



STATISTIQUE
SCIENCE DES DONNÉES BIOSTATS
UNIVERSITÉ DE MONTPELLIER

MASTER 2 STATISTIQUES ET SCIENCES DES DONNÉES - BIOSTATS

Comparaison par paire généralisée pour plusieurs
critères de jugement classés par ordre de priorité :
win ratio vs. net benefit of treatment

TENA Arthur



Année 2024 - 2025

Abréviations

GPC	Generalized pairwise comparisons
tte	time-to-event
GGW	Gehan's Generalized Wilcoxon
WR, R_w	Win ratio
WO	Win Odds

TABLE 1 – Table des abréviations

Table des matières

1	Introduction	3
2	Articles	4
2.1	Critère de jugement	4
2.2	Comparaison par paire	4
2.2.1	Comparaison par paire généralisée	6
2.2.2	Win Ratio	8
2.2.3	Win odds	9
2.3	Conclusion	9
3	Simulations de données cliniques	10
A	Annexe	11
A.1	Article	11
A.2	Simulations de données cliniques	14

1 Introduction

2 Articles

Lors d'un essai clinique, plusieurs phases se succèdent, chacune ayant un objectif spécifique visant à évaluer la sécurité, l'efficacité et les effets secondaires d'un traitement ou d'une intervention.

- **Phase 0** : Étude préclinique exploratoire, visant à observer le comportement du médicament à très faible dose chez les animaux ou les cellules.
- **Phase I** : Évaluation de la sécurité et du dosage. Cette phase vise principalement à identifier les effets secondaires, la tolérance et la dose maximale tolérée.
- **Phase II** : Étude de l'efficacité et de la posologie. Elle permet d'affiner la dose optimale et d'évaluer les premiers signaux d'efficacité sur une période de quelques mois à un an.
- **Phase III** : Confirmation de l'efficacité et surveillance des effets secondaires sur une population plus large. Cette phase est souvent déterminante pour l'autorisation de mise sur le marché et peut durer plusieurs années.
- **Phase IV** : Surveillance post-commercialisation, destinée à suivre les effets à long terme, détecter d'éventuels effets indésirables rares et évaluer l'utilisation du traitement en conditions réelles.

Au cours de cette étude, plusieurs notions ont nécessité un approfondissement. Tout d'abord, il était important de comprendre la définition d'un **critère de jugement** que nous verrons dans la partie 2.1. Ensuite, dans la partie 2.2, nous allons étudier la **comparaison par paire**, ce que c'est et quelles sont ses utilités. Et nous regarderons trois méthodes de calculs, la **comparaison par paire généralisée** (GPC) dans la partie 2.2.1, les **win ratio** dans la partie 2.2.2 et enfin une approche un peu différente des win ratio, les **win odds** dans la partie 2.2.3.

Nous suivrons notre étude à l'aide d'un exemple d'accident cardiovasculaire.

2.1 Critère de jugement

Le critère de jugement (en anglais *outcomes* ou *endpoint*) est expliqué par McCoy (2018), lors de l'évaluation des résultats d'une étude clinique.

Lors d'un essai clinique, les **critères de jugement** sont des indicateurs permettant d'évaluer l'efficacité et la sécurité d'un traitement. On distingue généralement un **critère de jugement principal** (*primary endpoint*), qui représente l'événement clinique le plus important pour l'étude, et des **critères de jugement secondaire** (*secondary endpoint*), qui sont des événements pertinents permettant d'augmenter la justesse statistique et de mieux refléter l'impact global du traitement. Il est tout de même possible d'avoir plusieurs critères de jugement principaux ou secondaires, on parlera alors de **critères de jugement composé** (*endpoint composite*).

Par exemple, dans une étude évaluant l'efficacité d'un traitement sur la prévention des événements cardiovasculaires majeurs, le critère de jugement principal pourrait être la mortalité cardiovasculaire, c'est-à-dire le décès dû à une cause d'origine cardiaque. Cependant, comme la mortalité cardiovasculaire seule peut être un événement relativement rare sur la durée de l'étude, on peut également définir un critère de jugement principal composé, pouvant inclure plusieurs événements cliniquement significatifs, tels que l'hospitalisation pour infarctus du myocarde (IDM) ou pour un accident vasculaire cérébral (AVC).

