

Prise en main et application du test de Buyse dans une étude en vie réelle un tutoriel

INTRODUCTION

La comparaison de traitements se fait souvent sur plusieurs critères de jugement. En règle générale, ces critères sont évalués indépendamment, avec un critère principal et des critères secondaires. Le test de Buyse¹, est une généralisation des comparaisons non-paramétriques des paires permettant de comparer deux groupes de traitement sur plusieurs critères de jugement de nature différentes ou de manière hiérarchisée.

OBJECTIFS

Décrire l'utilisation de la méthode et la prise en main du package avec une application sur des données anonymisées issues des bases du Système National des Données de Santé (SNDS).

Ce travail est une présentation pédagogique simplifiée de la méthode de Buyse qui s'appuie sur des données en vie réelle d'une étude comparant la survenue d'événements indésirables.

MÉTHODES

La méthode des paires généralisées permet d'étudier, entre autres, le temps jusqu'à la survenue des événements indésirables en présence de données censurées à droite².

Au sein d'une cohorte rétrospective, des patients initiant un traitement A ont été appariés sur un score de propension à des patients initiant un traitement B de référence avec un ratio 1:3.

Les comparaisons des patients ont été réalisées en appliquant le test de Buyse pour comparer plusieurs critères de jugements comprenant les temps jusqu'à la survenue de cancers, d'infections et d'événements cardiovasculaires.

Rappel des règles de classement des paires en fonction du type de variable

Nous noterons x_i^0 la valeur du temps jusqu'à événement du i -ème patient issu du groupe A, et y_j^0 la valeur du temps jusqu'à événement du j -ème patient issu du groupe B. Les événements ne sont pas observés pour tous les individus du fait des censures à droite (censure = 1 dans le **Tableau 1**). Les temps jusqu'à observation sont notés $x_i = \min(x_i^0, u_i)$ et $y_j = \min(y_j^0, v_j)$ où u_i et v_j sont les temps jusqu'à censure des patients i et j respectivement.

On définit les indicatrices d'événement comme suit :

$$y_i = \begin{cases} 1: x_i = x_i^0 \\ 0: x_i < x_i^0 \end{cases} \text{ et } \varepsilon_j = \begin{cases} 1: y_j = y_j^0 \\ 0: y_j < y_j^0 \end{cases}$$

Soit τ une valeur seuil que doit dépasser la différence entre les valeurs de X et Y pour être considérée comme significative (**voir Tableau 1**).

Scoring des paires et évaluation de la propension au succès

Un score est défini pour le l -ième critère de jugement priorisé et pour chaque paire de patients incluant un patient i issu du groupe A et un patient j issu du groupe B tel que :

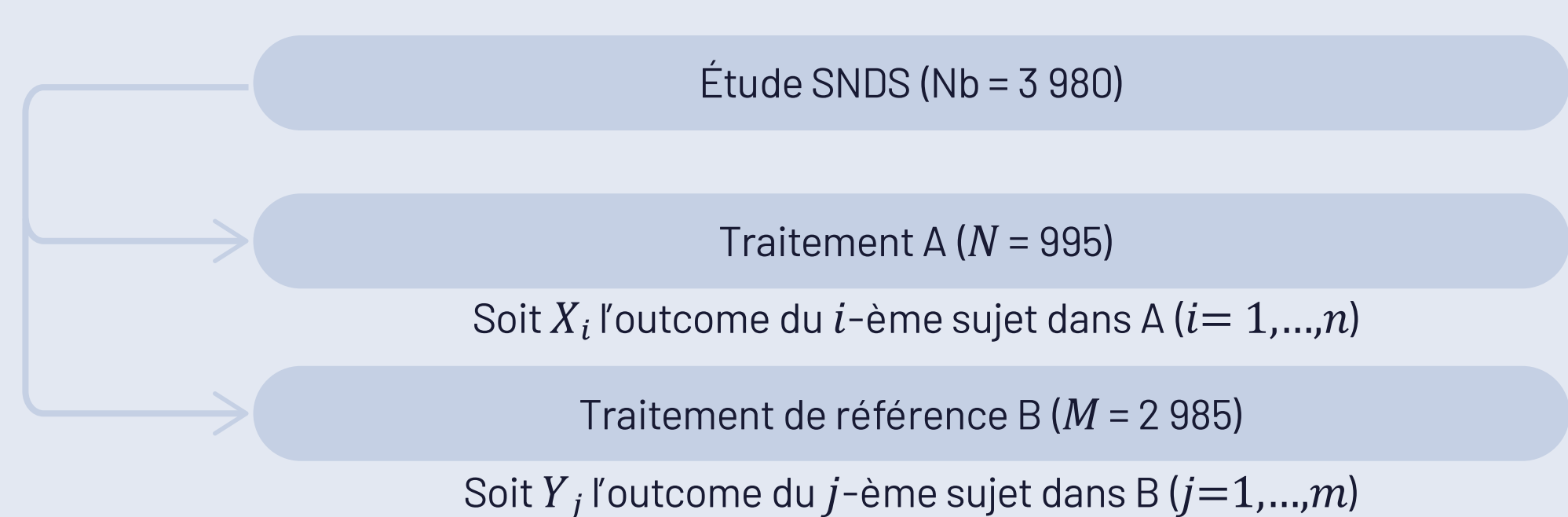
$$p_{ij}(l) = \begin{cases} 1: \text{si la paire est favorable} \\ -1: \text{si la paire est défavorable} \\ 0: \text{si la paire est neutre,} \\ \text{non-informative ou déjà classée} \end{cases}$$

