplant diabetes (type I or II), pretransplant HCV antibodies, new onset diabetes within the first year (NODAT), 1-year creatinemia, 1-year daily proteinuria and delivery of Tacrolimus or Mycophenolate Mofetil (MMF) within the first year. This retrospective study was conducted on 2348 Spanish patients receiving a kidney allograft in 1990, 1994, 1998 and 2002. Only adult patients at the time of transplantation (>18 years) were considered. The mean age was 46.3 years (± 13.0) and 64.0% were men. The median follow-up was 6.8 years (interquartile range : from 4.0 to 8.6]). At the end of the study, 221 patients had died.

Table 10.4 presents the results obtained by Hernández and others using the Cox model [67]. The score was based on the sum of the products between regression coefficients and explicative variables, i.e. the linear predictor of the model. The largest positive coefficients were obtained for patients above the age of 50 with a daily proteinuria above 1g at 1-year. In contrast, the use of Tacrolimus or MMF was associated with a lower mortality.

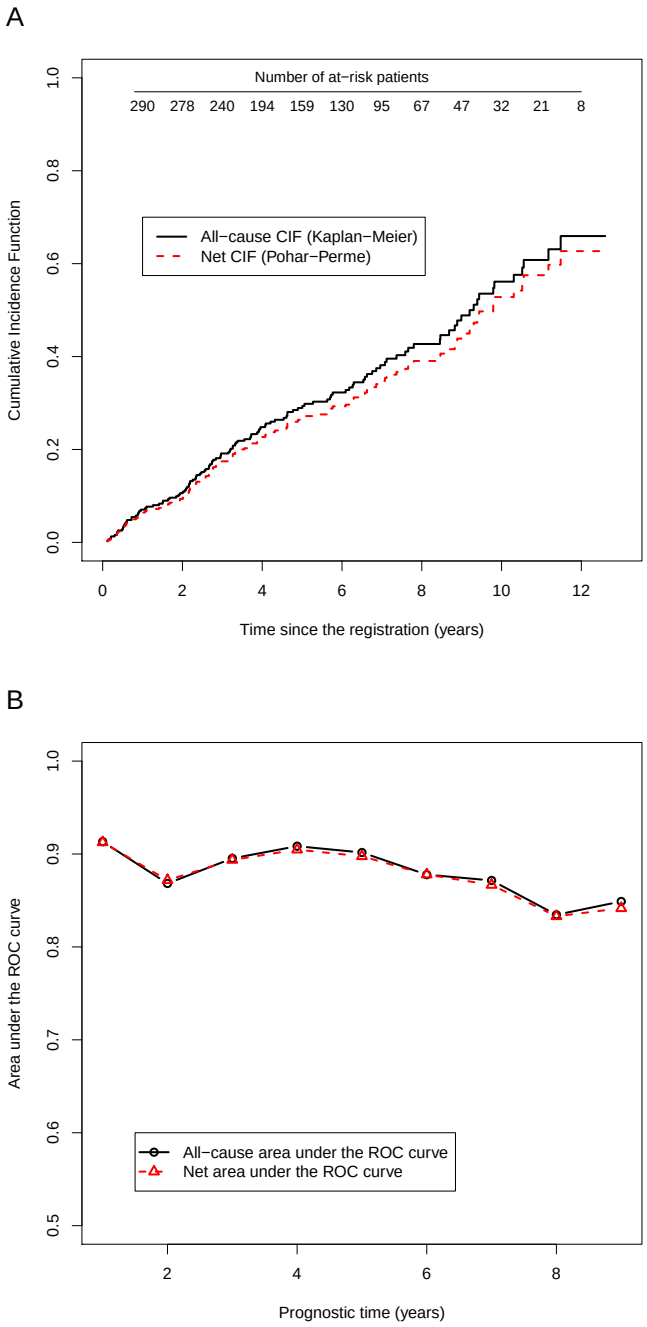

Figure 10.1: *Analysis of mortality in PBC patients (n=312).* ***(A)*** *All-cause and net estimations of the cumulative incidence functions (CIF) for death according to the time since registration on the waiting list.* ***(B)*** *Area under the all-cause and net time-dependent ROC curves according to the prognostic time.*

Table 10.4: *Multivariate analysis of risk factors for death beyond 1 year of renal transplantation obtained using a Cox proportional hazard model in the study by Hernández and others (n = 2348)[67].*

Variables	Coefficients	HR	CI95%
Age (reference class :<40 years)			
40-50 years	0.80	2.2	[1.5 ;3.3]
50-60 years	1.32	3.7	[2.6 ;5.4]
>60 years	1.91	6.7	[4.6 ;9.9]
Pretransplant diabetes	0.58	1.8	[1.1 ;2.9]
Positive HCV antibodies	0.44	1.5	[1.1 ;2.1]
NODAT at 1-year	0.45	1.5	[1.1 ;2.3]
Serum creatinemia at 1-year ($mg.dl^{-1}$)	0.56	1.7	[1.5 ;2.1]
Proteinuria >1g at 1-year	0.99	2.7	[1.8 ;4.0]
Use of tacrolimus at the first year	-0.48	0.6	[0.4 ;0.9]
Use of MMF at the first year	-0.78	0.4	[0.3 ;0.6]

We tested this score by using the prospective DIVAT cohort from Nantes University Hospital (n = 1230, www.divat.fr). The aim of this analysis was to validate and evaluate its capacity to predict the KTR-related mortality. Adults receiving a kidney transplant alone between 1996 and 2009 in Nantes were included. Returns to dialysis were considered as right censoring. Mean age at the time of transplantation was 49.0 years (± 13.8) and 62.4% of the patients were male. The median follow-up was 4.9 years (interquartile range : from 2.1 to 7.9]). Eighty-three deaths were observed. The prognostic time was defined at 10 years post transplantation.

The all-cause and net cumulative incidence functions for death are presented in Figure 10.2A. The 10-year all-cause CIF was 12.2% (95% CI = [9.3 ;14.9]) while the net CIF was 7.0% (95% CI = [3.9 ;10.0]), illustrating that among the observed deaths, a considerable number were not related to KTR. More precisely, around 40% of deaths up to 10 years were unrelated to KTR. In Figure 10.2B, the area under the all-cause time-dependent ROC curve was 0.68 at 10 years (95% CI = [0.62 ;0.74]). Therefore, the capacity of the score to predict the all-cause mortality was acceptable. The area under the net time-dependent ROC curve was estimated at 0.65 (95% CI = [0.56 ;0.72]). Consequently, it appears difficult to validate this score in the prediction of KTR-related mortality, especially considering the lower bound of the 95% CI.

10.6 Discussion

The prognostic of the all-cause mortality for patients with chronic disease is important for optimal healthcare management. Nevertheless, a tool to predict deaths related

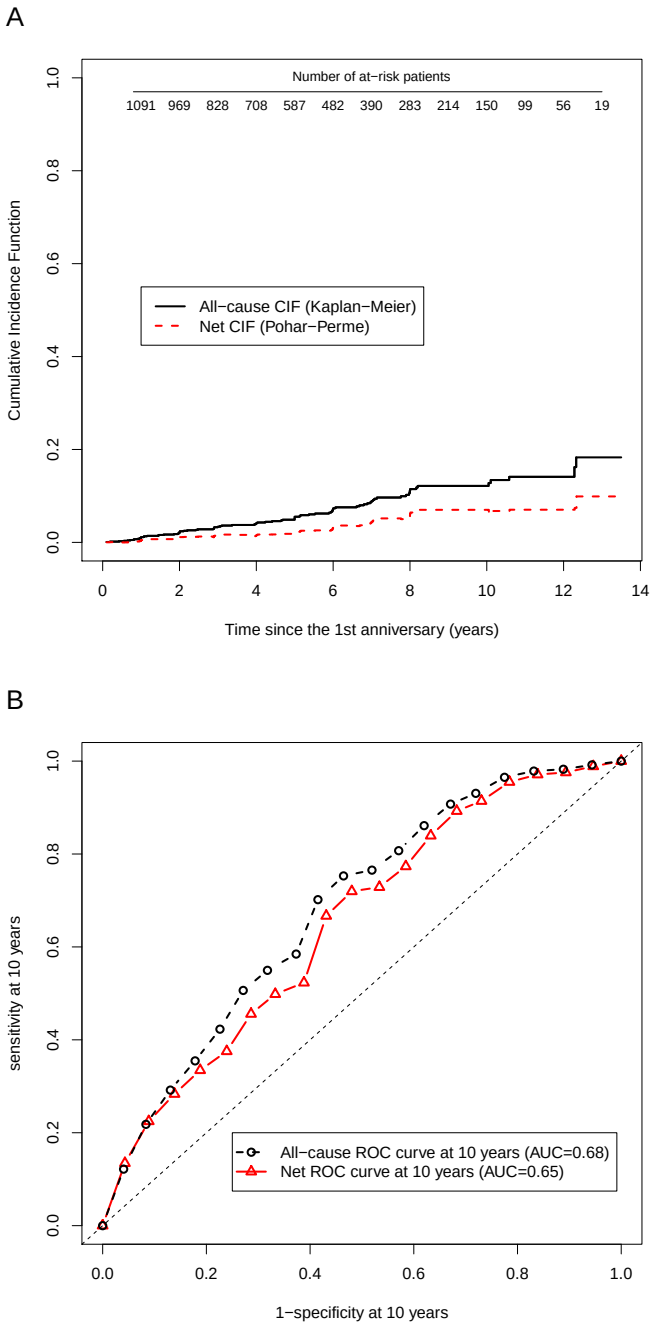

Figure 10.2: *Analysis of mortality in kidney transplant recipients (n=1230).* ***(A)*** *All-cause and net estimations of the cumulative incidence functions (CIF) for death according to time since the first anniversary.* ***(B)****. All-cause and net time-dependent ROC curves for a prognosis of mortality up to 10 years.*

to the disease may also constitute a complementary information for medical decision making. This paper describes an estimator of net time-dependent ROC curves to evaluate the accuracy of a marker to predict disease-related death by removing the expected mortality of a reference population. This methodology is useful when attribution of the deaths is impossible, which is often the case in long-term chronic disease analysis. In such situations, cause-specific or competing risk approaches cannot be used. The proposed method uses a non parametric estimator based on the net survival recently developed by Pohar and others [115] and the time-dependent ROC theory proposed by Heagerty and others [63]. The area under the net ROC curve at time t is interpretable : for two patients randomly selected, it corresponds to the probability that the patient with the higher value of the marker dies because of the disease, before the patient with the lower value. Two applications were tested, but the use of net time-dependent ROC curves can be applied to others areas of medicine and biology, in particular to the analysis of cancer registries for which the use of one additive relative survival model is well established.

We validated the proposed methodology by simulating the scenarios. If the sample size is sufficiently high, at least 250 individuals according to our results, the net AUC provided significant correction of the all-cause AUCt. Indeed, the proposed estimator was finally close to the theoretical values given by the cause-specific AUCt. As all simulation studies show, we do not expect to propose an exhaustive validation. The aim of these analyses was only to provide further information for special studies such as the second application of this paper devoted to kidney transplant recipients.

We first applied this estimator to study the score proposed by Heagerty and Zheng to predict mortality in primary biliary cirrhosis patients [64]. We demonstrated that the score had a similar capacity to predict all-cause and PBC-related mortality. This result was expected since the all-cause mortality is mainly related to the disease. In this context, the use of time-dependent ROC curves or net time-dependent ROC curves is equivalent.