Ainsi, le choix entre un critère de jugement principal unique et un critère composé dépend des objectifs de l'étude, de la fréquence des événements et de leur importance clinique. Une attention particulière doit être portée à la prise en compte des événements récurrents, afin d'assurer une interprétation fiable et robuste des résultats.

2.2 Comparaison par paire

La comparaison par paire, comme l'explique Marc Buyse (2010) au travers de deux exemples est une méthode de comparaison de deux échantillons, un échantillon T de n patients essayant un nouveau traitement et un échantillon de contrôle C de m patients prenant un traitement de contrôle (ancien traitement ou placebo).

Cette comparaison requiert alors de prendre un individu de chaque paire (T, C) , le résultat de ces deux individus est alors comparé et est dit "**favorable**" si le résultat de l'individu du groupe T est meilleur¹ que celui du groupe C . Il est dit "**défavorable**" si le résultat de T est moins bon que celui de C . "**Neutre**" si le résultat de T ne change pas de celui de C , et enfin "**non-informatif**" si on ne peut pas déterminer quel groupe a le meilleur résultat.

Explication dans le cas avec un seul outcome

Plusieurs types de variables peuvent être étudiés, des variables binaires, des variables continues ou des variable "tte" (time-to-event) et, suivant le type de variable. On peut alors voir que pour les variables binaire les quatre cas de figure et le résultat de la paire associé. Pour les variables continue et tte, un seuil d'acceptation τ est défini, pour les variables continues trois cas de figure sont étudiés, si la différence de la paire est supérieur à τ alors le résultat est favorable, si la différence est inférieur à $-\tau$ alors le résultat est défavorable et la paire sera désignée neutre si la valeur absolue de la différence est inférieure ou égal à τ . Enfin pour les variables tte, plus de cas de figure sont étudiées car ces variables combinent les variables binaires (valant 1 si l'observation est complète, donc non censurée et 0 si l'observation est censurée) et les variables continues (temps observée). Ainsi douze cas de figure sont à prendre en compte suivant s'il y a observation complète ou non et comment se comporte la différence de la paire par rapport à τ . Tout ceci est mieux détaillé en annexe dans la section A.1 dans les tableaux 5, 6 et 7.

Dans le cas particulier où $\tau = 0$, la comparaison par paire sera équivalente à un test de Wilcoxon pour les variables continues et à un test de Gehan (GGW, Gehan (1965)) qui est un test de Wilcoxon généralisé aux variables tte.

Remarque : Les variables peuvent également nécessiter une stratification, par exemple en fonction de l'âge, du sexe, du stade de la maladie ou d'un risque que le patient encoure. Cette stratification sera indiquée par l'indice k ($k \in \{1, \dots, K\}$), définissant ainsi des sous-groupes de taille n_k et m_k pour les groupes T et C , respectivement. Elle sera utile pour comparer des individus semblable selon ses caractéristiques.

L'idée globale est donc de comparer chaque paire d'individu afin de donner un score qui sera utilisé pour déterminer l'utilité du nouveau traitement.

Nous allons alors créer un score $p_{ijk} \in \{-1, 0, 1\}$ qui nous servira par la suite où,

$$p_{ijk} = \begin{cases} 1 & \text{si la paire est favorable} \\ -1 & \text{si la paire est défavorable} \\ 0 & \text{si la paire est neutre} \end{cases}$$

Dans l'exemple d'accident cardiovasculaire, la comparaison par paire va comparer 2 à 2 chaque patient de chaque groupe en attribuant ce score suivant si le critère de jugement principal est arrivé. Chaque patient étant comparé avec des patients similaire afin de ne pas fausser l'étude.

Extension dans le cas de plusieurs critères

La comparaison par paire peut être généralisée de plusieurs façons : en intégrant plusieurs seuils, en tenant compte d'observations répétées dans le temps pour un même critère de jugement, ou encore en considérant des critères de jugement composés.

Chaque extension présente un intérêt particulier. L'utilisation de seuils successifs permet, par exemple, de définir différents stades d'une maladie. L'intégration d'observations répétées offre un suivi des patients et facilite la prédiction des critères de jugement. Enfin, prendre en compte plusieurs critères de jugement est pertinent lorsque l'analyse ne peut se limiter à un seul indicateur. C'est cette dernière approche que nous explorerons dans la suite.

1. un meilleur résultat dépend de ce que l'on cherche. "Meilleur" peut être plus grand ou plus petit suivant l'étude.