La propension au succès $\delta(l)$ représente la différence entre le nombre de paires classées favorables au traitement et le nombre de paires classées défavorables au traitement divisée par le nombre total de paires et est calculée comme suit :

$$\delta(l) = \frac{\sum_{i=1}^n \sum_{j=1}^m p_{ij}(l)}{n \cdot m}$$

À noter qu'en présence de stratification, le calcul de est réalisé au sein de chacune de K strates.

La 'propension cumulée au succès', $\Delta(l)$ pour le critère de jugement de la dernière priorité est appelé et est notée Δ et se calcule comme suit : $\Delta(l) = \sum_{\lambda=1}^l \delta(\lambda)$.



CONCLUSION

Ce tutoriel fournit un exemple d'application de la méthode par paires généralisées avec différents calculs du bénéfice net du traitement. Cette méthode permet d'apporter une réponse globale à la question de l'efficacité ou de la sûreté des traitements en examinant cette différence entre deux traitements en combinant l'information sur plusieurs critères avec le bénéfice net. Elle peut remplacer avantageusement l'étude des critères composites car elle permet d'avoir l'impact de chaque outcome séparément tout en ayant un critère global d'évaluation.

Références

- Buyse M. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. Stat Med. 2010 Dec 30;29(30):3245-57. doi: 10.1002/sim.3923. PMID: 21170918.
- Péron J, Buyse M, Ozenne B, Roche L, Roy P. An extension of generalized pairwise comparisons for prioritized outcomes in the presence of censoring. Stat Methods Med Res. 2018 Apr;27(4):1230-1239. doi: 10.1177/0962280216658320. Epub 2016 Aug 2. PMID: 27487842.

Remerciements

Les auteurs tiennent à remercier les équipes de la Direction de la stratégie, des études et des statistiques, du Département Accès, Traitement et Analyse de la Donnée et des Demandes Externes de la Caisse nationale de l'assurance maladie (CNAM) pour l'extraction des données.