The second application concerned the score proposed by Hernández and others to predict mortality beyond the first year of kidney transplantation [67]. By applying this score to a French prospective cohort, we validated the capacity of this score to predict all-cause mortality up to 10-years post transplantation (AUC = 0.68, 95% confidence interval [0.62 ; 0.74]). In contrast with the application to PBC data, a significant part of the all-cause mortality seemed to be unrelated to the kidney transplant status. It may have been due to ageing : 25% of the patients were over 60 years of age. Nevertheless, the score obtained by Hernández and others is mainly based on recipient age [67]. However, knowing that these older recipients have a greater risk of dying than younger patients is not informative. A more suitable approach would be to evaluate whether transplantation in older recipients is associated with excess mortality. By using the net time-dependent

ROC method, we were unable to validate the score by Hernández to predict excess mortality related to KTR.

The main limitation of the method is that the censoring process was not considered as informative, i.e. it does not depend on the event being studied. As for many models in time-to-event analysis, the use of the proposed net time-dependent ROC curves in the presence of informative censoring can have a noticeable effect on the results of the analysis.

Acknowledgment

We thank Dr. J.Ashton-Chess for editing the manuscript. We also wish to thank S. Le Floch for the collection of the data in the DIVAT cohort. This work was supported by a grant from the French National Agency of Research (ANR-11-JSV1-0008-01) and by a grant from the French Ministery of Health (PHRC, PROG/11/85, 2011).

- Chapitre 11 -

A multiplicative-regression model to compare the effect of factors associated with the time to graft failure between first and second renal transplant

This work constitutes the second part of the thesis of Katy Trébern-Launay. Her developments have been submitted in *BMC Medical Research Methodology*.

11.1 Introduction

In patients facing a first allograft loss, repeat kidney transplantation provides a better chance for both long-term survival and quality of life than a return to dialysis [106, 128]. The prognosis of second kidney transplant recipients (STR) compared to first kidney transplant recipients (FTR) has been frequently studied. To date the analysis of studies aimed at determining whether STR have a worse prognosis than FTR have mainly focused on older literature [143, 6]. However recent analyses with adjustments for confounding factors have challenged this generally accepted idea [11, 154], with the exception of one large, adjusted study [96]. By modelling the time-dependent hazard between FTR and STR, we also recently demonstrated that STR have a lower patient and graft survival than FTR several years after transplantation [148]. According to the literature, there is no doubt that the excess risk to STR compared to FTR is negligible considering the improvements in life expectancy and quality of life compared to dialysis therapy. Nevertheless, as the demand for kidney transplants largely exceeds the supply [157], it is necessary to evaluate the differences in risk factors between STR and FTR so as to improve graft allocation.

For this purpose, traditional survival models can be used by merging STR and FTR.

However, it should be noted that : (i) a comparison of risk factors between both groups would imply testing interactions of all the other explicative variables with the graft rank and (ii) STR-specific explicative variables (survival time of the first transplant, transplantectomy or time in dialysis before re-transplantation) cannot be analyzed, despite the knowledge that their use would improve risk evaluation [1, 55, 11, 154].

To overcome these difficulties, we proposed a multiplicative-regression model to evaluate differences in risk factors between a specific group and a reference group. This type of relative approach is often used to study the net survival of patients with cancer, i.e. after the elimination of all other causes of death [33, 60, 29, 115]. The principle is to introduce expected mortality rates adjusted for gender, age and calendar year, using life tables. Andersen et al. [9] proposed a multiplicative model also using life tables. However, to our knowledge, the applications of such methodology to endpoints other than mortality, with a reference group without a life time, has never been explored despite its potential to offer original results.

11.2 The multiplicative-regression model for relative survival

Let the individuals be indexed by j ($j = 1, ..., n_e$) in the reference group and by i ($i = 1, ..., n_r$) in the relative group. n_e and n_r represent the sample sizes of the respective groups. We note $h^o(t_i|z_i)$ the observed instantaneous hazard function at time t_i for the ith individual, where z_i is the vector of explicative variables. The observed hazard can be decomposed in the multiplication of two hazards [17, 9]

$$h^o(t_i|z_i) = h^e(t_i|z_i^e)\, h^r(t_i|z_i^r) \tag{11.1}$$

where $h^e(t_i|z_i^e)$ is the expected hazard for a reference individual with similar characteristics to the ith individual of the relative group. z_i^e is a subset of z_i and represents these common characteristics. $h^r(t_i|z_i^r)$ is the relative hazard with z_i^r being a subset of z_i.

11.2.1 Estimation of the expected hazard function

Parameters of the expected hazard can be estimated assuming a semi-parametric and proportional hazards (PH) model [23]. For the jth individual ($j = 1, ..., n_e$)

$$h^e(t_j|z_j^e) = h_0^e(t_j) \exp(\beta^e z_j^e) \tag{11.2}$$

where $h_0^e(t_j)$ is an unknown expected baseline hazard function and β^e is the vector of regression coefficients associated with z_j^e. The estimations $\hat{\beta}^e$ are obtained by maximizing the partial log-likelihood :

$$\log \mathcal{PL}_e(\beta^e) = \sum_{j=1}^{n_e} \delta_j \left\{ \beta^e z_j^e - \log \left(\sum_{k:t_k \geq t_j} \exp(\beta^e z_k^e) \right) \right\} \tag{11.3}$$

where δ_j equals 1 if the failure was observed for the subject j and 0 otherwise. The variance-covariance matrix, referred to as $\hat{V}(\hat{\beta}^e)$, is obtained via the corresponding information matrix.

11.2.2 Estimation of the relative hazard function

Parameters of the relative hazard can also be estimated using a semi-parametric PH model. For the ith individual ($i = 1, ..., n_r$), the instantaneous hazard is defined as

$$h^r(t_i|z_i) = h_0^r(t_i) \exp(\beta^r z_i^r) \tag{11.4}$$

where $h_0^r(t_i)$ is an unknown relative baseline hazard function and β^r is the vector of regression coefficients associated with z_i^r. Of note, for explicative variables not taken into account in the expected hazard, $\exp(\beta^r)$ represents the ratios of the observed hazards, comparable to the observed hazard ratios (HR) estimated by the PH model (11.2) applied to the relative group. In contrast, for explicative variables taken into account in the expected hazard model, $\exp(\beta^r)$ represents the weighting factors between the expected HR (i.e. $\exp(\hat{\beta}^e)$) and the observed HR in the relative group (i.e. $\exp(\hat{\beta}^e) \times \exp(\beta^r)$). In other words, for explicative variables involved in both models (11.2) and (11.4), $\beta^r = 0$ means that the variable has the same effect in both groups. If $\beta^r > 0$, the hazard ratio increases in the relative group compared to the reference group. If $\beta^r < 0$, the hazard ratio decreases.

By adapting the partial likelihood function (11.3) and assuming the previous estimations of expected parameters as constants, the regression coefficients β^r are estimated by maximizing

$$\log \mathcal{PL}_r(\beta^r) = \sum_{i=1}^{n_r} \delta_i \left\{ \hat{\beta}^e z_i^e + \beta^r z_i^r - \log \left(\sum_{k:t_k \geq t_i} \exp(\hat{\beta}^e z_k^e) \exp(\beta^r z_k^r) \right) \right\} \tag{11.5}$$

Nevertheless, in contrast to the traditional relative survival models based on life tables, the expected hazards cannot be reasonably assumed as constants since the corresponding parameters were estimated from the reference sample. To take into account

the variability associated with the expected model (11.2) in the estimation of the relative model, we used Monte-Carlo simulations associated with bootstrap resampling. At each of the B iterations, this procedure can be divided into the following steps ($b = 1, \ldots, B$) :

(a) Generation of a vector of parameters $\hat{\beta}_b^{e*}$ using the multivariate normal distribution $\mathcal{N}\left(\hat{\beta}^e, \hat{V}(\hat{\beta}^e)\right)$.

(b) Generation of a bootstrap sample from the relative sample comprising n_r subjects.

(c) Model (11.4) is estimated by maximizing (11.5) in which the simulated parameters $\hat{\beta}_b^{e*}$ are used instead of $\hat{\beta}^e$. $\hat{\beta}_b^r$ is the resulting estimation of the relative regression coefficients.

Means, standard deviations and 95% confidence intervals can be calculated from the B estimations $\hat{\beta}_b^{e*}$.

11.2.3 Evaluation of the proportional hazards assumption

For models (11.2) and (11.4), the only assumption is the proportionality of the hazard. Hazards proportionality was checked for each explicative variable by plotting log-minus-log survival curves obtained by the Kaplan and Meier estimator [81] and by testing the scaled Schoenfeld residuals [57] separately in the reference and relative samples. Indeed, if the observed hazard ratios are constant regardless of time in both groups, the ratio between both observed hazard ratios, i.e. the weighting factor $exp(\beta^r)$, will also be constant.

11.2.4 Software

All statistical analyses were performed using R version 2.15.1 [124]. The proposed multiplicative-regression model for relative survival was implemented in an R package `MRsurvival` available at `www.divat.fr/en/softwares/MRsurvival`.

11.3 Risk factor differences between first and second kidney transplants

11.3.1 Study population

Second transplant recipients (STR) constituted the relative group of interest. Recipients older than 18 years at the date of transplantation between 1996 and 2010 were selected from the French DIVAT (`www.divat.fr/en`) multicentric prospective cohort

[86]. Only recipients with a maintenance therapy with calcineurin inhibitors, mammalian target of rapamycin inhibitors or belatacept, in addition to mycophenolic acid and steroids were included. Simultaneous transplantations were excluded. Recipients with at least one missing data for all the variables taken into account in the expected hazard (listed below) were excluded. The same criteria were applied to the reference group composed of first transplant recipients (FTR).

Among FTR meeting the inclusion criteria, some patients were also part of the STR group as they had received two transplants during the observation period. These 37 patients, who were included in both cohorts, represented 2% and 7% of the FTR and STR groups respectively. Given the large number of explicative variables, it seemed reasonable to assume conditional independence of the two transplantations of a given patient. In order to validate this assumption, we performed a frailty Cox model [129] based on the 37 individuals who were included in both groups. The frailty term was assumed to be Gamma distributed. The variance of the random variable was estimated at 5.10^{-9} ($p = 0.9948$). Therefore, no intra-individual dependency was demonstrated.

We identified 2462 FTR who met the inclusion criteria. We excluded 256 FTR (10.3%) with one missing data for at least one of the variables taken into account in the expected hazard model. Finally, 2206 FTR made up the reference group. The principal outcome was the time between transplantation and graft failure, which was the first event between return to dialysis and patient death with a functioning graft. The mean follow-up was 3.4 years with a maximum of 13.7 years. During the observation period, 191 returns to dialysis and 109 deaths were observed. 641 STR potentially made up the relative group of interest, but 75 STR (11.7%) with missing data for explicative variables of the expected hazard were excluded. Finally, 566 STR were included in the group of interest. The mean follow-up was 3.1 years with a maximum of 13.1 years. During the observation period, 72 returns to dialysis and 34 deaths were observed.

The demographic and baseline characteristics at the time of transplantation are presented in Table 11.1. Regardless of the group, the majority of patients received a transplant from a deceased donor and the recipient gender was comparable between groups. However, STR were younger and their transplants were provided by younger donors. Recurrent nephropathy, past history of cardiac disease, hepatitis and malignancy were more frequent among STR, but STR had less diabetes and were less likely to be obese at the time of transplantation. Compared to FTR, STR received better HLA-matched transplants, but their cold ischemia time was longer and they were more immunized against HLA class I and class II antigens (historical Panel Reactive Antibodies) than FTR. They were also more frequently exposed to induction therapy with a lymphocyte-depleting agent.

