

METHODES STATISTIQUES POUR PRENDRE EN COMPTE L'OCCURRENCE DE DONNEES MANQUANTES ALEATOIRES CONJOINTEMENT AVEC LA METHODE DU TEMPS JUSQU' A DETERIORATION D'UN SCORE DE QUALITE DE VIE : UNE ETUDE DE SIMULATION

Amélie Anota¹, Francesco Cottone², Fabio Efficace³ & Franck Bonnetain⁴

¹ *Unité de méthodologie et de qualité de vie en cancérologie (EA3181), Plateforme Nationale Qualité de Vie et Cancer, CHRU de Besançon, 2 place Saint-Jacques, 25000 Besançon; aanota@chu-besancon.fr*

² *Italian Group for Adult Hematologic Diseases (GIMEMA) Data Center and Health Outcomes Research Unit, 6 Via Benevento, 00161 Rome, Italy; f.cottone@gimema.it*

³ *Italian Group for Adult Hematologic Diseases (GIMEMA) Data Center and Health Outcomes Research Unit, 6 Via Benevento, 00161 Rome, Italy; f.efficace@gimema.it*

⁴ *Unité de méthodologie et de qualité de vie en cancérologie (EA3181), Plateforme Nationale Qualité de Vie et Cancer, CHRU de Besançon, 2 place Saint-Jacques, 25000 Besançon; franck.bonnetain@univ-fcomte.fr*

Résumé. L'objectif de ce projet était de comparer trois approches statistiques pour prendre en compte les données manquantes aléatoires conjointement avec la méthode du temps jusqu'à détérioration d'un score de qualité de vie (QdV) utilisée en cancérologie via une étude de simulations. Les méthodes comparées étaient les méthodes de probabilités inversées et d'appariement du score de propension ainsi qu'une méthode d'imputations multiples basée sur l'algorithme de Monte-Carlo par Chaîne de Markov. Ces méthodes ont été comparées selon le nombre d'évènements, la médiane de détérioration, l'erreur de type I et la puissance statistique du test du Log-rank, le coefficient Hazard Ratio mesurant l'effet traitement, le coefficient Kappa de Cohen pour comparer les évènements et le coefficient de corrélation de Spearman pour comparer les temps de survie. Les données de QdV longitudinales ont été simulées selon un modèle à crédit partiel longitudinal en considérant que la QdV suit une loi normale multivariée. Cinq items à 4 modalités de réponse ont été considérés, reflétant la dimension physique du questionnaire EORTC QLQ-C30. Cinq covariables associées à l'occurrence de données manquantes ont été simulées. Les simulations sont en cours d'implémentation et les résultats seront présentés lors du congrès.

Mots-clés. Cancérologie, qualité de vie relative à la santé, analyse longitudinale, données manquantes, temps jusqu'à détérioration

Abstract. The objective of this project was to compare three statistical approaches to take into account the occurrence of missing data at random in conjunction with the time to health-related quality of life (HRQoL) score deterioration method used in oncology using a simulation study. The methods compared were the inverse probability of treatment weighting method and the matching methods of the propensity score and a multiple imputations method based on the Monte Carlo Markov Chain

algorithm. These methods were compared according to the number of events, the median of deterioration, the type I error rate and the statistical power of the Log-rank test, the coefficient Hazard Ratio evaluating the treatment effect, the Cohen's Kappa coefficient to compare the events and the Spearman coefficient of correlation to compare the survival duration. The longitudinal HRQoL data were simulated according to a longitudinal partial credit model considering that HRQoL follows a multivariate normal distribution. Five items with 4 response categories were considered, reflecting the physical dimension of the EORTC QLQ-C30 questionnaire. Five covariates associated with the occurrence of missing data were simulated. Simulations are ongoing and the results would be presented at the time of the congress.

Keywords. Oncology, health-related quality of life, longitudinal analysis, missing data, time to deterioration

1 Introduction

Dans les essais cliniques en cancérologie, les résultats des données de qualité de vie relative à la santé (QdV) demeurent peu utilisés pour changer les standards de prise en charge des patients, particulièrement en raison de la complexité de l'analyse longitudinale de la QdV. Un enjeu majeur de l'analyse longitudinale de données de QdV est l'occurrence des données manquantes (DM) qui peuvent entraîner un biais dans l'analyse si elles ne sont pas correctement prises en compte, comme démontré par Fairclough (1998 ; 2010). Dans les études de QdV, la majorité des DM sont :

- soient aléatoires (« Missing At Random », MAR) selon la classification de Little et Rubin (1987), i.e. dépendent de covariables observées du patient telles que les caractéristiques socio-démographiques et cliniques du patients recueillies à l'inclusion ;
- soient non aléatoires (« Missing Not At Random », MNAR) i.e. dépendent du niveau de QdV manquant. Une part importante des DM MNAR sont les sorties d'étude précoces (avant l'arrêt planifié de l'étude) en raison d'une détérioration de l'état de santé du patient ou du décès.