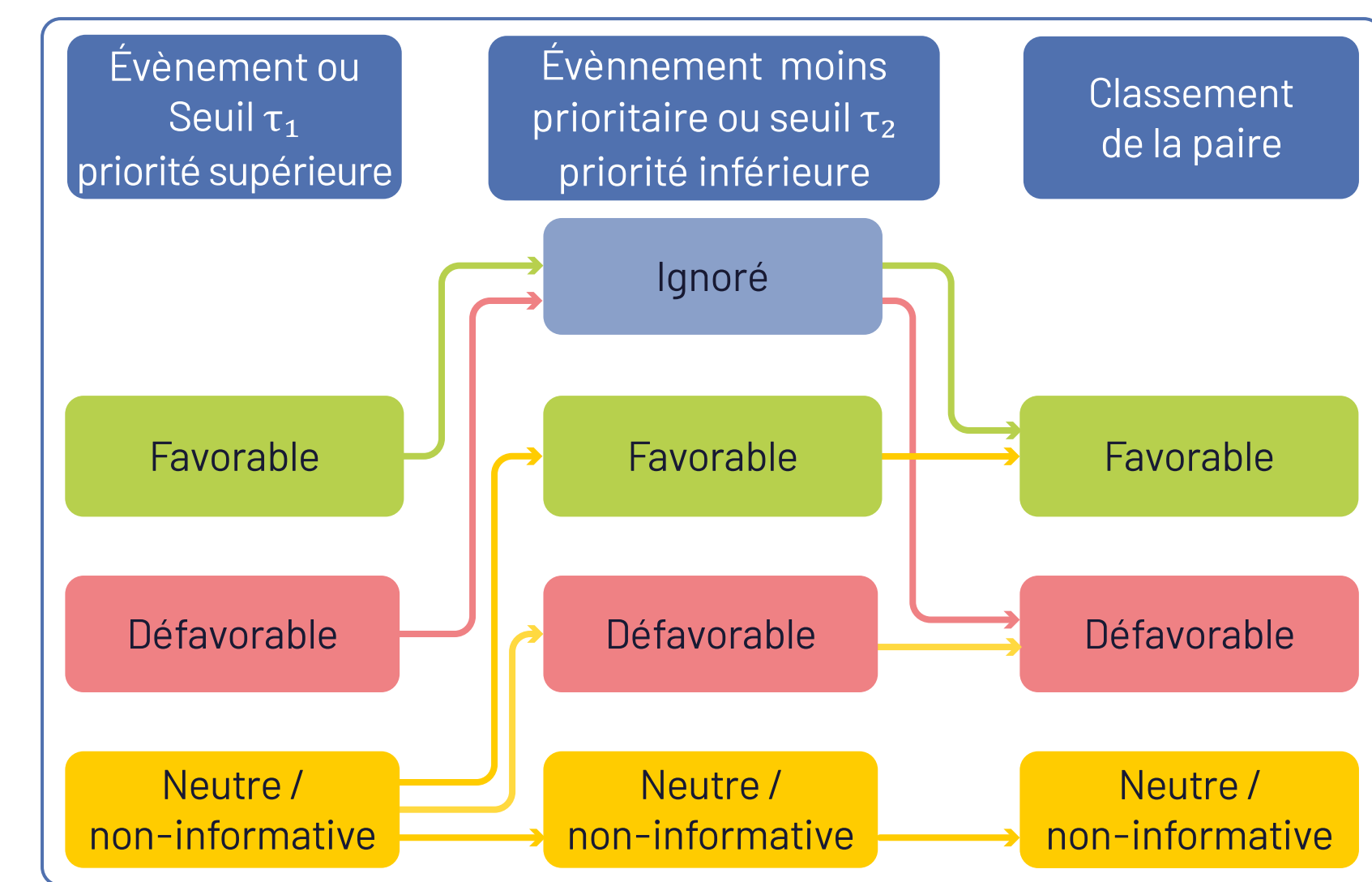
RÉSULTATS

Un total de 3 980 patients sont inclus dans cette étude. Les comparaisons ont été réalisées sur plusieurs critères de jugement, en les hiérarchisant et/ou en adoptant une pondération subjective en fonction de la criticité clinique. Les résultats seront exprimés en termes de bénéfice net, soit la différence entre le nombre de paires en faveur du traitement d'intérêt et le nombre de paires en défaveur du traitement d'intérêt, en prenant en compte les paires neutres ou en les traitant comme non-informatives.