à voir
comment
définir
mieux
les va-
riables
tte

On notera alors l ($l \in \{1, \dots, L\}$) le l -ième critère de jugement classé par priorité où 1 est la plus grosse priorité et L la plus basse. On aura alors une mesure $u_{ijk}(l)$ comme suit

$$u_{ijk}(l) = \begin{cases} 1 & \text{si le } l\text{-ième critère est non-informatif} \\ 0 & \text{sinon} \end{cases}$$

et un score associé :

$$p_{ijk}(l) = \begin{cases} 1 & \text{si le } l\text{-ième critère est favorable} \\ -1 & \text{si le } l\text{-ième critère est défavorable} \\ 0 & \text{si le } l\text{-ième critère est neutre} \end{cases} ; \quad \text{et} \quad u_{ijk}(h) = 1 \quad \forall h < l \quad (1)$$

Cette deuxième partie $u_{ijk}(h) = 1 \quad \forall h < l$ joue un rôle clé dans la priorisation des critères de jugement. En d'autres termes, cela signifie que si tous les critères plus prioritaires que l (celui utilisé pour calculer $p_{ijk}(l)$) sont considérés comme non-informatifs, alors on passe au critère suivant. Les conditions précises définissant un critère comme non-informatif sont détaillées en annexe dans les tableaux 5, 6 et 7.

Ce score sera utile pour calculer une proportion en faveur du traitement évoqué dans la partie suivante sur l'estimation et les tests.

Dans notre exemple d'accident cardiovasculaire avec comme différents critères la mort, l'hospitalisation pour infarctus du myocarde et l'AVC. Nous allons devoir comparer chaque patient étant dans les strates identiques 2 à 2 afin de voir si les 3 critères, par ordre de priorité, sont en faveur du nouveau traitement ou non. Pour ce faire, nous regarderons d'abord si le critère prioritaire (la mort) est non-informatif ou non, s'il ne l'est pas alors on calculera le score p de ce critère, alors que s'il est non-informatif on passera au critère suivant (infarctus du myocarde par exemple) et on calculera le score p correspondant etc. Des observations répétées seront effectuées au long de l'étude où les différents critères seront observés.

2.2.1 Comparaison par paire généralisée

La comparaison par paire généralisée (GPC) s'appuie sur 2 notions, la *proportion en faveur du traitement* (noté Δ) et la *proportion en faveur du traitement cumulé* (noté δ).

Dans la GPC, nous comparons toutes les paires d'individus, nous avons donc $n.m$ paires qui peuvent être rangées par strate, toujours dans le but de comparer des individus identiques.

La **proportion en faveur du traitement** Δ est la différence entre le nombre de paire favorable et le nombre de paire défavorable divisé par le nombre de paire totale sans prendre en compte les différents critères de jugement.

$$\Delta = \frac{\sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{j=1}^{m_k} p_{ijk}}{\sum_{k=1}^K n_k \cdot m_k} \quad (2)$$

Avec $n = \sum_{k=1}^K n_k$, $m = \sum_{k=1}^K m_k$ et $\sum_{k=1}^K n_k \cdot m_k \leq n \cdot m$.

Δ varie entre -1 et 1, où une valeur de -1 signifie que l'ancien traitement est uniformément meilleur que le nouveau, et une valeur de 1 signifie que le nouveau traitement est uniformément meilleur que l'ancien.

Ce qui est différent de la **proportion en faveur du traitement cumulé** dépendant elle du critère de jugement l ,

$$\delta(l) = \frac{\sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{j=1}^{m_k} p_{ijk}(l)}{\sum_{k=1}^K n_k \cdot m_k} \quad (3)$$

On a alors que $\Delta(l) = \sum_{i \leq l} \delta(i)$ et $\Delta = \Delta(L)$

Les tests seront réalisés sur la valeur de Δ sous l'hypothèse $H_0 : \Delta = 0$, correspondant à l'absence de différence entre les deux traitements et l'hypothèse alternative $H_1 : \Delta \neq 0$. Rejeter l'hypothèse

nulle signifie qu'il existe une différence significative entre les deux traitements qu'elle soit en faveur du traitement T ou en celui de C.

