



**HAL**  
open science

# Évaluation séquentielle de tests de diagnostics médicaux à réponse binaire

Mamadou Bassirou Diallo

► **To cite this version:**

Mamadou Bassirou Diallo. Évaluation séquentielle de tests de diagnostics médicaux à réponse binaire. Méthodologie [stat.ME]. 2011. dumas-00618523

**HAL Id: dumas-00618523**

<https://dumas.ccsd.cnrs.fr/dumas-00618523v1>

Submitted on 13 Sep 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# RAPPORT DE MEMOIRE

Evaluation séquentielle de  
tests de diagnostics  
médicaux à réponse binaire

**Etudiant: DIALLO Mamadou Bassirou**

**Maitre de mémoire : Mr Frédéric  
BERTRAND**

## Sommaire

I	Partie 1 .....	3
	I.1) Notion de Sensibilité .....	3
	I.2) Notion de spécificité .....	3
	I.2.1) Valeur prédictive positive (VPP).....	4
	I.2.2) Valeur prédictive négative (VPN) .....	4
	I.3) Phase 2 : hypothèse et modèle à un stade .....	4
	I.4) Test d'intersection-union .....	5
	I.4.1) Théorème 1 :.....	6
	I.4.2) Démonstration du théorème 1 :.....	6
	I.4.3) Théorème 2.....	6
	I.4.4) Démonstration du théorème 2.....	6
	I.5) Procédure de test à plusieurs stades : .....	8
	I.5.1) Modèle à deux stades .....	9
	I.5.2) Calcul des valeurs critiques .....	9
	I.5.3) Exemple .....	11
	I.6) Modèle optimal à deux stades :.....	11
II	RESULTAT .....	14
III	Bibliographie.....	15

## I Partie 1

### I.1) Notion de Sensibilité

La sensibilité d'un test est déterminée sur une population de patients dont on sait qu'elle est porteuse de la maladie M parce qu'elle a subi un test de référence. Elle est définie par la proportion (%) de patients qui ont la maladie recherchée et dont le test est positif, en d'autres termes par la proportion de patients malades de la maladie M que le test détecte correctement (vrais positifs). Par opposition, la proportion de patients porteurs de la maladie M que le test n'a pas identifiés est des résultats faussement négatifs.

La sensibilité du modèle est estimée par la formule :  $P^{\hat{D}} = X^D / n^D$

### I.2) Notion de spécificité

La spécificité d'un test est déterminée sur une population de patients dont on sait qu'elle n'est pas porteuse de la maladie M parce qu'elle a subi un test de référence. Elle est définie par la proportion (%) de patients qui n'ont pas la maladie recherchée et dont le test est négatif, en d'autres termes par la proportion de patients non malades de la maladie M que le test détermine correctement (vrais négatifs). Par opposition, la proportion de patients non porteurs de la maladie M chez qui le test est positif est un résultat faussement positif.

La spécificité du modèle est estimée par la formule :  $P^{\hat{\bar{D}}} = X^{\bar{D}} / n^{\bar{D}}$

Dans l'idéal, on aimerait que le test soit très sensible et très spécifique. Cependant, les tests qui ont à la fois une sensibilité et une spécificité très élevées sont rares. On sera donc conduit à sélectionner un test dont les caractéristiques sont appropriées à la situation clinique.

Plus le test est sensible, moins il comporte de faux négatifs, et mieux il permettra, s'il est négatif, d'exclure la maladie : un moyen mnémotechnique de s'en souvenir est le vocable « SnOut » en anglais, qui regroupe l'abréviation de la sensibilité (« Sn ») et le mot « out » qui désigne le fait d'exclure (« rule out ») un diagnostic.

Par exemple, dans des conditions de dépistage, on choisira de préférence un test très sensible, puisque l'objectif est la détection du plus grand nombre possible de sujets porteurs de la maladie. En revanche, on exigera du test de confirmation effectué chez les sujets dont le test de dépistage s'est révélé positif une spécificité très élevée, afin de minimiser le risque de faux positifs. De même, il est fréquent de devoir combiner plusieurs tests aux propriétés complémentaires dans une stratégie diagnostiques.

Sensibilité et spécificité décrivent les caractéristiques d'un outil diagnostique et sont dérivées de populations de patients dont on connaît le statut de porteur ou non de la maladie M. Or, en pratique, le clinicien est confronté au résultat du test, mais ne sait évidemment pas si le patient a réellement la maladie recherchée ou non. Il doit donc connaître, si le test est positif, quelle est la probabilité que le patient soit réellement porteur de la maladie M ; et si le test est négatif, quelle est la probabilité que le patient ne soit pas porteur de la maladie M. Ces renseignements sont donnés par les notions de valeur prédictive positive et valeur prédictive négative d'un test.