Tableau 1 : Classement des paires en fonction de la censure (Y_i, ε_j) et du seuil tau (τ)

(Y_i, ε_j)	$x_i - y_j > \tau$	$x_i - y_j < -\tau$	$ x_i - y_j \leq \tau$
(1,1)	Favorable	Défavorable	Neutre
(0,1)	Favorable	Non-informative	Non-informative
(1,0)	Non-informative	Défavorable	Non-informative
(0,0)	Non-informative	Non-informative	Non-informative

Tableau 2 : Classement des paires pour des critères de jugement priorisés ou pour les seuil de priorité (exemple pour deux événements)



Test sur un outcome – survenue de cancer

```
BT=BuyseTest(data=data, endpoint = "delai_cancer",
type="timeToEvent", treatment = "Traitement",
status = "top_cancer", threshold = 90 )
Or
BT.f =BuyseTest(Categorie = tte(delai_cancer,
threshold =90, status="top_cancer"), data=data)
```

Critère de jugement	seuil	total	favorable	défavorable	neutre	non-informative	Delta (Δ)	IC 95%	P-valeur
Délais jusqu'à survenue du cancer	90	100 %	1,14 %	3,5 %	0,01 %	95,35 %	0,0237	[-0,0534; 0,0061]	0,120
Délais jusqu'à survenue du cancer	90	2 970 075 paires	33 769 paires	104 043 paires	181 paires	2 832 082 paires	0,0237	[-0,0534; 0,0061]	0,120

Hiérarchisation des paires

```
#f2 = Traitement ~ tte(delai_cancer, threshold = 90, status = "top_cancer") + tte(delai_cardio, threshold = 60, status = "top_cardio") + tte(delai_infections, threshold = 30, status="top_infections") + tte(delai_thrombose, threshold = 60, status="top_thrombose")
BT.H = BuyseTest(#f2, data = data, trace = 0)
summary(BT.H)
BT.nH = BuyseTest(#f2, hierarchical = FALSE, data = data, trace = 0)
summary(BT.nH)
```

delta (δ) est la propension au succès qui moyenne les paires favorables et celles défavorables rapportée au nombre de paires totales. Une valeur positive signifie que l'issue est en faveur du traitement A pour ce critère de jugement. Une valeur négative signifie que l'issue est en faveur du traitement B – contrôle.

Delta (Δ) est l'indicateur cumulatif qui permet de conclure sur l'ensemble des critères de jugement évalués.

Ici le **Delta** global est négatif, cela signifie que sur l'ensemble des critères de jugement examinés, en moyenne le traitement témoin B est meilleur.

Hierarchical = True

Critère de jugement	seuil	total	favorable	défavorable	neutre	non-informatif	delta (δ)	Delta (Δ)	IC 95%	P-valeur
Cancer	90	100 %	1,14 %	3,50 %	0,01 %	95,35 %	-0,0237	-0,0237	[-0,0534; 0,0061]	0,120
Cardio	60	95,36 %	0,74 %	1,35 %	0,00 %	93,26 %	-0,0061	-0,0298	[-0,0634; 0,0039]	0,083
Infections	30	93,26 %	4,12 %	2,13 %	0,01 %	87,01 %	-0,0198	-0,0100	[-0,0487; 0,0289]	0,615
Thrombose	60	87,01 %	0,25 %	0,12 %	0,00 %	86,65 %	-0,0013	-0,0087	[-0,0476; 0,03031]	0,663

Hierarchical = False

Critère de jugement	seuil	total	favorable	défavorable	neutre	non-informatif	delta (δ)	Delta (Δ)	IC 95%	P-valeur
Cancer	90	100 %	1,14 %	3,50 %	0,01 %	95,35 %	-0,0237	-0,0059	[-0,0134; 0,0015]	0,120
Cardio	60	100 %	0,81 %	1,42 %	0,00 %	97,77 %	-0,0061	-0,0074	[-0,0160; 0,0011]	0,088
Infections	30	100 %	4,48 %	2,26 %	0,01 %	93,24 %	-0,0222	-0,0019	[-0,0117; 0,0011]	0,707
Thrombose	60	100 %	0,31 %	0,13 %	0,00 %	99,56 %	-0,0017	-0,0015	[-0,0113; 0,0084]	0,773