Table 11.1: *Characteristics at the date of transplantation : (i) whole cohort and (ii) FTR and STR separately. The last three rows of the table concern covariates specific for STR.*

	All grafts (N = 2772)		FTR (N = 2206)		STR (N = 566)		
Demographic characteristics	Number	(%)	Number	(%)	Number	(%)	p-value
Transplantation period < 2005	594	(21.4)	457	(20.7)	137	(24.2)	0.0806
Recipient ≥ 55 years of age	1175	(42.4)	994	(45.1)	181	(32.0)	<0.0001
Male recipient	1705	(61.5)	1362	(61.7)	343	(60.6)	0.6536
Recurrent causal nephropathy	906	(32.7)	666	(30.2)	240	(42.4)	<0.0001
History of diabetes	306	(11.0)	269	(12.2)	37	(6.5)	0.0002
History of hypertension	2263	(81.6)	1804	(81.8)	459	(81.1)	0.7545
History of vascular disease	352	(12.7)	272	(12.3)	80	(14.1)	0.2804
History of cardiac disease	903	(32.6)	686	(31.1)	217	(38.3)	0.0012
History of dyslipemia	799	(28.8)	661	(30.0)	138	(24.4)	0.0104
History of malignancy	228	(8.2)	147	(6.7)	81	(14.3)	<0.0001
History of hepatitis B or C	168	(6.1)	96	(4.4)	72	(12.7)	<0.0001
Recipient BMI ≥ 30 kg.m^{-2}	263	(9.5)	235	(10.7)	28	(4.9)	<0.0001
Positive anti-class I PRA	706	(25.5)	355	(16.1)	351	(62.0)	<0.0001
Positive anti-class II PRA	733	(26.4)	319	(14.5)	414	(73.1)	<0.0001
Donor ≥ 55 years of age	1172	(42.3)	973	(44.1)	199	(35.2)	0.0002
Deceased donor	2470	(89.1)	1940	(87.9)	530	(93.6)	0.0002
Donor serum creatinine ≥ 133 μmol/l	342	(12.5)	279	(12.8)	63	(11.4)	0.3807
Positive donor EBV serology	2613	(94.3)	2087	(94.6)	526	(92.9)	0.1540
HLA-A-B-DR incompatibilities > 4	365	(13.2)	326	(14.8)	39	(6.9)	<0.0001
Cold ischemia time ≥ 24h	754	(27.2)	552	(25.0)	202	(35.7)	<0.0001
Lymphocyte-depleting induction	1223	(44.1)	793	(35.9)	430	(76.0)	<0.0001
First graft survival < 1 year	-	-	-	-	131	(24.1)	-
Waiting time before retransplantation ≥ 3 years	-	-	-	-	272	(49.8)	-
Transplantectomy of the first graft	-	-	-	-	220	(38.9)	-

BMI, body mass index ; PRA, panel reactive antibody ; EBV, Epstein-Barr virus ; HLA, human leukocyte antigen

11.3.2 Analysis of risk factors in the FTR population

As previously illustrated, it is well-established that FTR and STR are not intrinsically comparable. Thus, for the analysis of risk factors in the FTR population, adjustments were made *(i)* for all of the possible pre- or per-transplant immunological and non-immunological confounding factors according to experts and *(ii)* for all the baseline parameters differentially distributed between FTR and STR (Table 11.1). Thus, the expected hazard of graft failure was estimated according to recipient age and gender, causal nephropathy, comorbidities (including history of diabetes, hypertension, cardiac or vascular disease, dyslipemia, B or C hepatitis and malignancy), obesity, pre-transplantation immunization (PRA) against class I and class II antigens), donor age, deceased or living donor status, Epstein-Barr Virus (EBV) serology, period of transplantation, level of HLA-A-B-DR mismatches, induction therapy and cold ischemia time according to a threshold of 24 hours. This modelling is explained in detail in the paper by Trébern-Launay et al. [148]. The final multivariate model in the reference group of FTR is presented in Table 11.2.

11.3.3 Relative hazard modelling

The final relative model is presented in Table 11.3. Expected HR are also presented in the first column (but estimated previously in FTR) to enable a direct comparison between FTR and STR. Donor gender and waiting time before re-transplantation were not taken into account in the expected hazards. As a validation, the corresponding HRs were thus similar between the multiplicative relative model and the Cox model applied to the STR group. Therefore, we calculated a 1.5-fold increase in risk of graft failure for STR with grafts from males compared to STR with grafts from females ($p = 0.0320$). This risk factor was not identified for FTR ($p > 0.05$, Table 11.2). Moreover, STR who waited more than 3 years in dialysis before retransplantation had a 1.9-fold increased risk compared to STR with a shorter waiting time ($p < 0.0001$). Of note, for both explicative variables, 95% confidence intervals were also similar between the relative and the Cox model.

In contrast, the effect of recipient age and donor age seemed significantly different between FTR and STR ($p < 0.05$). More precisely, the expected HR associated with recipient age $\geq$ 55 years would be 1.39 in the STR group regarding the HR observed in the FTR group (11.2). In fact, the relative model showed that this HR was 1.6-fold higher for STR compared to FTR (CI95% = [1.01-2.72], $p = 0.0480$). This results was validated by the HR estimated from the Cox model applied in the STR group ($2.65 \approx 1.39 \times 1.65$). Similarly, the effect of donor age $\geq$ 55 years was nearly two fold lower

Table 11.2: *Multivariate Cox model analysis of graft failure risk factors for FTR (N = 2206).*

Covariates	Hazard Ratio	[95% CI]	p-value
Transplantation period (< 2005 / ≥ 2005)	1.33	[0.97-1.82]	0.0693
Recipient age (≥ 55 years / < 55 years)	1.39	[1.05-1.83]	0.0204
Recipient gender (male / female)	1.17	[0.91-1.51]	0.2186
Causal nephropathy (recurrent / non recurrent)	1.24	[0.96-1.59]	0.0987
History of diabetes (positive / negative)	1.34	[0.96-1.85]	0.0819
History of hypertension (positive / negative)	0.77	[0.57-1.05]	0.0986
History of cardiac disease (positive / negative)	1.41	[1.11-1.79]	0.0051
History of vascular disease (positive / negative)	1.10	[0.81-1.51]	0.5351
History of dyslipemia (positive / negative)	1.12	[0.87-1.45]	0.3828
History of hepatitis B/C (positive / negative)	0.82	[0.45-1.47]	0.4969
History of malignancy (positive / negative)	1.25	[0.84-1.86]	0.2698
Body mass index (≥ 30 kg.m-2 / < 30 kg.m-2)	1.58	[1.12-2.14]	0.0084
Anti-class I PRA (positive / negative)	1.45	[1.07-1.97]	0.0182
Anti-class II PRA (positive / negative)	1.09	[0.78-1.52]	0.6299
Donor age (≥55 years / <55 years)	1.34	[1.03-1.74]	0.0313
Donor status (deceased/living)	2.50	[1.41-4.43]	0.0016
Donor EBV serology (positive / negative)	1.65	[0.98-2.78]	0.0606
Number of HLA-A-B-DR mismatches (> 4 / ≤ 4)	1.30	[0.97-1.76]	0.0824
Induction therapy (depleting / non depleting)	0.79	[0.60-1.05]	0.1091
Cold ischemia time (≥ 24 h / < 24 h)	1.29	[1.01-1.66]	0.0441

PRA, panel reactive antibody ; EBV, Epstein-Barr virus ; HLA, human leukocyte antigen

for STR than for FTR (CI95% = [0.33-0.99], p = 0.0440), while it was identified as a significant risk factor for FTR (Table 11.2, HR = 1.34, p = 0.0313). This results was also concordant with the HR estimated from the Cox model applied to the STR group ($0.83 \approx 1.34 \times 0.59$). Of note, the relationship between the recipient gender and the risk of graft failure was not found to be significantly different between FTR and STR (p = 0.0720). In order to validate the variance estimations of these relative parameters, we compared the 95% CI of the observed HR obtained by a Cox model applied to the STR-group with the 95% CI of the expected HR deduced from the sum of the B sampled expected parameters and the B simulated relative estimations (last columns of the Table 11.3).

11.4 Discussion

Although the comparison of survival between first and second kidney transplants has been frequently performed, no study has addressed the issue of comparing the risk factors associated with the time to graft failure between both groups. Understanding the factors influencing the long-term evolution of STR compared to FTR would benefit the medical management of graft attribution by identifying patients with the best outcomes.

The absence of literature focusing on this question may be partially explained by the methodological issues associated with such studies. Indeed, the Cox model is classically used to explore risk factors influencing graft survival and interactions can be included to evaluate risk factor differences between FTR and STR. However this approach has several limitations. Firstly, it implies testing interactions between the graft rank and each explicative variable, increasing the number of parameters and making interpretations difficult. Secondly, as only covariates common to both groups can be taken into account, this excludes explicative variables specific for one group. Concerning our application, this constitutes a limitation as several STR-specific explicative variables are known to be associated with second graft prognosis : the first graft transplantectomy [1], the first graft survival duration [55, 154] or the time in dialysis before re-transplantation [11].

This paper describes an innovative use of a multiplicative-regression model for relative survival with the aim of accurately comparing risk factors between two groups of patients. This approach was initially used to analyze relative mortality using life tables (in which the expected rates are considered as constants). Thus, we proposed a semi-parametric model based on partial likelihood maximisations : one Cox model for the expected hazard and one multiplicative-regression for the relative hazard. The standard deviations of the relative parameters were obtained by Monte-Carlo simulations associated with bootstrap re-sampling. Moreover, the estimations of the relative parameters

Table 11.3: *Results of the relative survival model based on 540 STR (26 recipients presenting missing data for the waiting time before re-transplantation were excluded). (i) The first column provides the expected HR obtained from the semi-parametric PH model fully described in Table 11.2 and obtained for the FTR. (ii) The next three columns provide results estimated using the relative survival model. (iii) The "Cox model (STR)" columns provide results estimated from the Cox model applied in the STR group (N=540). (iv) The "Expected Cox model" columns provide results expected for STR obtained from the sum of* $\hat{\beta}_b^{e*}$ *and* $\hat{\beta}_b^r$.

	Cox model (FTR)	Relative model (STR)			Cox model (STR)			Expected Cox model	
	HR	HR	[95% CI]	p-value	HR	[95% CI]	p-value	HR	[95% CI]
Transplantation < 2005	1.33	0.97	[0.55-1.74]	0.9360	1.30	[0.80-2.10]	0.2868	1.30	[0.73-2.32]
Male recipient	1.17	0.61	[0.38-1.05]	0.0720	0.76	[0.51-1.14]	0.1860	0.72	[0.47-1.12]
Recipient age ≥ 55 years	1.39	1.65	[1.01-2.72]	0.0480	2.65	[1.69-4.14]	<0.0001	2.23	[1.40-3.56]
Donor age ≥ 55 years	1.34	0.59	[0.33-0.99]	0.0440	0.83	[0.52-1.34]	0.4551	0.79	[0.48-1.31]
Male donor	-	1.53	[1.03-2.48]	0.0320	1.56	[1.01-2.43]	0.0461	1.53	[1.03-2.48]
Waiting time >3 years	-	1.92	[1.22-3.00]	<0.0001	2.14	[1.41-3.26]	0.0004	1.92	[1.22-3.00]

and the standard deviations of the multiplicative-regression model were validated by recomputing the HRs and the confidence intervals of the Cox model estimated in the relative group. In this multiplicative modelling, the regression coefficients are straightforward interpreted in terms of interactions. Moreover, specific explicative variables can be included in the relative model.