Seules les DM complètement aléatoires (« Missing Completely At Random », MCAR) sont non informatives et n'entraînent pas de biais dans l'analyse. Les DM MAR sont non informatives du niveau de QdV du patient conditionnellement aux covariables associées à l'occurrence de DM. Les DM MNAR sont informatives et peuvent entraîner un biais dans l'analyse en surestimant par exemple le niveau de QdV du patient. Des méthodes utilisant le score de propension sont souvent utilisées dans les études observationnelles pour réduire le biais dû à l'absence de randomisation, comme illustré par Trojano (2009). La méthode de probabilités inversées (IPTW) est la méthode recommandée pour les modèles de survie par Austin (2013). Ces méthodes ont également été investiguées pour tenir compte de l'occurrence de DM comme l'a démontré Seaman (2013). Des méthodes d'imputations multiples sont également souvent utilisées en présence de DM MAR tel que par Mojita (2005).

Depuis quelques années, des modèles de type « temps jusqu'à évènement » tels que le temps jusqu'à détérioration (TJD) d'un score de QdV ont été proposés comme modalité d'analyse longitudinale en cancérologie par Bonnetain (2010), Hamidou (2011) et Anota (2015a). Ces modèles ont l'avantage de proposer des résultats cliniquement pertinents tels que des courbes de survie estimées selon la méthode de Kaplan-Meier et un effet traitement résumé sous forme d'Hazard Ratio. Dans les définitions de TJD généralement utilisées, les DM sont ignorées considérant le niveau de QdV du patient constant entre deux mesures de QdV disponibles, supposant ainsi les DM MCAR. Or, dans ce cas, la censure peut être informative de l'état de santé du patient ce qui peut biaiser l'analyse. Ainsi, il paraît nécessaire de proposer des méthodes statistiques permettant de prendre en compte l'occurrence des DM MAR et MNAR conjointement avec la méthode du TJD. Un premier travail a ainsi été mené en investiguant la méthode IPTW conjointement avec la méthode du TJD dans un essai clinique de phase II par Anota (2015b). Une étude de simulation est donc indispensable pour valider l'utilisation de cette méthode.

Dans cette optique, l'objectif de ce projet est d'évaluer et de comparer trois approches statistiques conjointement avec la méthode du TJD pour tenir compte de l'occurrence de DM MAR :

- les méthodes de probabilités inversées et d'appariement du score de propension
- et une méthode d'imputations multiples basée sur l'algorithme de Monte-Carlo par Chaîne de Markov.

De plus, les analyses seront répétées en présence de DM monotones de type MNAR, correspondant à des patients décédés durant l'étude dans le but de refléter les conditions réelles des essais cliniques en cancérologie.

2 Méthodes

2.1 Simulations des données longitudinales de qualité de vie

Nous avons considéré que la QdV était évaluée selon 5 items construits selon une échelle de Likert à 4 modalités de réponse afin de refléter la structure des questionnaires de QdV de l'EORTC dont le QLQ-C30 développé par Aaronson (1993). La simulation est réalisée en deux étapes :

1. le niveau de QdV latent est simulé selon une loi normale multivariée avec une matrice de variance-covariance autorégressive de premier ordre. La simulation est réalisée par bras de traitement en introduisant ou non un effet bras de traitement.
2. les réponses aux items sont obtenues en utilisant un modèle longitudinal à crédit partiel issu de la théorie de réponse à l'item. Ainsi, la probabilité pour un individu n de choisir la modalité de réponse k parmi les 4 réponses possibles à l'item j au temps t connaissant le niveau de QdV latent $\theta_n^{(t)}$ et les paramètres de difficulté $\delta_{j,1}, \dots, \delta_{j,m_j}$ de l'item est :

$$P\left(X_{n,j} = k \mid \theta_n^{(t)}, \delta_{j,1}, \dots, \delta_{j,m_j}\right) = \frac{\exp(k\theta_n^{(t)} - \sum_{i=1}^k \delta_{j,i})}{\sum_{h=1}^{m_j} \exp(h\theta_n^{(t)} - \sum_{i=1}^h \delta_{j,i})}$$