### **I.2.1) Valeur prédictive positive (VPP)**

La probabilité d'avoir la maladie M en cas de test positif s'appelle la valeur prédictive positive (VPP) d'un test. Elle est donnée par le rapport des vrais positifs sur l'ensemble des tests positifs.

### **I.2.2) Valeur prédictive négative (VPN)**

La probabilité de ne pas souffrir de la maladie M en cas de test négatif s'appelle la valeur prédictive négative (VPN) d'un test. Elle est donnée par le rapport des vrais négatifs sur l'ensemble des tests négatifs

En résumé, les notions de sensibilité et spécificité sont déterminées sur une population dont on connaît la présence ou l'absence de la maladie recherchée et représentent donc les caractéristiques intrinsèques d'un test donné. Les valeurs prédictives positive et négative, en revanche, fournissent la probabilité qu'un patient ait ou n'ait pas la maladie recherchée en fonction du résultat positif ou négatif du test.

Sensibilité et spécificité décrivent les qualités intrinsèques du test.

Les valeurs prédictives dépendent des qualités intrinsèques du test et de la probabilité pré-test, c'est à dire de la prévalence de la maladie dans la population étudiée. Elles décrivent l'utilité diagnostique du test dans une population donnée.

## **I.3)Phase 2 : hypothèse et modèle à un stade**

Dans ce modèle, on sélectionne au hasard deux échantillons  $n^D$  et  $n^{\bar{D}}$  dans une population malade et non malade respectivement. Donc nous observons  $X^D$  et  $X^{\bar{D}}$  qui sont deux variables binomiales indépendantes. Ces deux variables représentent le nombre de sujet qui sont correctement classé par le test selon leurs vrais statuts.

Si on dépasse les valeurs critiques, alors on rejette  $H_0$ .

La taille de l'échantillon doit être assez grand pour assurer un niveau  $\alpha$  et une puissance  $1 - \beta$  pour rejeter  $H_0$  si l'exactitude du diagnostique du test médical est à des niveaux précis de sensibilité  $P_1^D$  et de spécificité  $P_1^{\bar{D}}$  dans l'espace de paramètre alternative.

La fonction puissance d'un modèle à un stade est une fonction à des paramètres  $P^D$  et  $P^{\bar{D}}$  :

$$\begin{aligned}\beta(P^D, P^{\bar{D}}) &= P(X^D > b^D, X^{\bar{D}} > b^{\bar{D}}) \\ &= [1 - B(b^D, n^D, P^D)][1 - B(b^{\bar{D}}, n^{\bar{D}}, P^{\bar{D}})]\end{aligned}\quad (1)$$

Où  $b^D$  et  $b^{\bar{D}}$  sont les valeurs critiques du test statistiques  $(X^D, X^{\bar{D}})$

et  $B(b^D, n^D, P^D) = \sum_{i=0}^b \binom{n}{i} P^i (1-p)^{n-i}$  est la fonction de probabilité cumulative.

A cause de l'indépendance des cas et des contrôles, la fonction puissance dans (1) est monotone car elle augmente avec l'un ou l'autre des paramètres tandis que l'autre reste constant.

La caractéristique de cette augmentation de la fonction puissance garantie que la puissance atteint son minimum dans l'espace des paramètres nulles; donc la fonction puissance ne devrait pas être plus contrôlée que le niveau  $\alpha$ . Par conséquent les valeurs critiques doivent satisfaire les équations suivantes :

$$\begin{aligned}\underbrace{\sup}_{\theta_0} \beta(P^D, P^{\bar{D}}) &= \alpha \\ \beta(P_1^D, P_1^{\bar{D}}) &= 1 - \beta\end{aligned}$$

Pour déterminer la région de rejet d'un test de diagnostique, nous allons étudier le test d'intersection-union proposé par Casella et Berger.

#### I.4) Test d'intersection-union

On donne  $X \rightarrow f(x|\theta)$ , On suppose que  $H_0$  est exprimée comme une union de  $k$  ensembles.

$$H_0 : \theta \in \theta_0 \text{ contre } H_A : \theta \in \theta_0^c = \bigcap_{i=1}^k \theta_i^c \quad (1)$$

On suppose que pour chaque  $i$ ,  $R_i$  est les région de rejet pour le test de :

$$H_{0i} : \theta \in \theta_i \text{ contre } H_{Ai} : \theta \in \theta_i^c$$

Casella et Berger ont prouvé que la région de rejet pour un test d'intersection-union de  $H_0$  contre  $H_1$  est  $R = \bigcap_{i=1}^k R_i$ .