Nous pouvons alors créer, à l'aide d'un nombre de test suffisant (S), des intervalles de confiance à $(1 - \alpha)\%$ basé sur Δ_{obs} , la valeur de Δ observée de Δ sous l'hypothèse H_0 . Cet intervalle de confiance est calculé comme suit :

$$[\Delta_{obs} + \Delta_{\alpha/2}, \Delta_{obs} + \Delta_{1-\alpha/2}]$$

où $\Delta_{\alpha/2}$ est la valeur de Δ_i ($i \in \{1, \dots, S\}$) qui laisse au plus $\alpha/2$ des valeurs de Δ_i à sa gauche et $\Delta_{1-\alpha/2}$ la valeur qui laisse au plus $\alpha/2$ des valeurs de Δ_i à sa droite.

Les tests sur la proportion en faveur du traitement peuvent être effectués de la même façon pour la proportion en faveur du traitement cumulé dans le cas où nous avons plusieurs critères de jugements.

Exemple

Dans cette partie, nous allons voir un exemple de calcul de proportion en faveur du traitement sur 6 individus en prenant un critère de jugement principal, la mort. Et un critère de jugement secondaire, l'hospitalisation.

revoir
l'exemple

Traitement	T		C	
	Mort	Hospitalisation	Mort	Hospitalisation
individus	1	0	0	0
	0	0	0	1
	0	0	0	1
	0	1	1	0
	1	1	1	0
	0	1	1	0

TABLE 2 – Exemple utilisé

Dans cet exemple, nous avons $L=2$ et $u_{ij}(l) = 0 \forall i, j, l$ par soucis de simplicité. On aura alors $l = 1$ correspondant à la mort et $l = 2$ à l'hospitalisation.

$$\delta(2) = \frac{2 + 2 + 2 - 4 - 4 - 4}{36} = \frac{-6}{36} \quad \delta(1) = \frac{0 + 3 + 3 + 3 + 0 + 3}{36} = \frac{12}{36}$$

Et,

$$\Delta = \frac{12 - 6}{36} = \frac{1}{6}$$

Ici, nous avons donc que la proportion en faveur du traitement est positif ce qui nous amène à penser que le traitement est bénéfique. En revanche, lorsque l'on regarde les proportion en faveur du traitement cumulé $\delta(l)$, on se rend compte que le nouveau traitement diminue le nombre de mort mais augmente le nombre d'hospitalisation.

Relation entre les tests de Wilcoxon, de Gehan et de Fisher avec la GPC

Des rappels sur les tests utilisés se situent en annexe, section A.1.

Dans le cas de variables binaire, des tests de Fisher sont effectués.

On peut montrer que Δ est en fait la U-statistique pour le test de Wilcoxon,

$$U = \Delta = \frac{1}{n.m} \sum_{i=1}^n \sum_{j=1}^m U_{ij}$$

où

Pour des variables continues :

Pour des variables tte :

$$U_{ij} = \begin{cases} 1 & \text{si } X_i > Y_j \\ -1 & \text{si } X_i < Y_j \\ 0 & \text{sinon} \end{cases} \quad U_{ij} = \begin{cases} 1 & \text{si } x_i > y_j \text{ ou } x'_i \geq y_j \\ -1 & \text{si } x_i < y_j \text{ ou } x_i \leq y'_j \\ 0 & \text{si } x_i = y_j \text{ ou } (x'_i, y'_j) \text{ ou } x'_i < y_j \text{ ou } y'_j < x_i \end{cases}$$

Où les valeurs de x (resp. y) sont les temps jusqu'à l'échec des personnes ayant pris le traitement T (resp. C), x' (resp. y') est le temps jusqu'à la censure des personnes ayant pris le traitement T (resp. C).

Lors d'une GPC, on peut donc étendre les résultats en considérant des seuils successifs pouvant correspondre à différents stade d'une maladie par exemple, mais on peut aussi considérer des mesures observées répétées dans le temps, c'est-à-dire de prendre des mesures régulière (ou pas) dans le temps pour un meilleur suivi du traitement. Finkelstein & Schoenfeld (1999) proposent un test non-paramétrique combinant les variables tte et les variables longitudinale² donnant des scores aux patients survivant le plus longtemps.

2.2.2 Win Ratio

Explication

Pocock (2012) propose une approche de l'analyse des critères de jugement basés sur leurs priorité en définissant le **Win Ratio**.

Lors d'une approche Win Ratio, nous allons utiliser une méthode comparable à la GPC, il faut donc prendre en compte que deux patients n'auront pas les mêmes risques suivant leur profil (âge, sexe, diabète etc) afin de s'assurer que la différence soit due aux traitements et non pas aux profils.