Après obtention des réponses aux items, un score variant de 0 à 100 est généré comme la proposé Fayers (2001) pour les questionnaires de l'EORTC. Pour un individu n et une échelle construite avec J items, le score X_n est :

$$X_n = \frac{\sum_j X_{n,j}}{J}$$

2.2 Simulations des données manquantes aléatoires

Cinq covariables associées à l'occurrence de DM sont simulées selon Cottone (2013) :

- Variable âge : $X_1 = 18 + (50 \times K)$, $K \sim Beta(3, 2)$
- Variable sexe : $X_2 \sim Binomial(0.5)$
- Variable #3 : $X_3 \sim Binomial(0.5)$
- Variable #4 : $X_4 \sim 5 \times J$, $J \sim Beta(1, 4.6)$
- Variable #5 : $X_5 \sim Binomial(0.5)$

La probabilité pour chaque sujet de présenter au moins une DM est une variable aléatoire MD suivant une loi binomiale de paramètre p_{MD^*} telle que :

$$p_{MD^*} = P(MD^* = 1 \mid \alpha, x) = \frac{\exp(\alpha_0 + \sum_{j=1}^5 \alpha_j x_j)}{1 + \exp(\alpha_0 + \sum_{j=1}^5 \alpha_j x_j)}$$

Les coefficients α_j pour $j \geq 1$ mesurent l'impact de chaque covariable sur la propension à présenter des DM. La constante α_0 est fixée de sorte d'avoir $k\%$ de patients avec des DM. Les DM sont ensuite générées pour les patients pour lesquels $MD = 1$, uniquement pour les mesures de suivi (la mesure de QdV initiale est toujours présente), selon le mécanisme suivant :

$$MD^t = \begin{cases} 1 & \text{si } t = t^*, t^* \sim U(2,5) \\ \text{Bin}(\pi^t) & \text{si } t \neq t^*, \pi^t \sim U(0,1) \end{cases}, \forall t > 1 \mid MD^* = 1$$

2.3. Méthode du temps jusqu'à détérioration

Différentes définitions de temps jusqu'à détérioration ont été proposées selon les situations thérapeutiques et les localisations cancéreuses. L'évènement considéré peut être défini selon le score de référence choisi, la différence minimale cliniquement importante (DMCI), les scores manquants et incluant ou non le décès comme évènement. De plus, la détérioration considérée peut être un état transitoire (généralement considéré en situation adjuvante) ou définitif (plus pertinent pour les situations avancées ou métastatiques).

Ainsi, deux définitions de détérioration ont été retenues pour ce projet :

- Le temps jusqu'à détérioration (TJD) en tant qu'état transitoire, défini selon Hamidou (2011) comme l'intervalle de temps entre la date d'inclusion du patient et l'observation d'une première détérioration d'au moins 5 points (DMCI=5) du score de QdV par rapport au score à l'inclusion. Les patients ne présentant pas de détérioration sont donc censurés au moment du remplissage du dernier questionnaire de QdV disponible.
- Le TJD définitive (TJDD) d'un score de QdV, défini selon Bonnetain (2010) comme l'intervalle de temps depuis la date d'inclusion et l'observation d'une première détérioration d'au moins 5 points du score de QdV par rapport au score à l'inclusion, sans amélioration ultérieure d'au moins 5 points par rapport au score à l'inclusion. Les patients ne présentant pas de détérioration, ou présentant une détérioration mais suivie d'une amélioration significative, sont censurés au moment du remplissage du dernier questionnaire de QdV disponible.

L'estimation du TJD/TJDD est réalisée selon la méthode d'estimation de Kaplan-Meier. Pour les simulations intégrant des DM monotones correspondant au décès de patients, l'évènement décès est pris en compte dans une analyse de sensibilité si aucune détérioration de QdV n'a été observée préalablement. De plus, les analyses seront également réitérées en utilisant la méthode de censure par intervalle en présence de DM afin d'investiguer l'impact de l'utilisation de cette méthode sur l'estimation du TJD.