#### I.4.1) Théorème 1 :

Si  $R_i$  est de niveau  $\alpha$  pour  $H_{0i}$ , Pour  $i = 1, \dots, k$ , alors le test intersection-union avec région de rejet  $R = \bigcap_{i=1}^k R_i$  est de niveau  $\alpha$  pour le test de  $H_0$  contre  $H_1$  dans (1).

#### I.4.2) Démonstration du théorème 1 :

On donne  $\theta \in \theta_0 = \bigcup_{i=1}^k \theta_i$ . Ainsi, pour certains  $i = 1, \dots, k$ , ( $i = i'$ ),  $\theta \in \theta_{i'}$ .

Donc,  $P_\theta(\bigcap R_i) \leq P_\theta(\bigcap R_{i'}) \leq \alpha$

Comme  $\theta \in \theta_0$  a été choisi arbitrairement, le test d'intersection-union est de niveau  $\alpha$  avec  $\sup_{\theta \in \theta_0} P_\theta(\bigcap R_i) \leq \alpha$

#### I.4.3) Théorème 2

Pour  $i = 1, \dots, k$ , on suppose que  $R_i$  est la région de rejet de niveau  $\alpha$  pour tester  $H_0$  contre  $H_{1i}$  pour tout  $i = 1, \dots, k$ ,  $j \neq i$ , on suppose  $R_j$  la région de rejet est de niveau  $\alpha$  pour tester  $H_{0j}$  contre  $H_{A1}$ . On suppose  $\dots, \theta_l, l = 1, \dots, k$ , in  $\theta_i$

$$\lim_{l \rightarrow +\infty} P_{\theta_l}(R_i) = \alpha$$

Et pour tout  $j = i, \dots, k$ ,  $j \neq i$

$$\lim_{l \rightarrow +\infty} P_{\theta_l}(R_j) = 1$$

Alors le test intersection-union avec la région de rejet  $R = \bigcap_{i=1}^k R_i$  est de niveau  $\alpha$  pour tester  $H_0$  contre  $H_1$ .

#### I.4.4) Démonstration du théorème 2

$$\lim_{l \rightarrow +\infty} P_{\theta_l}(\bigcap_{m=1}^k R_m) \geq \underbrace{\lim_{l \rightarrow +\infty} \sum_{m=1}^k P_{\theta_l}(R_m) - (k-1)}_{\text{Inégalité de Bonferroni}}$$

Inégalité de Bonferroni

$$= \alpha + (k+1) - (k-1) = \alpha$$

Donc le test ne rejette que seulement si tous les tests rejettent.

Nous utilisons la distribution de probabilité exacte pour déterminer les valeurs critiques, depuis qu'elle produit plus de résultat exact que l'approximation normal à cause de la petitesse de la taille des échantillons dans la deuxième phase d'étude.

Cependant, la distribution de probabilité binomiale exacte ne fournit pas une forme fermée de valeur critique avec la puissance  $1 - \beta$ .

Une méthode numérique pour résoudre la taille de l'échantillon requis est proposée comme suit :

Nous supposons que le ratio d'allocation de cas et contrôle dans l'étude est spécifié ; dans ce cadre  $n^D = n^{\bar{D}}$  si le ratio est égal 1.

Pour un test de taille d'échantillons donnée avec  $n^D$  et  $n^{\bar{D}}$  qui représente les malades et les sains, les valeurs critiques  $b^D$  et  $b^{\bar{D}}$  peuvent être obtenues en inversant les 2 fonctions suivantes :

$$P((X^D > b^D) | P_0^D) = 1 - B(b^D, n^D, P_0^D) = \alpha$$

$$P((X^{\bar{D}} > b^{\bar{D}}) | P_0^{\bar{D}}) = 1 - B(b^{\bar{D}}, n^{\bar{D}}, P_0^{\bar{D}}) = \alpha$$

Alors,  $\beta$  au point  $(P_1^D, P_1^{\bar{D}})$  dans l'espace des paramètres alternatives peut être calculé ( $n = n^D + n^{\bar{D}}$ ).

L'augmentation de la taille totale de l'échantillon augmente la puissance de l'étude.

La taille dont on a besoin pour l'étude à un seul stade est choisi pour être les valeurs minimales de  $n^D$  et  $n^{\bar{D}}$  qui satisfont les critères d'erreur.