Le *Win Ratio* propose deux approches : une approche **appariée** et une approche **non-appariée**. L'approche appariée compare un nombre fixe n de paires d'individus, chaque paire étant formée d'un individu du premier groupe et d'un individu du second groupe selon leur risque. L'approche non-appariée compare toutes les combinaisons possibles de paires entre les deux groupes, c'est-à-dire que chaque individu du premier groupe est comparé à tous les individus du second groupe.

Le Win Ratio sera aussi calculé à l'aide des $p_{ijk}(l)$ mais d'une façon différente de la GPC, cette fois-ci, trois quantités sont calculées, le nombre de win N_w le nombre de lose N_l et le nombre d'égalité N_t .

$$\begin{cases} N_w = \text{Card}(\{p_{ijk}(l) = 1\}) \\ N_l = \text{Card}(\{p_{ijk}(l) = -1\}) \\ N_t = \text{Card}(\{p_{ijk}(l) = 0\}) \end{cases}$$

Où $p_{ijk}(l)$ est défini à l'équation (1)

Le Win Ratio R_w sera alors $R_w = N_w/N_l$ et on pourra aussi calculer la proportion de win p_w , valant $p_w = \frac{N_w}{N_w + N_l}$ on pourra alors établir des intervalles de confiance à 95% :

$$\left[p_w \pm 1.96 \left(\frac{p_w(1-p_w)}{N_w + N_l} \right)^{1/2} \right] = [p_L, p_U]$$

où p_L est la borne inférieure de l'intervalle de p_w et p_U est la borne supérieure.

Comme le win ratio vaut $\frac{p_w}{1-p_w}$, on établit l'intervalle à 95% du win ratio comme étant :

$$\left[\frac{p_L}{(1-p_L)}, \frac{p_U}{(1-p_U)} \right]$$

De même, pour la significativité du test, on peut calculer le z-score, $z = (p_w - 0.5)/[p_w(1-p_w)/(N_w + N_l)]^{1/2}$ suivant une distribution normale standardisée. Nous pouvons alors trouver la p-valeur dans la table de la loi normale-centrée réduite ??.

2. Variables mesurées à plusieurs reprises sur les mêmes individus ou unités d'observation au fil du temps

2.2.3 Win odds

Le **Win Odds** évoqué par Dong et al. (2020) et Brunner et al. (2021) est une adaptation des win ratio incluant la possibilité d’avoir des égalités. Les deux papiers proposent d’inclure dans le ratio la moitié du nombre d’égalité au numérateur et au dénominateur comme suit :

$$R_{wo} = \frac{N_w + 0.5N_t}{N_l + 0.5N_t}$$

Exemples

Dans cette section nous allons voir trois exemples simple tiré du papier de Brunner et al. (2021). Un premier exemple pour traiter de l’approche apparié, un exemple pour traiter l’approche non-apparié et un dernier exemple pour montrer la différence entre les wins ratio et les wins odds avec l’approche non-apparié.

Cas	Traitement T	Traitement C
1	(1.7 ; 3.3 ; 3.8 ; 4.9 ; 6.3)	(1.4 ; 2.6 ; 3.7 ; 5 ; 6.4)
2	(2 ; 3 ; 4 ; 5 ; 6)	(1 ; 2 ; 3 ; 4 ; 5)

TABLE 3 – Exemples utilisés

Approche apparié :

On utilise le cas 1, dépourvu d’égalité.

Dans cet exemple simple, nous allons juste comptabiliser $\mathbb{P}(X \succ Y)$ et $\mathbb{P}(X \prec Y)$, ce qui nous donne :

$$N_w = 3$$

$$N_l = 2$$

Et alors, $R_w = 1.5$, ce qui signifie que le traitement T est 1.5 fois meilleur que le traitement C.

Approche non-apparié :

On utilise une nouvelle fois le cas 1 dépourvu d’égalité.

$$N_w = 13$$

$$N_l = 12$$

Et alors, $R_w = 1.08$ et le traitement T est 1.08 fois meilleur que le traitement C.

Win Odds :

On utilisera cette fois le cas 2 et une approche non-apparié.

$$N_w = 15$$

$$N_l = 6$$

$N_t = 4$ Et alors, $R_{WO} = 2.125$, cette fois, le nouveau traitement est 2.125 fois meilleur que le traitement C

Approche	λ_{wr}	λ_{wo}
Cas 1		
apparié	1.5	1.5
non-apparié	1.08	1.08
Cas 2		
apparié	∞	∞
non-apparié	2.5	2.125

TABLE 4 – Table des exemples

2.3 Conclusion

On peut ainsi apprécier l’utilité et la complémentarité de ces trois méthodes dans le cadre d’une étude clinique.