The results showed that male donor gender and long waiting time before re-transplantation were two specific-STR risk factors : donor gender was not significantly associated with the risk of graft failure in the FTR population and the waiting time before retransplantation was only related to STR. The interpretations were similar to hazard ratios from a Cox model preformed on the STR group. Conversely, two explicative variables appeared to be differently associated with the risk of graft failure between STR and FTR. These factors were involved in the regression of the expected hazard, therefore they represented weights of the expected hazard ratio. More precisely, we showed for the first time that the adverse effect of recipient age was enhanced for STR as compared to FTR. The main clinical explanation is a cumulative effect of the risk factors for STR, in particular because of the cumulative exposure to immunosuppressive drugs during the first transplantation period. From a clinical point of view, this result may imply that clinicians should pay a particular attention to recipient age in second kidney transplantations. Also, for the first time to our knowledge, this study identified an attenuation of the risk factor related to older transplants for STR as compared to FTR. Two explanations may be outlined : *(i)* an indication bias with only high-quality donors (without diabetes, hypertension or cardiovascular disease) proposed to STR ; *(ii)* a higher non-HLA immunization in STR, explaining why graft failure is generally due to immunological phenomena rather than transplant quality.

Although we illustrated the advantages of this approach in renal transplantation, this methodology may be useful in number of other clinical and epidemiological applications. We propose an R package for this purpose. Of course, the aim of this multiplicative relative approach is not to replace traditional survival models, but rather to provide a more suitable alternative when the main objective is to compare risk factors between two populations, in particular when population-specific covariates need to be included.

As always, there are several avenues for further work. The main limitation of the method is that only time-invariant explicative variables are considered. A generalization with time-dependent covariates is possible by adapting the partial likelihood in this context [84].

In conclusion, the proposed multiplicative-regression model for relative survival enabled the analysis of an outcome other than mortality, using a reference group other than the general population. This study highlighted novel risk factor differences bet-

ween first and second kidney transplant recipients. These results could help improve the management of patients waiting for a second graft. They may also encourage the widespread use of this original methodology in other medical fields.

Acknowledgment

We thank Dr. J.Ashton-Chess for editing the manuscript. We also wish to thank the DIVAT scientific council and the members of the clinical research assistant team (M. Kessler, M. Ladrière, JP. Soulillou, C. Legendre, H. Kreis, G. Mourad, V. Garrigue, L. Rostaing, N. Kamar, E. Morelon, F. Buron, S. Le Floch, C. Scellier, V. Eschbach, P. Przednowed, K. Zurbonsen, V. Godel, K. Zurbonsen, X. Longy, C. Dagot, F. M'Raiagh) for the collection of the data in the DIVAT cohort and Dr. M. Pohar-Perme for her helpful advice. This work was partly supported by the RTRS, the Fondation de Coopération Scientifique CENTAURE and Roche Laboratory.

- Chapitre 12 -

The extension to other medical fields

As illustrated in all the previous chapters, our developments were mainly motivated by the analysis of kidney transplant recipients. Nevertheless, these developments can also be useful to other medical fields.

12.1 Application in bone marrow transplant

12.1.1 Context

Allogeneic haematopoietic stem cell transplantation (HSCT) is a curative option for many haematological malignancies. However, HSCT is limited by its toxicity, especially graft-versus host disease (GVHD) both in its acute and chronic forms. Chronic GVHD (cGvHD) is a relatively common complication, with an incidence ranging from 40% to 70% and is one of the most severe consequences of HSCT. On the long term, occurrence of cGvHD is associated with non-malignant organ or tissue dysfunction, infections, abnormal immune reconstitution and secondary cancers. cGvHD is lethal in approximately 15% of transplant recipients.

Many studies have identified risk factors associated with the development of cGvHD. However, most of them were retrospective, single-centred or/and reported on a relatively limited number of patients. One of the largest studies was recently published by Flowers et al. [40] evaluating 2941 adult and paediatric patients. This study identified donor-recipient HLA mismatch and the use of unrelated donors as risk factors for developing cGvHD, whereas total body irradiation was strongly associated with acute GVHD (aGvHD). Nevertheless, the occurrence of aGvHD has been documented as a main risk factor for subsequent cGvHD. Moreover, G-CSF-mobilized peripheral blood

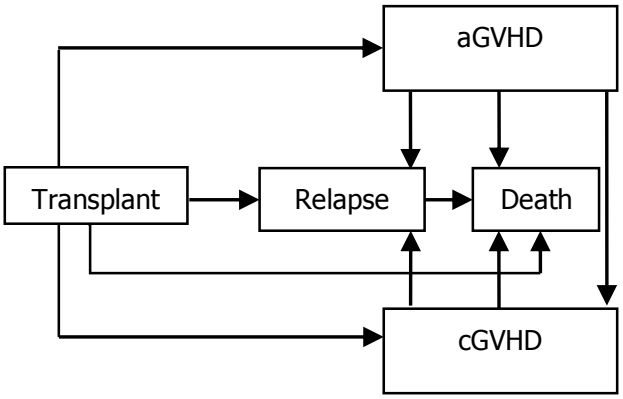

Figure 12.1: *Multistate evolution of patients after allogeneic haematopoietic stem cell transplantation*

stem cells are nowadays a well-established risk factor for cGvHD. Despite identification of these risk factors, no large multicenter or prospective cohorts have accurately and comprehensively assess risk factors for cGvHD.

Furthermore, the results from literature are limited by the relative weakness of the statistical methods, especially given the complexity and the dynamic nature of cGvHD evolution after HSCT (Figure 12.1). Cause-specific Cox model or competing risk models are often used in order to evaluate specific associations of factors on each transition (for instance in Flowers et al.), but multistate model would be more adapted to describe precisely the real dynamics. Factors can, for instance, firstly influence the time-to-aGvHD which secondly constitutes one of the risk factors for cGvHD development among the other factors. Risk factors for cGvHD without aGvHD can also be identified in the same model and all these estimations take into account the informative censoring of deaths. Relapses will also be considered as competing event. The development of such models can improve the knowledge about clinical risk factors of GVHD.

Despite this perfectible literature on risk factors for patients and GVHD, predicting GVHD still remains elusive to physicians. This is partially caused by the incorrect use of statistical terminology, resulting in overoptimistic interpretation of results. Therefore, there is an important similitude with kidney transplantation [42]. For instance, MacMillan et al. [95] compared the prognostic utility of the Minnesota versus the Center for International Blood and Marrow Transplant Research (CIBMTR) grading systems. As no GVHD grading system optimally predicted outcomes, a novel aGvHD risk score was proposed. These analyses were based on different logistic and Cox regressions which are not suitable to such multistate context. Moreover, a growing body of novel literature in biostatistics demonstrates that odds ratios or hazard ratios and the corresponding

p-values do not indicate whether a given variable (clinical factor, biomarker or a scoring system) will be a good predictor. New indicators like time-dependent Receiver Operating Characteristic (ROC) curves or other concordance indexes for multistate data are likely more appropriate [132, 47], in order to provide clinicians with a clearer and better understanding of the true clinical utility of GVHD predictive tests.

12.1.2 Preliminary results

We have begun this collaboration with Pr. M. Mothy and Dr. E. Brissot few years ago. The principal objective of this work was to identify cytokines associated with the time to extensive cGvHD (without considering the local cGvHD). 152 patients treated with allo-SCT between February 2005 and July 2008 has been included in the study. 40 cytokine levels have been collected at 3 months post-treatment.

The data have been collected at baseline, i.e. the time of cytokine measurements (age, gender, chemotherapy, donor characteristics, recipient and donor CMV serology, disease origin, prophylactic treatment, initial disease and ATG treatment), within the first 3 months (date of aGvHD) and prospectively afterword (date of cGvHD, characteristics of cGvHD, date and cause of death). The principal outcome was the time between the cytokine measurement and the extensive cGvHD. Cytokines were first evaluated by univariate Fine and Gray model. In such competing risk models, the deaths without extensive cGvHD were treated as competing event. The right censored observations consisted in patients alive without extensive cGvHD at the end of the follow-up. An important question is to identify the cytokines with a possible independent effect from the well-established clinical determinants. We thus performed univariate analyses of the clinical parameters. Significant clinical parameters ($p < 0.25$) were further analysed in a multivariate Fine and Gray model. We finally retained the clinical parameters if $p < 0.05$. Then, we computed the previous clinical multivariate model by adding each cytokine one-by-one. The prognostic capacities of the significant cytokines previously retained were evaluated by computing the non-parametric estimator of the time-dependent Receiver Operating Characteristic (ROC) curves proposed by Saha and Heagerty for competing events [132].

According to the previous methodology, the cytokine MDC was significantly associated with the time to extensive cGvHD regardless the analysis : (i) univariate, (ii) by adjusting on clinical risk factors and (iii) by adjusting on clinical risk factors and cytokines. We thus decided to study its capacity to predict extensive cGvHD. The area under the ROC curves for a prognostic up to 2 years was 0.69 (CI95%=[0.57,0.78]).

12.1.3 Description of the project plan

To support this project, we answered to the call for proposals in Translational Cancer Research (Institut National du CAncer, INCA 2013).

a. The construction of the data base.

The first part of the project concerns the epidemiological analyses. The data will be extracted from the registry of the European group for Blood and Marrow Transplantation (EBMT). EBMT registry is a voluntary working group of more than 500 transplants centers, whose participants are requested to report all consecutive stem cell transplantations and follow-ups once a year. We have decided to select a cohort with restricted criteria in order to obtain a homogeneous population : data of adult acute myeloid leukemia in first and second complete remission at transplantation and given G-CSF-mobilized peripheral blood stem cells (PBSC) from HLA-matched sibling or HLA-matched unrelated donors, between 2000 and 2009 after educed-intensity conditioning regimen. Patients given ex-vivo T-cell depleted grafts, those given other stem cell sources than PBSC, those given pre-emptive donor lymphocyte infusions and those who failed to engraft will be excluded. The date and severity (limited versus extensive) of cGvHD are prospectively collected using the EBMT Minimum Essential Data-A Form. The sample will be randomly divided into two groups : 2/3 of the patients in the training sample (for the risk factors analysis and for the clinical-based scoring systems for event prediction) and 1/3 in the validating sample. Regarding the following statistical analysis, no sample size can be a priori calculated. We will include all the patients regarding the previous inclusion criteria. This study will concern more than 1800 patients.

b. The estimation of multistate model, the construction of predictors and the evaluation of the prognostic capacities.

Semi-Markov models will be used to model the evolution of the patients within several states. Covariates can be included in the waiting time hazard function or in the embedded Markov chain. The waiting times will be assumed generalized-Weibull distributed, with a possible simplification if necessary into Weibull or even Exponential distributions. No interval-censoring will be observed. The estimations of parameter will be performed by likelihood maximization. The goodness-of-fit of the model will be done according to the methodology we proposed in 2010 [46]. Univariate and multivariate analyses will be performed respecting type I error at 0.20 and 0.05, respectively. The scoring systems for the predictions of each time-to-transition will be obtained by the expected semi-Markov hazard function (the join process of the sequence of states and the time of the transition). These estimations will be done based on the patients in the training sample.

Time-dependent ROC curves will then be adapted to this multistate context in order to evaluate each prognostic capacity of scoring system for the corresponding transition and for the overall capacity to predict cGvHD. This step is associated with the continuity of our previous developments [47]. These ROC curves will be estimated based on the patients in the validation sample.

c. Serum samples and cytokine measurements

For the purpose of this study, peripheral blood samples have been collected around day 100 after allo-SCT. After blood collection, serum has been immediately obtained by centrifugation, transferred into cryotubes, and is stored at -80^{0}C until further processing. A total of 40 cytokines will be studied on blood samples. These 40 cytokines are chosen based on previously published data and based on their putative mechanistic role in the pathophysiology of cGvHD. They will be determined by the bead-based multiplex protein array technology. These analyses will be performed by using the biocollection constructed in the Nantes university hospital. A total of 250 patients will be included. These patients will not be included in the previous analyses based on the multi-centric cohort constituted by 1800 patients. Therefore, it will be possible to independently evaluate if the cytokines can improve the clinical-based prognostic of cGvHD and to propose a specific bio-clinical-based signature of cGvHD.