Exemple avec pondération

On peut attribuer un poids à chaque critère de jugement afin de pondérer son impact en fonction de son importance (e.g., pour prendre en compte la gravité). Ici dans le cadre de l'analyse du temps jusqu'à la survenue du critère de jugement, la pondération intervient sur le temps moyen.

```
#f2w = Traitement ~ tte(delai_cancer, threshold = 90, status = "top_cancer", weight = 0.9)
#f2w = update.formula(#f2w, . . . + tte(delai_cardio, threshold = 60, status = "top_cardio", weight = 0.7))
#f2w = update.formula(#f2w, . . . + tte(delai_infections, threshold = 30, status = "top_infections", weight = 0.5))
#f2w = update.formula(#f2w, . . . + tte(delai_thrombose, threshold = 60, status = "top_thrombose", weight = 0.7))
BT.nHw = BuyseTest(#f2w, hierarchical = FALSE, data = data, trace = 0)
summary(BT.nHw, print = FALSE)$table.print[, -13]
```

Critère de jugement	poids	seuil	total	favorable	défavorable (%)	neutre	non-informative	delta (δ)	Delta (Δ)	IC 95%	P-valeur
Cancer	0.9	90	100 %	1,14 %	3,50 %	0,01 %	95,35 %	-0,0237	-0,0213	[-0,0481; 0,0055]	0,120
Cardio	0.7	60	100 %	0,81 %	1,42 %	0,00 %	97,77 %	-0,0061	-0,0256	[-0,0548; 0,0037]	0,087
Infections	0.5	30	100 %	4,48 %	2,26 %	0,01 %	93,24 %	-0,0222	-0,0145	[-0,0452; 0,0163]	0,358
Thrombose	0.7	60	100 %	0,31 %	0,13 %	0,00 %	99,56 %	-0,0017	-0,0132	[-0,0442; 0,0177]	0,402

Changements

Évaluation simultanée de plusieurs critères de jugement

Prise en compte globale des critères de jugement, ici par type, qui sont hiérarchisés par priorités successives (ordre d'inclusion dans la commande R) (**voir Tableau 2**).

Le seuil de significativité clinique pour le temps jusqu'à survenue du critère de jugement est fixé à 30 jours pour les pathologies infectieuses dans cet exemple.

```
#mf= Traitement ~ tte(delai_infection1, threshold = 90, status = "top_infection1")+
tte(delai_infection2, threshold = 90, status = "top_infection2")+
tte(delai_infection3, threshold = 90, status = "top_infection3")+
tte(delai_infection4, threshold = 90, status = "top_infection4")+ ...
BT.inf = BuyseTest(#mf, data = data, trace = 0)
summary(BT.inf)
```

Critère de jugement	seuil	total	favorable	défavorable	neutre	non-informative	delta (δ)	Delta (Δ)	IC 95%	P-valeur
Pathologie infectieuse 1	30	100,00 %	0,97 %	0,00 %	0 %	99,03 %	0,0097	0,0097	[0,005; 0,0143]	4,3571e-05
Pathologie infectieuse 2	30	99,03 %	1,45 %	0,88 %	0 %	96,70 %	0,0056	0,0153	[0,002; 0,0286]	0,024
Pathologie infectieuse 3	30	96,70 %	0,90 %	0,12 %	0 %	95,68 %	0,0078	0,0231	[0,0089; 0,0374]	0,001
Pathologie infectieuse 4	30	95,68 %	0,44 %	0,41 %	0 %	94,83 %	0,0003	0,0234	[0,0081; 0,0387]	0,003
Pathologie infectieuse 5	30	94,83 %	0,21 %	0,00 %	0 %	94,62 %	0,0021	0,0255	[0,01; 0,041]	0,001
Pathologie infectieuse 6	30	94,62 %	0,12 %	0,00 %	0 %	94,49 %	0,0012	0,0267	[0,011; 0,0424]	< 0,001
Pathologie infectieuse 7	30	94,49 %	0,49 %	0,82 %	0 %	93,19 %	0,0033	0,0235	[0,004; 0,043]	0,018
Pathologie infectieuse 8	30	93,19 %	0,08 %	0,43 %	0 %	92,68 %	0,0035	0,0200	[-0,005; 0,045]	0,117

Pour les pathologies infectieuses 1 à 7 la propension cumulée au succès est en faveur du traitement A et est statistiquement significative, cependant dès lors qu'on ajoute la pathologie infectieuse 8 la propension au succès globale diminue et devient en faveur du traitement B pour cette pathologie.