Ces méthodes ont en commun l'attribution d'un score permettant d'évaluer si le traitement testé est supérieur, inférieur ou équivalent au traitement de référence. La différence entre elles réside dans la manière dont cette évaluation est effectuée.

3 Simulations de données cliniques

Cette section est en 2 parties. Une partie sur le rapport et une partie sur mon github (<https://github.com/Arthur-tena/Comparaison-par-paire>).

Lors de simulation d'étude clinique, il est important dans un premier temps de dégager des scénarios pouvant arriver. C'est pourquoi, les trois scénarios que nous avons choisis sont les suivants :

1. **scénario 1** : le traitement T est uniformément meilleur que le traitement C.
2. **scénario 2** : le traitement C est uniformément meilleur que le traitement T.
3. **scénario 3** : les deux traitements sont similaires.

Nous nous sommes concentrés sur deux critères de jugement, simulés sous forme de variables. Étant donné les trois types de variables disponibles (tte, continue et binaire), nous avons exploré six paires différentes afin d'examiner toutes les combinaisons possibles.

Il faut ensuite générer les variables suivants les différents scénarios. Les façons de générer les variables sont exprimé dans la section A.2. Nous avons choisis comme généralités qu'une valeur supérieure est meilleure parmi les variables continues.

Afin de vérifier que les variables générées suivent bien le scénario choisis, nous avons visualisé la distribution des vecteurs T et C.

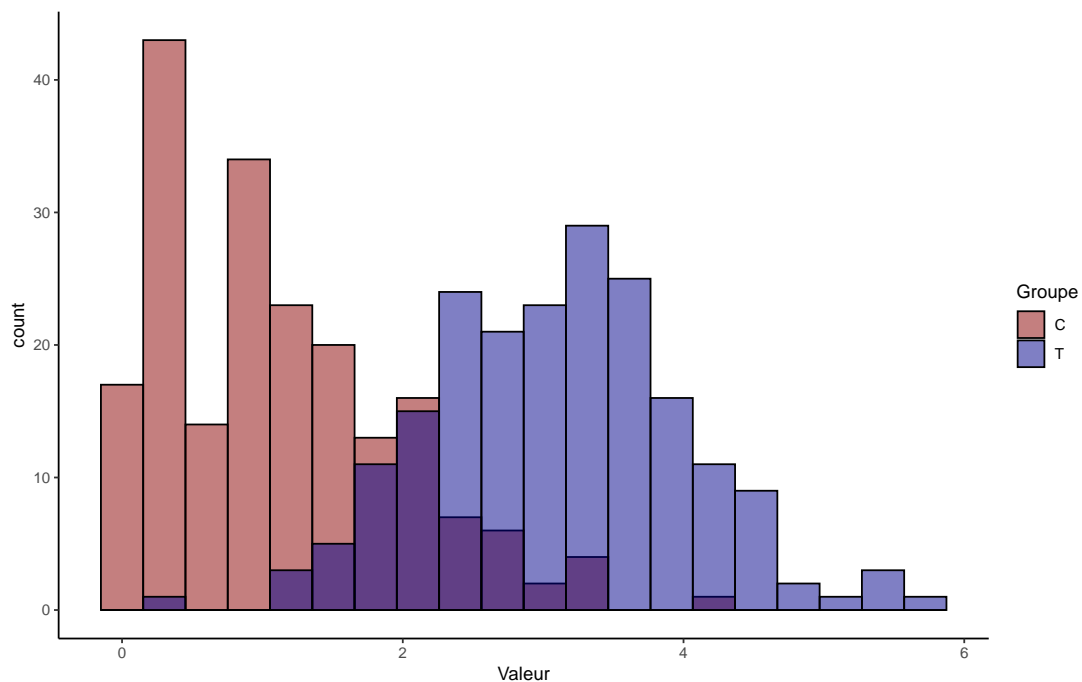


FIGURE 1 – Exemple de visualisation dans le scénario 1 pour le couple de variable (continue,continue) du premier critère de jugement

On constate ici que ce couple de variables du critère de jugement prioritaire correspond bien au scénario 1, car la distribution des valeurs de T est supérieure à celle de C.