12.2 Prognostic of valvular heart disease after replacement by bioprosthetic valve

Valvular heart disease can be acquired or congenital. Irrespective of the etiology of valvular heart disease, deterioration of the native valve can result in mitral or aortic regurgitation or stenosis, necessitating replacement with a prosthetic valve [74]. The main issues with bioprosthetic valves are their finite lifespan and their risk of Structural Valve Deterioration (SVD). SVD does not occur for mechanical valve, but it requires lifelong anticoagulation to prevent thromboembolic complications [37]. One advantage of bioprosthetic valves is that they are much less thrombogenic, but serious SVD can require reoperation. Each year, 300.000 patients world-wide receive a bioprosthetic valves.

The objective of the PhD thesis of Thomas Senage, junior cardiac surgeon, is to better evaluate this SVD risk for patients with bioprosthetic valves. Indeed, on can expect a large under-estimation of the SVD incidence since only valve biospies from re-operated patients constitute the official diagnostic of SVD. Nevertheless, patients who cannot be re-operated or who died are not taken into account. In order to avoid this important bias due to competing various health deteriorations, Dr. T. Senage and Dr.

J.C. Roussel are proposing a systemic diagnostic of SVD by regular echography. They have constituted a cohort with more than 500 patients. T. Senage has begun his thesis in December 2012 in the SPHERE laboratory.

In parallel to this clinical-based project, Dr. J.C. Roussel is in charge of a clinical part of a European project (FP7) devoted to the investigation of the role played by the immune system in the premature structural deterioration, calcification, and failure of bioprosthetic valves. Such immunological responses may be underestimated according to the little attention paid in the literature. If the project is accepted, the SPHERE laboratory will be central in the statistical and methodological expertise regarding the coherence with our previous developments in renal transplantation.

12.3 Meta-analysis of prognostic marker in cancer

Cancer is a major public health problem but personalized medicine is a promise of future cures. This will completely change how treatments are given. Risk stratification at diagnosis permits the identification of patients who may benefit from active surveillance, those who might benefit from immediate local treatment and those who require aggressive multi-modal therapy. Numbers of important prognostic factors have been identified in the literature, some generic to all cancers and some specific for different cancer types. However, these studies differ from each other with regard to population studied, study design, sample size and factors adjusted for in the analyses. Therefore, recent meta-analyses were performed to summarize the role of such prognostic markers. Meta-analysis results for the purpose of integrating the findings of different studies. All these meta-analyses are based on pooled hazard ratios. Nevertheless, hazard ratio only indicates an increase in failure risk, not the prognostic accuracy.

As previously explained (chapter 5), we proposed a solution to this methodological issue. We developed a method, called SROCt, to assess time-dependent summary ROC curves from published results (Combescure et al., 2012). In order to compute the SROCt estimator, the following data have to be collected from each paper : the survival curves par strata defined by the marker, the threshold(s) used for defining the strata and the number of individuals in each strata. As always, there are several avenues for further work regarding these first biostatistical developments : this approach needs parametric assumptions on the distribution of the marker and on the survival process.

The overall objective of this project is to improve the results from published studies and meta-analysis of prognostic marker in cancer research. These improvements can be divided into two complementary parts. 1) We will continue the methodological developments with developments to derive time-dependent sensitivities, specificities and

likelihood ratios from published survival curves. It is also important to develop a precise non-parametric estimation of summary survival curves based on such aggregate data. Then, these developments will be used to improve the SROCt framework, especially by exploring a non-parametric estimation. 2) This SROCt methodology will be applied to the eight meta-analyses. According to the interest of the results in oncology and the expertise area of the ICO (Institut de Cancérologie de l'Ouest), we will particularly focus on two studies : the human epidermal growth factor receptor 3 (HER3) in solid tumor [105] and vascular endothelial growth factor (VEGF) in ovarian cancer [159].

The eight meta-analyses will offer the main list of publications for each topic. In order to potentially complete these lists for the two main analyses of our project, we will perform additional researches. Only the studies with the description of survival curves according to the level of the marker will be included. In order to estimate these models and compute SROCt curves, we will use the R package SROCt we have developed (`www.divat.fr`) and we will extend. First, by using utility scores of the possible outcomes of a prognostic test, one can compute the optimal threshold, i.e. the point on the ROC curve that best achieves the goal desired by the decision maker. Second, a methodology will be developed for non-parametric and time-dependent sensitivities/specificities from published survival curves. A ROC plot can represent these sensitivity and specificity pairs. Based on this plot, regression may be performed to obtain a continuous ROC curve, which may constitute alternative estimator of SROCt curve. Third, one can also improve the non-parametric estimation of pooled survival curves.

It is possible that the SROCt estimations based on these meta-analyses will not validate the prognostic capacity of the different markers. Nevertheless, we hope to validate few markers regarding their prognostic capacity and to additionally propose validated cut-offs for medical decision making. Regardless the negativity or positivity of the results we will obtain, this project will improve the evaluation of the clinical utility of these markers. The 8 selected meta-analyses appear important in terms of public health and were recently published, underlying the relevance of our project.

To support this project, we answered to the call for proposals in "Projets libres de recherche en Sciences Humaines et Sociales, Epidémiologie et Santé Publique" (Institut National du CAncer, INCA 2013).

Conclusions et Perspectives

L'ensemble de ce document résume 9 ans de travail de recherche depuis l'obtention de mon DEA. Il apparait évident que ce travail a été nourri et guidé par les projets issus de la cohorte DIVAT. Les patients transplantés représentent un terreau riche pour la recherche épidémiologique et Biostatistique. Il s'agit en effet d'une pathologie qui commence à une date précise et dont les paramètres du suivi sont connus et codifiés. Le suivi dès la greffe, l'évolution complexe et la volonté de personnaliser la prise en charge, nous ont amené successivement à développer trois axes de recherche dans le domaine de l'analyse des données incomplètes : les modèles multi-états, les courbes ROC dépendantes du temps et les modèles de survie relative. Les passerelles entre ces thématiques sont nombreuses. Nous développons par exemple les courbes ROC en présence de risques compétitifs, les courbes ROC nettes ou les modèles semi-Markoviens avec prise en compte de la mortalité attendue.

L'objectif à court terme est de finaliser ces projets. Certains de ces travaux ne sont pas présentés ici, en particulier les développements avec Christophe Combescure (Hôpitaux Universitaires de Genève, ex : courbes ROC pronostiques ou courbes de survie moyennes à partir de méta-analyses) ou ceux d'Etienne Dantan (modèles conjoints, interprétations et développements d'indicateurs pronostiques).

Je crois à la continuité des travaux méthodologiques que nous menons en parallèle des travaux plus appliqués de recherche clinique. Le score KTFS, obtenu à partir de l'analyse de DIVAT et de l'utilisation pour la première fois en transplantation rénale de courbes ROC dépendantes du temps, est un bon exemple. Ce travail a abouti aujourd'hui à un nouvel essai clinique financé par PHRC national (TELEGRAFT). Cette étude randomisée en cours devra valider l'utilité du score pour adapter la prise en charge du patient transplanté rénal. TELEGRAFT ouvre de nouvelles perspectives de recherche : la qualité de vie, l'adaptation des patients à leur maladie et les coûts de leur prise en charge étant des paramètres très importants à prendre en compte. De nouvelles collaborations s'ouvrent ainsi avec nos collègues de l'équipe SPHERE (EA-4275) : Philippe Tessier (MCU) pour l'économie de la santé, Angélique Bonnaud-Antignac (PU) pour la

psychologie clinique, Véronique Sébille-Rivain (PUPH) et Jean-Benoit Hardouin (MCU) pour le développement de questionnaires et l'interprétation des données subjectives des patients (PRO, Patient Reported Outcome). Un projet autour de l'évolution des receveurs d'une greffe préemptive renforcera aussi cette cohérence (projet PHRC 2013).

Les autres outils pronostiques en cours de construction ouvriront sur des projets similaires. Ces outils étant issus de données observationnelles, il est important de les valider à partir d'études expérimentales. Les cinq prochaines années seront donc axées sur cette poursuite scientifique.

Je souhaite conserver cette direction, cela sous-entend de consolider notre équipe regroupée autour de l'analyse de données censurées. L'obtention de l'HDR va dans ce sens. La place que prend et pendra Etienne Dantan est essentielle pour ce renforcement. Je suis cependant convaincu que la seule dimension universitaire que nous apportons sera limitée, que ce soit pour le temps que nous pouvons consacrer à nos activités de recherche en plus de l'enseignement et du soutien méthodologique au CHU de Nantes, que pour l'augmentation de l'effectif de ce groupe de travail. Notre évolution est liée à celle de l'équipe EA. L'obtention d'une qualification INSERM serait par exemple un moyen de pérenniser certains jeunes et de permettre des recrutements extérieurs de chercheurs statutaires. A nous de travailler pendant les années à venir (contrat quadriennal 2017) pour atteindre ce niveau d'exigence. De nos résultats scientifiques dépendra cette progression.

Peut-on prévoir au-delà un projet scientifique qui sera d'autant plus aléatoire qu'il dépend de variables extérieures ? Il s'agit en particulier d'une structuration scientifique et cohérente de l'épidémiologie clinique à Nantes. Une solution abordée dans le dernier chapitre est l'extension à d'autres cohortes.

Il peut s'agir d'extensions autour de l'insuffisance rénale avec d'autres bases de données comme le registre national REIN (patients en dialyse) ou comme DIVA-Neph (patients en insuffisance rénale chronique, M. Hourmant). Les travaux épidémiologiques retraçant les trajectoires complètes des patients en insuffisance rénale seront très riches. De telles études sont peu nombreuses car les systèmes d'information sont souvent cloisonnés. En misant sur l'avenir de l'interconnexion de ces réseaux, la modélisation de ces trajectoires nécessitera de développer des modèles statistiques adaptés.

Il peut aussi s'agir de l'extension vers d'autres pathologies. Cette ouverture est importante. Je pense en particulier à la greffe de moelle osseuse (M. Mothy), à la cancérologie (J.M. Classe), à la sclérose en plaque (D. Laplaud) ou aux prothèses valvulaires cardiaques (J.C. Roussel). Ces extensions vont bien sur lever à leur tour d'autres difficultés méthodologiques qui viendront nourrir notre propre recherche.

Ces projets ne devront pas remettre en question nos collaborations étroites avec l'équipe INSERM U1064 de S. Brouard et la recherche de biomarqueurs. Cette recherche translationnelle est très structurée grâce à une équipe internationalement reconnue et à la biocollection attenante à DIVAT. C'est à ce stade de structuration que la Biostatistique devient centrale, faisant le lien entre la biologie fondamentale et les retombées en termes de médecine personnalisée et de santé publique.

Pour conclure sur un plan plus personnel, et bien que les projets liés à l'analyse de données longitudinales incomplètes m'apparaissent toujours aussi passionnants, je pense avoir besoin de changements pour conserver la motivation et le dynamisme indispensables à tout chercheur. Avant d'envisager cette seconde période, il reste important de consolider notre thématique et son groupe de travail.