2.4. Méthodes pour prendre en compte les données manquantes

2.4.1. Méthodes du score de propension

L'objectif est de déterminer les variables associées à l'occurrence des DM en comparant les patients présentant au moins une DM (répondeurs partiels) versus les patients avec l'ensemble des données disponibles jusqu'à leur sortie d'étude ou le décès (répondeurs complets) selon les covariables recueillies à l'inclusion. Les variables pour lesquelles une P-value ≤ 0.20 est obtenue en analyse de régression logistique univariée sont introduites dans le modèle multivarié. Toutes les variables recueillies à l'inclusion sont testées excepté la variable X_4 considérée comme non observée. Le score de propension est ensuite construit à partir des valeurs ajustées extraites du modèle multivarié.

Les méthodes IPTW et d'appariement du score de propension sont ensuite appliquées. La méthode IPTW consiste à réaliser les analyses de TJD en assignant un poids aux patients égal à l'inverse du score de propension pour les répondeurs partiels et à l'inverse de l'opposé du score de propension pour les répondeurs complets. Concernant la méthode d'appariement, les répondeurs complets et les répondeurs partiels sont appariés avec un ratio 1 :1 à partir du score de propension obtenu. Un modèle de Cox utilisant des techniques de modélisation de fragilité (« frailty models ») est ensuite réalisé pour estimer l'Hazard-Ratio correspondant selon le modèle de Hougaard (1995).

2.4.2. Méthode d'imputations multiples

La méthode d'imputations multiples choisie est basée sur l'algorithme de Monte-Carlo par Chaîne de Markov comme décrit par Azur (2011). Cinq imputations sont réalisées en tenant compte des covariables associées à l'occurrence des DM.

2.5. Critères de comparaison et scénarios considérés

Les critères de comparaison de ces différentes méthodes sont : 1) le nombre d'évènements considérés ; 2) la médiane de détérioration par bras de traitement avec son intervalle de confiance à 95% ; 3) le coefficient Hazard Ratio mesurant l'effet traitement avec son intervalle de confiance à 95% ; 4) le taux d'erreur de type I et la puissance statistique du test d'une différence entre les deux bras de traitement au cours du temps, selon le test du Log-rank ; 5) la concordance entre les évènements et censures entre les résultats obtenus par la méthode d'imputation multiple et les résultats sur les données complètes selon le coefficient Kappa de Cohen ; et 6) le coefficient de corrélation de Spearman pour comparer les temps de survie.

Les différents paramètres et scénarios considérés sont résumés dans le Tableau 1.

Tableau 1 : Paramètres utilisés dans l'étude de simulations et valeurs correspondantes

Paramètres	Valeurs
Nombre de temps de mesure	5 ; 10
Effet temps	linéaire : $\mu_0 = -0.4$ et $\mu_{fin} = 0.4$
Effet traitement	$\Delta = 0 \forall t$ ou $\Delta = 0$ pour $t = 0$; $\Delta = 0.4 \forall t > 1$
Corrélation entre les mesures de QdV	0.4 ; 0.7 ; 0.9
Matrice de variance-covariance	Autorégressive de premier ordre Symétrie composée
Nombre de patients	100 ; 200 ; 300 ; 500
Nombre d'items	5
Nombre de modalités de réponse par item	4
Pourcentage de patients présentant des DM MAR	10% ; 20% ; 30% ; 50% ; 60%

Ces paramètres ont été choisis en tenant compte des résultats d'une précédente étude de simulations réalisée par Anota (2014).

3 Résultats

Les simulations sont en cours d'implémentation sous le logiciel R (version 3.1) et les résultats seront présentés lors du congrès. Le temps de réalisation des simulations sera recueilli et une estimation du temps de calcul pour chaque scénario pourra également être donnée.

4 Conclusions et perspectives

Cette étude de simulations est la première étude comparant trois approches pour prendre en compte l'occurrence de DM MAR conjointement avec la méthode du TJD d'un score de QdV. D'autres méthodes pour prendre en compte les DM pourront également être investiguées, telle que la méthode « Rank Preserving Structural Failure Time » proposée par Robins (1992). Les résultats vont permettre

de proposer des recommandations sur les méthodes à utiliser selon les situations et le design de l'étude. Une analyse sur données réelles sera ensuite réalisée afin de compléter ces résultats.