A Annexe

A.1 Article

Explication de la GPC dans le cas avec un seul outcome

Comparaison par paire	Paire
$X_i = 1, Y_j = 0$	favorable
$X_i = 1, Y_j = 1$	neutre
$X_i = 0, Y_j = 0$	neutre
$X_i = 0, Y_j = 1$	défavorable

TABLE 5 – GPC pour des variables binaires

Comparaison par paire	Paire
$X_i - Y_j > \tau$	favorable
$ X_i - Y_j \leq \tau$	neutre
$X_i - Y_j < -\tau$	défavorable

TABLE 6 – GPC pour des variables continues

Où X_i dénote le résultat du i -ème individu $i \in \{1, \dots, n\}$ du groupe T et Y_j dénote le résultat du j -ème individus $j \in \{1, \dots, m\}$ du groupe C . Pour les variables continues, on prendra un seuil τ dépendant de l'étude et de ce que l'on veut.

Dans le cadre de variable tte, des censures à droite peuvent arriver, c'est pourquoi les variables ϵ et η dénotent la non-censure de X et Y respectivement c'est à dire les observations non-censurées. Et on note X'_i (resp. Y'_j) les observations censurées du groupe T (resp. C).

Censure de X et Y	Comparaison par paire	Paire
$\epsilon_i = 1, \eta_j = 1$	$X_i - Y_j > \tau$	favorable
	$ X_i - Y_j \leq \tau$	neutre
	$X_i - Y_j < -\tau$	défavorable
$\epsilon_i = 0, \eta_j = 1$	$X'_i - Y_j > \tau$	favorable
	$ X'_i - Y_j \leq \tau$	non-informative
	$X'_i - Y_j < -\tau$	non-informative
$\epsilon_i = 1, \eta_j = 0$	$X_i - Y'_j > \tau$	non-informative
	$ X_i - Y'_j \leq \tau$	non-informative
	$X_i - Y'_j < -\tau$	défavorable
$\epsilon_i = 0, \eta_j = 0$	$X'_i - Y'_j > \tau$	non-informative
	$ X'_i - Y'_j \leq \tau$	non-informative
	$X'_i - Y'_j < -\tau$	non-informative

TABLE 7 – GPC pour des variables tte

Rappel sur les tests de Wilcoxon, Mann-Witney et GGW

Comme mentionné plus haut, la GPC peut être vu comme des cas particulier des tests de Wilcoxon et de GGW dans le cas où $\tau = 0$. C'est ce que nous allons regarder. Dans le cadre de variables binaires, un test de Fisher sera effectué à la condition qu'il y ait le même nombre d'individus dans les deux groupes T et C .

Dans le cadre de variable continue, nous utiliserons des tests de rang de Wilcoxon ou des des tests de Mann-Whitney.

Pour rappel, le test de rang de Wilcoxon est construit comme suit. Par soucis de simplicité, nous n'aborderons pas le cas où les variables sont stratifiées. Soit $S_1 < S_2 < \dots < S_n$ les rangs des observa-

tions du groupe T, la statistique de Wilcoxon est alors

$$W_S = \sum_{i=1}^n S_i$$

La statistique de Mann-Witney s'écrit

$$W_{MW} = W_S - \frac{n \cdot (n+1)}{2}$$

Dans le cadre de variable tte, Gehan Gehan (1965) a proposé une généralisation de la statistique de Wilcoxon en incluant la possibilité d'avoir des événements censurés. Gehan propose alors de séparer les observations de X et de Y suivant si elles sont censurées ou non. On aura alors :

$$\begin{aligned} x'_1, \dots, x'_{r_1}, & \quad r_1 \text{ censurés} \\ x_{r_1+1}, \dots, x_n, & \quad n - r_1 \text{ échec} \\ y'_1, \dots, y'_{r_2}, & \quad r_2 \text{ censurés} \\ y_{r_2+1}, \dots, y_m, & \quad m - r_2 \text{ échec} \end{aligned}$$

où les valeurs de x (resp. y) sont les temps jusqu'à l'échec des personnes ayant pris le traitement T (resp. C), x' (resp. y') est le temps jusqu'à la censure des personnes ayant pris le traitement T (resp. C).