Table des figures

Liste des tableaux

Bibliographie

[1] MS Abouljoud, MH Deierhoi, SL Hudson, and AG Diethelm. Risk factors affecting second renal transplant outcome, with special reference to primary allograft nephrectomy. *Transplantation*, 60(2) :138–144, Jul 1995. 138, 145

[2] W Adler and B Lausen. Bootstrap estimated true and false positive rates and ROC curve. *Computional Statistics and Data Analysis*, 53 :718–729, 2009. 37

[3] R Aguirre-Hernandez and VT Farewell. A Pearson-type goodness-of-fit test for stationary and time-continuous Markov regression models. *Statistics in Medicine*, 21 :1899–1911, 2002. 90

[4] MG Akritas. Nearest neighbor estimation of a bivariate distribution under random censoring. *Ann Stat*, 22(3) :1299–1327, Sep 1994. 36, 66, 125

[5] AA Alizadeh, MB Eisen, RE Davis, C Ma, IS Lossos, A Rosenwald, JC Boldrick, H Sabet, T Tran, X Yu, JI Powell, L Yang, GE Marti, T Moore, J Hudson, L Lu, DB Lewis, R Tibshirani, G Sherlock, WC Chan, TC Greiner, DD Weisenburger, JO Armitage, R Warnke, R Levy, W Wilson, MR Grever, JC Byrd, D Botstein, PO Brown, and LM Staudt. Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling. *Nature*, 403(6769) :503–11, 2000. 33

[6] PS Almond, AJ Matas, K Gillingham, C Troppmann, W Payne, D Dunn, D Sutherland, and JS Najarian. Risk factors for second renal allografts immunosuppressed with cyclosporine. *Transplantation*, 52(2) :253–258, Aug 1991. 137

[7] C Ambroise and GJ McLachlan. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proceedings of the National Academy of Sciences*, 99(10) :6562–6566, 2002. 37

[8] PK Andersen. Multistate models in survival analysis : a study of nephropathy and mortality in diabetes. *Statistics in Medicine*, 7 :661–670, 1988. 89

[9] PK Andersen, K Borch-Johnsen, T Deckert, A Green, P Hougaard, N Keiding, and S Kreiner. A cox regression model for the relative mortality and its application to diabetes mellitus survival data. *Biometrics*, 41(4) :921–932, Dec 1985. 138

[10] LR Arends, MGM Hunink, and T Stijnen. Meta-analysis of summary survival curve data. *Statistics In Medicine*, 27 :4381–96, 2008. 48, 54

[11] M Arnol, JC Prather, A Mittalhenkle, JM Barry, and DJ Norman. Long-term kidney regraft survival from deceased donors : risk factors and outcomes in a single center. *Transplantation*, 86(8) :1084–1089, Oct 2008. 137, 138, 145

[12] J Ashton-Chess, E Dugast, RB Colvin, M Giral, Y Foucher, A Moreau, K Renaudin, C Braud, A Devys, S Brouard, and Soulillou JP. Regulatory, effector, and cytotoxic t cell profiles in long-term kidney transplant patients. *J Am Soc Nephrol*, 20 :1113–22, 2009. 30

[13] J Ashton-Chess, HL Mai, V Jovanovic, K Renaudin, Y Foucher, M Giral, A Moreau, E Dugast, M Mengel, M Racapé, R Danger, C Usal, H Smit, M Guillet, W Gwinner, L Le Berre, J Dantal, JP Soulillou, and S Brouard. Immunoproteasome beta subunit 10 is increased in chronic antibody-mediated rejection. *Kidney Int*, 77 :880–90, 2010. 30

[14] V Bagdonavicius and M Nikulin. *Accelerated Life Models*. Chapman and Hall-CRC, 2002. 81, 94

[15] HM Bovelstad, S Nygard, HL Storvold, M Aldrin, O Borgan, A Frigessi, and OC Lingjaerde. Predicting survival from microarray data - a comparative study. *Bioinformatics*, 23 :2080–7, 2007. 34

[16] DI Buckley, R Fu, M Freeman, K Rogers, and M Helfand. C-reactive protein as a risk factor for oronary heart disease : A systematic review and meta-analyses for the u .s. preventive services task force. *Annals of internal medicine*, 151 :483–495, 2009. 48

[17] JD Buckley. Additive and multiplicative models for relative survival rates. *Biometrics*, 40(1) :51–62, Mar 1984. 123, 138

[18] JM Campisto, J Albanell, W Arns, I Boletis, J Dantal, JW de Fijter, SA Mortensen, H Neumayer, O Oyen, J Pascual, E Pohanka, FP Schena, D Seron, V Sparacino, and JR Chapman. Use of proliferation signal inhibitors in the management of post-transplant malignancies - clinical guidance. *Nephrol. Dial. Transplant*, 22 :36–41, 2007. 123

[19] C Castelli, C Combescure, Y Foucher, and JP Daures. Cost-effectiveness analysis in colorectal cancer using a semi-Markov model. *Statistics in Medicine*, 26 :5557–5571, 2007. 27, 90

[20] J Cecka and P Terasaki. Scientific renal transplant registry : United network for organ sharing. *Clinical Transplants*, 21 :153–161, 2010. 112

[21] C Combescure, JP Daures, and Y Foucher. A literature-based approach to evaluate the predictive capacity of a marker using time-dependent summary receiver operating characteristics. *Stat Methods Med Res*, page in press, 2012. 29

[22] C Contal and J O'Quigley. An application of changepoint methods in studying the effect of age on survival in breast cancer. *Computational Statistics & Data Analysis*, 30 :253–70, 1999. 63

[23] DR Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2) :187–220, Jan 1972. 26, 35, 138

[24] DM Dabrowska, G Sun, and MM Horowitz. Cox Regression in a Markov Renewal Model : an application to the analysis of bone transplant data. *Journal of the American Statistical Association*, 89 :867–877, 1994. 89, 90

[25] R Danger and Y Foucher. Time dependent roc curves for the estimation of true prognostic capacity of microarray data. *Statistical Applications in Genetics and Molecular Biology*, 11, 2012. 30

[26] E de Azambuja, F Cardoso, G de Castro, M Colozza, M Mano, V Durbecq, C Sotiriou, D Larsimont, M Piccart-Gebhart, and M Paesmans. Ki-67 as prognostic marker in early breast cancer : a meta-analysis of published studies involving 12 155 patients. *British Journal of Cancer*, 96 :1504–1513, 2007. 48, 53

[27] KBG Dear. Iterative generalized least-squares for meta-analysis of survival-data at multiple times. *Biometrics*, 50 :989–1002, 1994. 48

[28] N Degauque, F Boeffard, Y Foucher, C Ballet, S Brouard, and JP Soulillou. The blood of healthy individuals exhibits cd8 t cells with a highly altered tcr vb repertoire but with an unmodified phenotype. *PLoS One*, 6 :e21240, 2011. 30

[29] PW Dickman, A Sloggett, M Hills, and T Hakulinen. Regression models for relative survival. *Stat Med*, 23(1) :51–64, Jan 2004. 123, 138

[30] M Dowsett, TO Nielsen, and R A'Hern. Assessment of ki67 in breast cancer : Recommendations from the international ki67 in breast cancer working group. *J Natl Cancer Inst*, 103 :1–9, 2011. 55

[31] L Duchateau, L Collette, R Sylvester, and JP Pignon. Estimating number of events from the kaplan-meier curve for incorporation in a literature-based meta-analysis : What you don't see you can't get ! *Biometrics*, 56 :886–92, 2000. 56

[32] L Duchateau, JP Pignon, L Bijnens, S Bertin, J Bourhis, and R Sylvester. Individual patient-versus literature-based meta-analysis of survival data : Time to event and event rate at a particular time can make a difference, an example based on head and neck cancer. *Controlled Clinical Trials*, 22 :538–547, 2001. 48

[33] F Ederer, LM Axtell, and SJ Cutler. The relative survival rate : a statistical methodology. *Natl Cancer Inst Monogr*, 6 :101–121, 1961. 123, 138

[34] B Efron. Estimating the error rate of a prediction rule - improvement of cross-validation. *Journal of tha American Statistical Association*, 78(382) :316–331, 1983. 38

[35] B Efron. Better bootstrap confidence intervals (with discussion). *Journal of the American Statistical Association*, 82 :171–200, 1987. 66

[36] B Efron and R Tibshirani. Improvements on cross-validation : The .632+ bootstrap method. *Journal of tha American Statistical Association*, 92(438) :548–560, 1997. 38

[37] U Elkayam and F Bitar. Valvular heart disease and pregnancy : part ii : prosthetic valves. *J Am Coll Cardiol*, 46 :403–10, 2005. 153

[38] S Escolano, JL Golmard, AM Korinek, and A Mallet. A multi-state model for evolution of intensive care unit patients : prediction of nosocomial infections and deaths. *Stat Med*, 19(24) :3465–3482, Dec 2000. 90

[39] J Estève, E Benhamou, M Croasdale, and L Raymond. Relative survival and the estimation of net survival : elements for further discussion. *Stat Med*, 9(5) :529–538, May 1990. 117

[40] ME Flowers, Y Inamoto, PA Carpenter, SJ Lee, HP Kiem, EW Petersdorf, SE Pereira, RA Nash, M Mielcarek, ML Fero, EH Warren, JE Sanders, RF Storb, FR Appelbaum, BE Storer, and PJ Martin. Comparative analysis of risk factors for acute graft-versus-host disease and for chronic graft-versus-host disease according to national institutes of health consensus criteria. *Blood*, 117 :3214–9, 2011. 149

[41] Y Foucher. *Modèles semi-markoviens : Application à l'analyse de l'évolution de pathologies chroniques (PhdThesis)*. PhD thesis, Université Montpellier 1, http ://tel.archives-ouvertes.fr/tel-00267683/fr/, October 2007. 91

[42] Y Foucher, C Combescure, J Ashton-Chess, and M Giral. Prognostic markers : data misinterpretation often leads to overoptimistic conclusions. *Am J Transplant*, 12(4) :1060–1061, 2012. 29, 111, 150

[43] Y Foucher, P Daguin, A Akl, M Kessler, M Ladrière, C Legendre, H Kreis, L Rostaing, N Kamar, and G Mourad. A clinical scoring system highly predictive of long-term kidney graft survival. *Kidney International*, 78(12) :1288–1294, 2010. 29, 111, 124

[44] Y Foucher, P Daguin, M Kessler, M Ladrière, C Legendre, H Kreis, D Durand, L Laveyssiere, G Mourad, V Garrigue, JP Daurès, JP Soulillou, and M Giral. How to evaluate the long-term prognostic capacity of pre-graft variables in transplantation ? *Clinical Transplant*, pages 113–8, 2008. 29

[45] Y Foucher, M Giral, JP Soulillou, and JP Daures. A semi-Markov model for multistate and interval-censored data with multiple terminal events. Application in renal transplantation. *Statistics in Medicine*, 26 :5381–5393, 2007. 27, 90, 119

[46] Y Foucher, M Giral, JP Soulillou, and JP Daurès. A flexible semi-markov model for interval-censored data and goodness-of-fit testing. *Statistical Methods in Medical Research*, 19 :127–45, 2010. 27, 81, 119, 152

[47] Y Foucher, M Giral, JP Soulillou, and JP Daurès. Time-dependent roc analysis for a three-class prognostic with application to kidney transplantation. *Stat. Med*, 30 :3079–87, 2010. 29, 119, 151, 153

[48] Y Foucher, M Giral, JP Soulillou, and JP Daurès. Cut-off estimation and medical decision making based on a continuous prognostic factor : The prediction of kidney graft failure. *The International Journal of Biostatistics*, 7, 2011. 29

[49] Y Foucher, E Mathieu, P Saint-Pierre, JF Durand, and JP Daurès. A semi-markov model based on generalized weibull distribution with an illustration for hiv disease. *Biometrical Journal*, 47(6) :825–833, 2005. 27, 76, 77, 90, 114, 119

[50] Y Foucher, P Saint-Pierre, P Puglièse, and JP Daurès. A semi-markov frailty model for multistate survival data : Illustration on hiv disease. *Far East Journal of Theoretical Statistics*, 47(19) :185–201, 2006. 27

