

Thème: Statistique et médecine personnalisée

Titre: Validation d'un biomarqueur prédictif de la réponse à un traitement en présence de données censurées

Bassirou Mboup (ISUP Université Pierre et Marie Curie)

Paul Blanche (LMBA/UMR6205 Université Bretagne Sud Vannes)

Aurélien Latouche (Inserm U900, Institut Curie et Cnam)

8 juin 2017

Des nouvelles méthodes statistiques sont utilisées pour évaluer et valider un biomarqueur prédictif d'une réponse à un traitement. Janes et al. (2014) ont proposé une approche formelle et compréhensible pour une évaluation d'un biomarqueur. Ces méthodes permettent notamment d'aller au delà du simple test d'une interaction entre le traitement et le biomarqueur qui est couramment employé. En effet, il a été démontré que deux biomarqueurs ayant la même interaction avec le traitement peuvent avoir des capacités prédictives très différentes Janes et al. (2011). La présence d'une interaction entre le biomarqueur et le traitement est une condition nécessaire mais pas suffisante.

Soit D l'indicateur de la réponse au traitement, $Treat$ l'indicateur de traitement et Y un biomarqueur continu, F sa fonction de répartition. L'effet absolu du traitement conditionnellement au biomarqueur est défini comme $\Delta(Y) = P(D = 1|Treat = 0, Y) - P(D = 1|Treat = 1, Y)$ et on note F_Δ sa fonction de répartition. L'étape descriptive pour guider les décisions de traitement repose sur l'estimation de deux courbes prédictivités, qui représentent les risques conditionnelles, $P(D = 1|Treat = 0, Y)$ et $P(D = 1|Treat = 1, Y)$ en fonction de la proportion de la population ayant une valeur au marqueur inférieure à Y , i.e. $F(Y)$. À partir de ces deux courbes, on peut estimer l'effet absolu du traitement sachant la valeur du biomarqueur Y (Figure 1).

Nous nous proposons d'étendre ces méthodes à des données censurées à droite. Soient T le temps d'évènement, C le temps de censure, $\tilde{T} = \min(T, C)$ le temps de suivi observé, $\delta = \mathbb{1}\{T \leq C\}$ l'indicatrice de survenue de l'évènement et $D(t) = \mathbb{1}\{T \leq t\}$ l'indicatrice de l'évènement à l'horizon de prédiction t . On définit l'effet absolu du traitement au temps t conditionnellement au biomarqueur Y comme

$$\Delta(t, Y) = P(D(t) = 1|Treat = 0, Y) - P(D(t) = 1|Treat = 1, Y).$$

Pour estimer $\Delta(t, Y)$, nous avons utilisé le modèle de régression logistique dépendant du temps suivant :

$$\text{logit}P\{T \leq t|Treat, Y\} = \beta_0(t) + \beta_1(t) \times Treat + \beta_2(t) \times Y + \beta_3(t) \times Treat \times Y.$$

L'estimation du modèle repose sur la résolution d'équations du score associées à la méthode du maximum de vraisemblance pour données non censurées, auxquelles des pondérations par l'inverse de la probabilité de censure sont ajoutées, pour tenir compte des données censurées

Oncotype-Dx est un test génomique fondé sur 21 gènes et permet de classer les patientes selon leur risque de rechute.

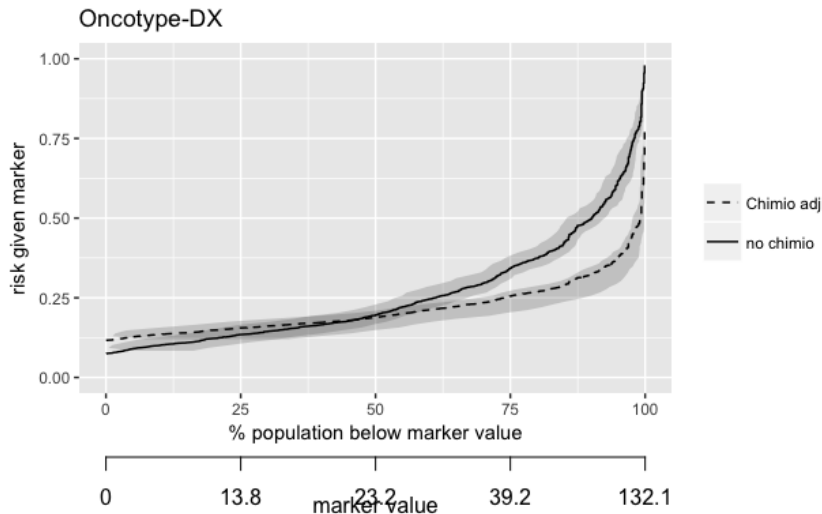


FIGURE 1 – Dans cet exemple, les données ont été simulées pour ressembler à celles d’un essai randomisé et contrôlé du cancer du sein, où le marqueur Y ressemble à Oncotype-DX.

(Zheng et al., 2006). Nous avons évalué la performance de nos méthodes par une étude de simulation et illustré sa pertinence avec des données réelles de cancérologie.

La méthode permet d’obtenir le seuil d’un biomarqueur au delà duquel nous recommandons le traitement et ainsi de déterminer un sous groupe de patients qui va bénéficier du traitement.

Ce travail est soutenu par le GDR statistique et Santé qui finance un stage de fin d’étude, partagé entre l’unité INSERM U900 de Paris (Saint Cloud) et l’Université de Vannes.

RÉFÉRENCES

- Janes, H., Brown, M. D., Huang, Y., and Pepe, M. S. (2014). An approach to evaluating and comparing biomarkers for patient treatment selection. *Int J Biostat*, 10(1) :99–121.
- Janes, H., Pepe, M. S., Bossuyt, P. M., and Barlow, W. E. (2011). Measuring the performance of markers for guiding treatment decisions. *Ann. Intern. Med.*, 154(4) :253–259.
- Zheng, Y., Cai, T., and Feng, Z. (2006). Application of the time-dependent ROC curves for prognostic accuracy with multiple biomarkers. *Biometrics*, 62(1) :279–287.