Gehan définit alors sa statistique comme suit :

$$U_{ij} = \begin{cases} 1 & \text{si } x_i > y_j \text{ ou } x'_i \geq y_j \\ -1 & \text{si } x_i < y_j \text{ ou } x_i \leq y'_j \\ 0 & \text{si } x_i = y_j \text{ ou } (x'_i, y'_j) \text{ ou } x'_i < y_j \text{ ou } y'_j < x_i \end{cases} \quad (4)$$

On peut alors calculer la statistique W , $W = \sum_{i,j} U_{i,j}$.

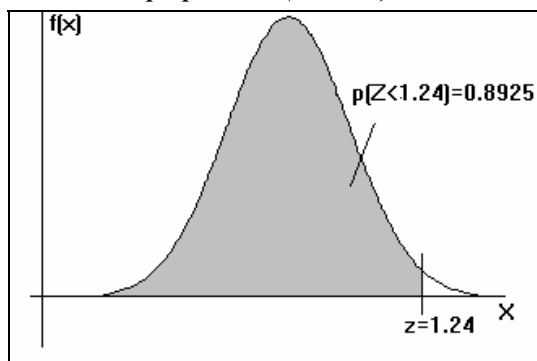
Table de la loi normale centrée réduite

©Université René Descartes

1/3

TABLE DE LA LOI NORMALE CENTREE REDUITE

Lecture de la table: Pour $z=1.24$ (intersection de la ligne 1.2 et de la colonne 0.04), on a la proportion $P(Z < 1,24) = 0.8925$



$P(Z > 1,96) = 0,025$
 $P(Z > 2,58) = 0,005$
 $P(Z > 3,29) = 0,0005$

Rappels:

1/ $P(Z > z) = 1 - P(Z < z)$ et 2/ $P(Z < -z) = P(Z > z)$

Exemple: Sachant $P(Z < 1,24) = 0,8925$, on en déduit:

1/ $P(Z > 1,24) = 1 - P(Z < 1,24) = 1 - 0,8925 = 0,1075$

2/ $P(Z < -1,24) = P(Z > 1,24) = 0,1075$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99987	0,99988	0,99989
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997
4,0	0,99997	0,99997	0,99997	0,99997	0,99997	0,99997	0,99998	0,99998	0,99998	0,99998

A.2 Simulations de données cliniques

```
gener_tte=function(mu, lambda=1){  
  X=rexp(n, rate=lambda)  
  C=rexp(n, rate = mu)  
  TT=pmin(X,C)  
  delta=as.numeric(X==TT)  
  return(data.frame(TT, delta))  
}  
  
gener_continue=function(mean=1, sd=0){  
  X=abs(rnorm(n, mean=mean, sd=sd))  
  X=factor(X)  
  return(data.frame(X))  
}  
  
gener_binom=function(prob){  
  X=rbinom(n, size = 1, prob = prob)  
  return(data.frame(X))  
}
```

Nous avons les possibilités de générer les variables **tte** en utilisant des lois exponentielles et les variables **continues** en utilisant la valeur absolue d'une lois normale dont les paramètres doivent être explicités.

Bibliographie

Références

- Pocock, Stuart J and Ariti, Cono A and Collier, Timothy J and Wang, Duolao(2012). The win ratio : a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *European heart journal*, 33(2), 176–182
- Buyse, Marc (2010). Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. *Statistics in Medicine*, 29(30), 3245–3257.
- Finkelstein, Dianne M. & Schoenfeld, David A. (1999). Combining mortality and longitudinal measures in clinical trials. *Statistics in Medicine*, 18(11), 1341–1354.
- Gehan, Edmund A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, 52(1-2), 203–224.
- McCoy, C. Eric (2018). Understanding the use of composite endpoints in clinical trials. *Western Journal of Emergency Medicine*, 19(4), 631.
- Brunner, Edgar, Vandemeulebroecke, Marc & Mütze, Tobias (2021). Win odds : an adaptation of the win ratio to include ties. *Statistics in Medicine*, 40(14), 3367–3384.
- Buyse, Marc, Saad, Everardo D., Peron, Julien, Chiem, Jean-Christophe, De Backer, Mickaël & Cantagallo, Eva & Ciani, Oriana (2021). The Net Benefit of a treatment should take the correlation between benefits and harms into account. *Journal of Clinical Epidemiology*, 137, 148–158.
- Dong, Gaohong, Hoaglin, David C., Qiu, Junshan, Matsouaka, Roland A., Chang, Yu-Wei, Wang, Jiuzhou & Vandemeulebroecke, Marc (2020). The win ratio : on interpretation and handling of ties. *Statistics in Biopharmaceutical Research*.