[51] WJJ Fu, RJ Carroll, and SJ Wang. Estimating misclassification error with small samples via bootstrap cross-validation. *Bioinformatics*, 21(9) :1979–1986, 2005. 37

[52] M Gail and R Pfeiffer. On criteria for evaluating models of absolute risk. *Biostatistics*, 6 :227–39, 2005. 73

[53] M Giral, JP Bertola, Y Foucher, D Villers, E Bironneau, Y Blanloeil, G Karam, P Daguin, L Lerat, and JP Soulillou. Effect of brain-dead donor resuscitation on delayed graft function : results of a monocentric analysis. *Transplantation*, 83(9) :1174–1181, May 2007. 91

[54] M Giral, C Taddei, JM Nguyen, J Dantal, M Hourmant, D Cantarovich, G Blancho, D Ancelet, and JP Soulillou. Single-center analysis of 468 first cadaveric kidney allografts with a uniform ATG-CsA sequential therapy. *Clinical Transplantation*, pages 257–64, 1996. 63

[55] DW Gjertson. A multi-factor analysis of kidney regraft outcomes. *Clin Transpl*, pages 335–349, 2002. 138, 145

[56] J Goeman. L1 penalized estimation in the cox proportional hazards model. *Biometrical Journal*, 52(1) :70–84, 2010. 37, 43

[57] PM Grambsch and TM Therneau. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81 :515–526, 1994. 140

[58] J Gui and H Li. Penalized Cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data. *Bioinformatics*, 21 :3001–8, 2005. 40

[59] P Guyot, AE Ades, M Ouwens, and N Welton. Enhanced secondary analysis of survival data : reconstructing the data from published. *BMC Medical Research Methodology*, 12(9), 2012. 56

[60] T Hakulinen and L Tenkanen. Regression analysis of relative survival rates. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 36(3) :309–317, Jan 1987. 123, 138

[61] S Hariharan, MA McBride, WS Cherikh, CB Tolleris, BA Bresnahan, and CP Johnson. Post-transplant renal function in the first year predicts long-term kidney transplant survival. *Kidney Int*, 62(1) :311–318, Jul 2002. 63, 69

[62] X He, CE Metz, BM Tsui, JM Links, and EC Frey. Three-class roc analysis-a decision theoretic approach under the ideal observer framework. *IEEE Transactions on Medical Imaging*, 25 :571–581, 2006. 76

[63] PJ Heagerty, T Lumley, and SP Pepe. Time-Dependent ROC Curves for Censored Survival Data and a Diagnostic Marker. *Biometrics*, 56 :337–44, 2000. 29, 34, 35, 48, 52, 64, 65, 66, 69, 75, 124, 125, 135

[64] PJ Heagerty and Y Zheng. Survival model predictive accuracy and roc curves. *Biometrics*, 61 :92–105, 2005. 64, 128, 135

[65] PS Heckerling. Parametric three-way receiver operating characteristic surface analysis using mathematica. *Medical Decision Making*, 21 :409–417, 2001. 76

[66] D Hernandez, M Rufino, S Bartolomei, V Lorenzo, A Gonzalez-Rinne, and A Torres. A novel prognostic index for mortality in renal transplant recipients after hospitalization. *Transplantation*, 79 :337–43, 2005. 123

[67] D Hernandez, A Sanchez-Fructuoso, JM Gonzalez-Posada, M Arias, JM Campistol, M Rufino, JM Morales, F Moreso, G Perez, A Torres, and D Seron. A novel risk score for mortality in renal transplant recipients beyond the first posttransplant year. *Transplantation*, 88 :803–809, 2009. 22, 123, 131, 133, 135, 166

[68] JPT Higgins and SG Thompson. Quantifying heterogeneity in a meta-analysis. *Statistics In Medicine*, 21 :1539–1558, 2002. 55

[69] T Hothorn and A Zeileis. Generalized Maximally Selected Statistics. *Biometrics*, 2008. 64, 67, 71, 114

[70] Y Huang. Censored regression with the multistate accelerated sojourn times model. *Journals of the Royal Statistical Society B*, 64 :17–29, 2002. 90

[71] WD Irish, JN Ilsley, MA Schnitzler, S Feng, and DC Brennan. A risk prediction model for delayed graft function in the current era of deceased donor renal transplantation. *American Journal of Transplantation*, 10 :2279–2286, 2010. 112

[72] H Ishwaran, UB Kogalur, EZ Gorodeski, AJ Minn, and MS Lauer. High-dimensional variable selection for survival data. *Journal of the Americain Statistical Association*, 105 :205–217, 2010. 46

[73] JP Jais, C Haioun, TJ Molina, DS Rickman, A de Reynies, F Berger, C Gisselbrecht, J Briere, F Reyes, P Gaulard, P Feugier, E Labouyrie, H Tilly, C Bastard,

B Coiffier, G Salles, and K Leroy. The expression of 16 genes related to the cell of origin and immune response predicts survival in elderly patients with diffuse large b-cell lymphoma treated with chop and rituximab. *Leukemia*, pages 1–8, 2008. 33

[74] FM Jeejeebhoy. Prosthetic heart valves and management during pregnancy. *Canadian Family Physician*, 55 :155–157, 2009. 153

[75] TK Jenssen, WP Kuo, T Stokke, and Hovig E. Associations between gene expressions in breast cancer and patient survival. *Human Genetics*, 111(4-5) :411–20, 2002. 33

[76] W Jiang and R Simon. A comparison of bootstrap methods and an adjusted bootstrap approach for estimating the prediction error in microarray classification. *Statistics in Medicine*, 26 :5320–34, 2007. 45

[77] P Joly and D Commenges. A penalized likelihood approach for a progressive three-state model with censored and truncated data : Application to aids. *Biometrics*, 55 :887–890, Sept 1999. 84

[78] JD Kalbfleisch and JF Lawless. The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association*, 80 :863–871, 1985. 90

[79] M Kang and SW Lagakos. Statistical methods for panel data from a semi-Markov process, with application to HPV. *Biostatistics*, 8 :252–264, 2007. 90, 105

[80] B Kaplan, J Schold, and HU Meier-Kriesche. Poor predictive value of serum creatinine for renal allograft loss. *American Journal of Transplantation*, 3 :1560–1565, 2003. 76, 83

[81] EL Kaplan and P Meier. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*, 53(282) :457–481, Jun 1958. 26, 33, 140

[82] BL Kasiske, AK Israni, JJ Snyder, MA Skeans, Y Peng, and ED Weinhandl. A simple tool to predict outcomes after kidney transplant. *Am J Kidney Dis*, 56 :947–60, 2010. 111

[83] R Kay. A Markov model for analyzing cancer markers and disease states in survival studies. *Biometrics*, 7 :855–866, 1986. 89

[84] JP Klein and ML Moeschberger. *Survival analysis : techniques for censored and truncated data*. Springer-verlag, New York, 1997. 147

[85] PA Lachenbruch, AS Rosenberg, E Bonvini, MW Cavaille-Coll, and RB Colvin. Biomarkers and surrogate endpoints in renal transplantation : Present status and considerations for clinical trial design. *American Journal of Transplantation*, 4 :451–457, 2004. 48

[86] M Ladrière, Y Foucher, C Legendre, N Kamar, V Garrigue, E Morélon, M Kessler, JP Soulillou, and Giral M. The western europe cohort of kidney transplanted recipients - the divat network. *Clin Transpl*, pages 460–461, 2010. 141

[87] SW Lagakos, CJ Sommer, and M Zelen. Semi-Markov models for partially censored data. *Biometrika*, 65 :311–317, 1978. 89

[88] V Lagani and I Tsamardinos. Structure-based variable selection for survival data. *Bioinformatics*, 26 :1887–94, 2010. 46

[89] JF Lawless. *Statistical Models and Methods for Lifetime Data*. Wiley, 1982. 101

[90] M LeBlanc and J Crowley. Survival trees by goodness of split. *Journal of the American Statistical Association*, 88 :457–67, 1993. 63

[91] AS Levey, JP Bosch, JB Lewis, T Greene, N Rogers, and D Roth. A more accurate method to estimate glomerular filtration rate from serum creatinine : a new prediction equation. Modification of Diet in Renal Disease Study Group. *Annals of Internal Medicine*, 130 :461–70, 1999. 69, 80, 91

[92] H Li and J Gui. Partial Cox regression analysis for high-dimensional microarray gene expression data. *Bioinformatics*, 20 :i208–i215, 2004. 34, 46

[93] N Li, R Elashoff, and G Li. Robust joint modeling of longitudinal measurements and competing risks failure time data. *Biometrical Journal*, 51 :19–30, 2009. 76

[94] JK Lindsey. A study of interval censoring in parametric regression models. *Lifetime Data Analysis*, 4 :329–354, 1998. 90

[95] ML MacMillan, TE DeFor, and DJ Weisdorf. What predicts high risk acute graft-versus-host disease (gvhd) at onset ? identification of those at highest risk by a novel acute gvhd risk score. *Br J Haematol*, 157 :732–41, 2012. 150

[96] JC Magee, ML Barr, GP Basadonna, MR Johnson, S Mahadevan, MA McBride, DE Schaubel, and AB Leichtman. Repeat organ transplantation in the united states, 1996-2005. *Am J Transplant*, 7 :1424–1433, 2007. 137

[97] E Mathieu, Y Foucher, P Dellamonica, and JP Daurès. A parametric and non homogeneous semi-markov process for hiv control. *Methodology and Computing in Applied Probability*, 47(3) :389–397, 2007. 27

[98] P McCullagh. *Generalized Linear Models*. Chapman and Hall, 1983. 97

[99] R Miller and D Siegmund. Maximally Selected Chi Square Statistics. *Biometrics*, 38 :1011–1016, 1982. 64

[100] AM Molinaro, R Simon, and RM Pfeiffer. Prediction error estimation : a comparison of resampling methods. *Bioinformatics*, 21(15) :3301–3307, 2005. 35, 37

[101] J Moore, X He, S Shabir, R Hanvesakul, D Benavente, P Cockwell, MA Little, S Ball, N Inston, A Johnston, and R Borrows. Development and evaluation of

a composite risk score to predict kidney transplant failure. *Am J Kidney Dis*, 57 :744–51, 2011. 111

[102] LE Moses, D Shapiro, and B Littenberg. Combining independent studies of a diagnostic-test into a summary ROC curve - data-analytic and some additional considerations. *Statistics In Medicine*, 12 :1293–316, 1993. 54

[103] D Mossman. Three-way roc. *Medical Decision Making*, 19 :78–89, 1999. 76

[104] D Nicol, A MacDonald, J Lawen, and P Belitsky. Early prediction of renal allograft loss beyond one year. *Transplantation International*, pages 153–7, 1993. 63

[105] A Ocana, F Vera-Badillo, B Seruga, A Templeton, A Pandiella, and E Amir. Her3 overexpression and survival in solid tumors : A meta-analysis. *J Natl Cancer Inst*, 2012. 155

[106] AO Ojo, RA Wolfe, LY Agodoa, PJ Held, FK Port, SF Leavey, SE Callard, DM Dickinson, RL Schmouder, and AB Leichtman. Prognosis after primary renal transplant failure and the beneficial effects of repeat transplantation : multivariate analyses from the united states renal data system. *Transplantation*, 66(12) :1651–1659, Dec 1998. 137

[107] J Orbe, E Ferreira, and V Nunez-Anton. Comparing proportional hazards and accelerated failure time models for survival analysis. *Stat Med*, 21(22) :3493–3510, Nov 2002. 96

[108] AD Oxman, MJ Clarke, and LA Stewart. Meta-analyses Using Individual Patient Data Are Needed . *Journal of the American Medical Association*, 274 :845–846, 1995. 48