Bibliographie

- [1] Fairclough D. L. (2010), Design and analysis of quality of life studies in clinical trials, *CRC press*.
- [2] Fairclough D. L., Peterson H. F. et Chang V. (1998) Why are missing quality of life data a problem in clinical trials of cancer therapy?, *Stat Med*, 17(5-7):667-77.
- [3] Little R. J. et Rubin D. B. (1987), Statistical analysis with missing data. *New York: John Wiley & Sons*.
- [4] Trojano M., Pellegrini F., Paolicelli D., Fuiani A., et Di Renzo V. (2009) Observational studies: propensity score analysis of non-randomized data. *Int MS J*. 16(3):90-7.
- [5] Austin P. C. (2013), The performance of different propensity score methods for estimating marginal hazard ratios. *Stat Med*, 32(16):2837-49.
- [6] Seaman S. R. et White I. R. (2013), Review of inverse probability weighting for dealing with missing data. *Statistical methods in medical research*, 22(3):278-95.
- [7] Morita S., Kobayashi K., Eguchi K., Matsumoto T., Shibuya M., Yamaji Y., et Ohashi Y. (2005), Analysis of incomplete quality of life data in advanced stage cancer: a practical application of multiple imputation. *Quality of Life Research*, 14(6):1533-44.
- [8] Bonnetain F., Dahan L., Maillard E., Ychou M., Mitry E., Hammel P., Legoux J. L., Rougier P., Bedenne L. et Seitz J.F. (2010) Time until definitive quality of life score deterioration as a means of longitudinal analysis for treatment trials in patients with metastatic pancreatic adenocarcinoma, *European Journal of Cancer*, 46(15):2753-62.
- [9] Hamidou Z., Dabakuyo T. S., Mercier M., Fraisse J., Causeret S., Tixier H. Padeano M. M., Loustalot C., Cuisenier J., Sauzedde J. M., Smail M., Combiér J. P., Chevillote P., Rosburger C., Arveux P. et Bonnetain F. (2011), Time to deterioration in quality of life score as a modality of longitudinal analysis in patients with breast cancer. *The Oncologist*, 16(10):1458-68.
- [10] Aota A., Hamidou Z., Paget-Bailly S., Chibaudel B., Bascoul-Mollevi C., Auquier P., Westeel V., Fiteni F., Borg C., et Bonnetain F. (2015a), Time to Health-related Quality of Life score deterioration as a modality of longitudinal analysis for health-related quality of life studies in oncology: do we need RECIST for quality of life to achieve standardization? *Quality of Life Research*, 24(1):5-18.
- [11] Aota A., Mouillet G., Trouilloud I., Dupont-Gossart A. C., Artru P., Lecomte T., Zaanán A., Gauthier M., Fein F., Dubreuil O., Paget-Bailly S., Taieb J., et Bonnetain F. (2015b), Sequential FOLFIRI3 + Gemcitabine improves health-related quality of life deterioration-free survival of patients with metastatic pancreatic adenocarcinoma: a randomized phase II trial. *PLOS ONE* (in press).
- [12] Aaronson N. K., Ahmedzai S., Bergman B., Bullinger M., Cull A., Duez N. J., Filiberti A., Flechtner H., Fleishman S. B., de Haes J. C., Kaasa S., Klee M., Osoba D., Razavi D., Rofe P. B., Schraub S., Sneeuw K., Sullivan M., Takeda F. for the European Organization for Research and Treatment of Cancer Study Group on. (1993), The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology, *Journal of the National Cancer Institute*, 85(5):365-76.
- [13] Fayers P. M., Aaronson N. K., Bjordal K., Groenvold M. et Curran D. (2001), AobotEQoLG. B. EORTC QLQ-C30 Scoring Manual (3rd edition). Brussels: EORTC 2001 ed2001.
- [14] Cottone F., Efficace F., Apolone G. et Collins G. S. (2013), The added value of propensity score matching when using health-related quality of life reference data. *Stat Med*, 32(29):5119-32.
- [15] Hougaard P. (1995), Frailty models for survival data, *Lifetime Data Anal*, 1(3):255-73.
- [16] Azur M. J., Sturat E. A., Frangakis C. et Leaf P. J. (2011), Multiple imputation by chained equations: what is it and how does it work?, *Int J Methods Psychiatr Res*, 20(1): 40-9.
- [17] Aota A., Barbieri A., Savina M., Pam A., Gourgou-Bourgade S., Bonnetain F., et Bascoul-Mollevi C. (2014), Comparison of three longitudinal analysis models for the health-related quality of life in oncology: a simulation study. *Health and Quality of Life Outcomes*, 12(1):1326.
- [18] Robins J. (1992), Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. *Biometrika*, 79, 321-334.