[109] W Pan and R Chappell. A nonparametric estimator of survival functions for arbitrarily truncated and censored data. *Lifetime Data Analysis*, 4 :187–202, 1998. 36

[110] MY Park and T Hastie. L1 regularization path algorithm for generalized linear models. *Journal of the Royal Statistical Society : Series B*, 69 :659–677, 2007. 37

[111] MKB Parmar, V Torri, and L Stewart. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Statistics in Medicine*, 17(24) :2815–2834, 1998. 47, 51, 54, 56

[112] MS Pepe, Y Zheng, Y Jin, Y Huang, C Parikh, and W Levy. Evaluating the roc performance of markers for future events. *Lifetime Data Analysis*, 14 :86–113, 2008. 76

[113] R Perez-Ocon and JE Ruiz-Castro. *Semi-Markov Models and Applications*. Kluwer Academic Publishers, Dordrecht, Netherlands, 1999. 76, 77, 89, 90

[114] R Perez-Ocon, JE Ruiz-Castro, and ML Gamiz-Perez. A piecewise Markov process for analysing survival from breast cancer in different risk groups. *Statistics in Medicine*, 20 :109–122, 2001. 101

[115] MP Perme, J Stare, and J Estève. On estimation in relative survival. *Biometrics*, 68 :113–20, 2012. 124, 135, 138

[116] Jose Pinheiro, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and the R Core team. *nlme : Linear and Nonlinear Mixed Effects Models*, 2009. R package version 3.1-93. 50

[117] S Pocock, T Clayton, and D Altman. Survival plots of time-to-event outcomes in clinical trials : good practice and pitfalls. *The Lancet*, 359 :1686–1689, 2002. 51

[118] M Pohar and J Stare. Goodness of fit of relative survival models. *Stat. Med*, 24 :3911–25, 2005. 124

[119] M Pohar and J Stare. Relative survival analysis in r. *Comput Methods Programs Biomed*, 3 :272–8, 2006. 124

[120] M Pohar and J Stare. Making relative survival analysis relatively easy. *Comput Biol Med*, 37 :1741–9, 2007. 124

[121] T Poisot. The digitize Package : Extracting Numerical Data from Scatterplots. *The R Journal*, 3(1) :25–26, June 2011. 51

[122] C Porzelius, M Schumacher, and H Binder. A general, prediction error-based criterion for selecting model complexity for high-dimensional survival models. *Statistics in Medicine*, 29 :830–8, 2010. 34

[123] RL Prentice, JD Kalbfleisch, AV Peterson, N Flournoy, VT Farewell, and NE Breslow. The analysis of failure times in the presence of competing risks. *Biometrics*, 34(4) :541–554, 1978. 76

[124] R Development Core Team. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2009. ISBN 3-900051-07-0. 50, 113, 140

[125] M Racapé, JP Duong Van Huyen, R Danger, M Giral, F Bleicher, Y Foucher, A Pallier, P Pilet, P Tafelmeyer, J Ashton-Chess, E Dugast, S Pettré, B Charreau, JP Soulillou, and S Brouard. The involvement of smile/tmtc3 in endoplasmic reticulum stress response. *PLoS One*, 6 :e19321, 2011. 30

[126] M Railo, S Nordling, K von Boguslawsky, M Leivonen, L Kyllonen, and K von Smitten. Prognostic value of Ki-67 immunolabelling in primary operable breast cancer. *British Journal of Cancer*, 68 :579–583, 1993. 54

[127] DF Ransohoff. Opinion - Rules of evidence for cancer molecular-marker discovery and validation. *Nature Reviews Cancer*, 4(4) :309–314, 2004. 37

[128] PS Rao, DE Schaubel, G Wei, and SS Fenton. Evaluating the survival benefit of kidney retransplantation. *Transplantation*, 82(5) :669–674, Sep 2006. 137

[129] S Ripatti and J Palmgren. Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, 56(4) :1016–1022, 2004. 141

[130] A Rosenwald, G Wright, WC Chan, JM Connors, E Campo, RI Fisher, RD Gascoyne, HK Muller-Hermelink, EB Smeland, JM Giltnane, EM Hurt, H Zhao, L Averett, L Yang, WH Wilson, ES Jaffe, R Simon, RD Klausner, J Powell, PL Duffey, DL Longo, TC Greiner, DD Weisenburger, WG Sanger, BJ Dave, JC Lynch, J Vose, JO Armitage, E Montserrat, A Lopez-Guillermo, TM Grogan, TP Miller, M LeBlanc, G Ott, S Kvaloy, J Delabie, H Holte, P Krajci, T Stokke, and LM Staudt. The use of molecular profiling to predict survival after chemotherapy for diffuse large-b-cell lymphoma. *New England Journal of Medicine*, 346(25) :1937–47, 2002. 33, 34, 40, 45

[131] V Rousseau, Y Foucher, M Giral, JP Soulillou, and JP Daurès. Informative censoring in a multiplicative relative survival model : application for kidney transplant recipients. *JP Journal of Biostatistics*, 3 :195–214, 2009. 28

[132] P Saha and PJ Heagerty. Time-dependent predictive accuracy in the presence of competing risks. *Biometrics*, 66 :999–1011, 2010. 76, 78, 84, 124, 151

[133] M Schumacher, H Binder, and T Gerds. Assessment of survival prediction models based on microarray data. *Bioinformatics*, 23 :1768–1774, 2007. 34, 37, 45

[134] J Sebaugh, J Wilson, M Tucker, and W Adams. A Study of the Shape of Dose-Response Curves for Acute Lethality at Low Response : A Megadaphnia Study. *Risk Analysis*, 11 :633–640, 1991. 63

[135] MR Segal. Microarray gene expression data with linked survival phenotypes : diffuse large-B-cell lymphoma revisited. *Bioinformatics*, 7 :268–285, 2005. 40, 41, 45

[136] R Simon, MD Radmacher, K Dobbin, and LM McShane. Pitfalls in the Use of DNA Microarray Data for Diagnostic and Prognostic Classification. *Journal of the National Cancer Institute*, 95 :14–18, 2003. 34, 43, 45

[137] R Simon, J Subramanian, MC Li, and S Menezes. Using cross-validation to evaluate predictive accuracy of survival risk classifiers based on high-dimensional data. *Briefings in Bioinformatics*, 12 :203–14, 2011. 39, 46

[138] X Song and XH Zhou. A semiparemtric approach for the covariate specific roc curve with survival outcome. *Statistica Sinica*, 18 :947–965, 2008. 84

[139] T Sorlie, R Tibshirani, J Parker, T Hastie, JS Marron, A Nobel, S Deng, H Johnsen, R Pesich, S Geisler, J Demeter, CM Perou, PE Lonning, PO Brown,

AL Borresen-Dale, and D Botstein. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci*, 100 :8418–23, 2003. 33

[140] DM Stablein, WH Carter, and JW Novak. Analysis of survival data with nonproportional hazard functions. *Control Clin Trials*, 2(2) :149–159, Jun 1981. 97

[141] J Stare, M Pohar, and R Henderson. Goodness of fit of relative survival models. *Stat Med*, 24(24) :3911–3925, Dec 2005. 117

[142] M Sternberg and G Satten. Discrete time nonparametric estimation for chain of events data subject to interval censoring and truncation. *Biometrics*, 55 :514–522, 1999. 90

[143] RJ Stratta, CS Oh, HW Sollinger, JD Pirsch, M Kalayoglu, and FO Belzer. Kidney retransplantation in the cyclosporine era. *Transplantation*, 45(1) :40–45, Jan 1988. 137

[144] A Thibault-Espitia, Y Foucher, R Danger, T Migone, A Pallier, S Castagnet, C Gueguen, A Devys, A Gautier, M Giral, JP Soulillou, and S Brouard. Baff and baff-r levels are associated with risk of long-term kidney graft dysfunction and development of donor-specific antibodies. *Am J Transplant*, 6 :2754–62, 2012. 30

[145] R Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society : Series B*, 58 :267–288, 1996. 35

[146] J Tierney, L Stewart, D Ghersi, S Burdett, and M Sydes. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials*, 8 :1–16, 2007. 47

[147] AC Titman and LD Sharples. A general goodness-of-fit test for Markov and hidden Markov models. *Statistics in Medicine*, DOI :10.1002/sim.3033, 2007. 90

[148] K Trébern-Launay, Y Foucher, M Giral, C Legendre, H Kreis, M Kessler, M Ladrière, N Kamar, L Rostaing, V Garrigue, G Mourad, E Morelon, JP Soulillou, and J Dantal. Poor long-term outcome in second kidney transplantation : a delayed event. 2012. 137, 143

[149] H Trihia, S Murray, K Price, RD Gelber, R Golouh, A Goldhirsch, AS Coates, J Collins, M Castiglione-Gertsch, and Gusterson BA. Ki-67 expression in breast carcinoma. Its association with grading systems, clinical parameters, and other prognostic factors - A surrogate marker ? *Cancer*, 97 :1321–31, 2003. 54

[150] LJ Van't Veer, H Dai, MJ van de Vijver, YD He, AA Hart, M Mao, HL Peterse, K van der Kooy, MJ Marton, AT Witteveen, GJ Schreiber, RM Kerkhoven, C Roberts, PS Linsley, R Bernards, and SH Friend. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415 :530–536, 2002. 33

[151] PJ Verweij and HC Van Houwelingen. Penalized likelihood in cox regression. *Statistics in medicine*, 13 :15–30, 1994. 46

[152] AJ Vickers and EB Elkin. Decision curve analysis : a novel method for evaluating prediction models. *Medical Decision Making*, 26 :565–74, 2006. 71, 73

[153] SD Walter. Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. *Statistics In Medicine*, 21 :1237–56, 2002. 54, 55

[154] D Wang, TZ Xu, JH Chen, WZ Wu, SL Yang, WH Lin, JQ Cai, and JM Tan. Factors influencing second renal allograft survival : a single center experience in china. *Transpl Immunol*, 20(3) :150–154, Jan 2009. 137, 138, 145

[155] PR Williamson, C Tudur Smith, JL Hutton, and AG Marson. Aggregate data meta-analysis with time-to-event outcomes. *Statistics in medicine*, 21 :3337–51, 2002. 47, 56

[156] RA Wolfe, VB Ashby, EL Milford, AO Ojo, RE Ettenger, LY Agodoa, PJ Held, and FK Port. Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. *N Engl J Med*, 341(23) :1725–1730, Dec 1999. 118

[157] RA Wolfe, KP McCullough, and AB Leichtman. Predictability of survival models for waiting list and transplant patients : calculating lyft. *Am J Transplant*, 9(7) :1523–1527, Jul 2009. 137

[158] SG Yarlagadda, SG Coca, RN Formica, ED Poggio, and CR Parikh. Association between delayed graft function and allograft and patient survival : A systematic review and meta-analysis. *Nephrol Dial Transplant*, 24 :1039–1047, 2008. 112

[159] L Yu, L Deng, J Li, Y Zhang, and L Hu. The prognostic value of vascular endothelial growth factor in ovarian cancer : A systematic review and meta-analysis. *Gynecol Oncol*, 2012. 155

[160] Y Zheng and PJ Heagerty. Semiparametric estimation of time-dependent roc curves for longitudinal marker data. *Biostatistics*, 5 :615–632, 2004. 75

[161] Y Zheng and PJ Heagerty. Prospective Accuracy for Longitudinal Markers. *Biometrics*, 63 :332–41, 2007. 74, 75

[162] XH Zhou, NA Obuchowski, and DK McClish. *Statistical Methods in Diagnostic Medicine*. Wiley-Interscience, New-York, 2002. 75