### I.5) Procédure de test à plusieurs stades :

Pour permettre une fin rapide d'une évolution d'un test médicale, une procédure de test à plusieurs stades peut-être employée.

On suppose que les sujets sont pris dans des groupes,  $n_k^D$  et  $n_k^{\bar{D}}$  étant respectivement la taille des échantillons cumulés pour les malades et les saints dans le premier groupe k.

Plutôt que d'utiliser  $X^D$  et  $X^{\bar{D}}$ , à partir de maintenant, nous utiliserons  $S_k^D$  et  $S_k^{\bar{D}}$  pour représenter le nombre de sujets malades et saints classés correctement, respectivement à partir des données observées jusqu'au  $k^{ième}$  groupe.

$k$  est le nombre maximal pré-spécifié des observations planifiées pour l'étude.

Si au  $k^{ième}$  stade,  $S_k^D > b_k^D$  et  $S_k^{\bar{D}} > b_k^{\bar{D}}$  on arrête le test et  $H_0$  est rejetée.

Autrement, les mesures de  $n_{k+1}^D - n_k^D$  sujets malades et  $n_{k+1}^{\bar{D}} - n_k^{\bar{D}}$  de sujets saints sont considérés et les 2 statistiques  $S_{k+1}^D$  et  $S_{k+1}^{\bar{D}}$  sont calculées en fonctions de tous les données disponibles jusqu'au  $(k + 1)^{ième}$  stade.

Si le test ne s'est pas arrêté tôt alors il s'arrêtera au  $k^{ième}$  stade et rejettera  $H_0$  si  $S_k^D > b_k^D$  et  $S_k^{\bar{D}} > b_k^{\bar{D}}$ . Ici  $b_k^D$  et  $b_k^{\bar{D}}$ ,  $k = 1, 2, 3 \dots \dots, K$ , sont les valeurs critiques qui satisfont les critères d'erreurs.

A la différence d'une phase deux typique d'un test clinique ou le centre est souvent une seule issue binaire (par exemple la réponse à un traitement thérapeutique), l'évaluation de la phase deux d'un test de diagnostics médical implique d'avantage de modèle paramétrique et par conséquent la fonction devient plus compliqués que le modèle à plusieurs stade.

De plus l'étude de la phase 2 contient souvent relativement un petit nombre de sujets. Nous nous focaliserons sur les modèles à deux stades qui sont pratiquement plus attrayant car moins compliqués à calculer et leur potentielle application en médecin.

Nous explorons en bas, les erreurs de Lan et Demets qui consiste à faire des approches pour déterminer la taille de l'échantillon et les valeurs critiques pour tester  $H_0$ .

### I.5.1) Modèle à deux stades

Dans une étude à deux stades, l'hypothèse nulle  $H_0$  est rejetée, si au premier stade on a  $\{S_1^D > b_1^D, S_1^{\bar{D}} > b_1^{\bar{D}}\}$  ou  $\{S_1^D \leq b_1^D, S_1^{\bar{D}} \leq b_1^{\bar{D}}\}$  au premier stade  $\{S_2^D > b_2^D, S_2^{\bar{D}} > b_2^{\bar{D}}\}$  au second stade.

En utilisant l'approche d'erreur de Lan et DeMets, on suppose que l'erreur de type 1 pour passer le stade 1 est  $\alpha_1$ . Alors :

$$\sup_{\theta \in \theta_0} P(S_1^D > b_1^D, S_1^{\bar{D}} > b_1^{\bar{D}}) < \alpha_1$$

$$\sup_{\theta \in \theta_0} P(\{S_1^D > b_1^D, \text{or } S_1^{\bar{D}} > b_1^{\bar{D}}\}, S_2^D > b_2^D, S_2^{\bar{D}} > b_2^{\bar{D}}) < \alpha - \alpha_1$$

A chaque stade les erreurs s'ajoutent entre elles pour donner  $\alpha$ . Les exigences de la puissance au point  $\theta_1 = (P_1^D, P_1^{\bar{D}})$  produisent  $1 - \beta$ .

$$P_{\theta_1}(S_1^D > b_1^D, S_1^{\bar{D}} > b_1^{\bar{D}}) + P_{\theta_1}(\{S_1^D > b_1^D, \text{or } S_1^{\bar{D}} > b_1^{\bar{D}}\}, S_2^D > b_2^D, S_2^{\bar{D}} > b_2^{\bar{D}}) = 1 - \beta$$

A cause de l'indépendance des cas et des contrôles, les probabilités peuvent être calculées en termes de fonctions de distribution binomiale. Les détails sont dans l'appendice numéro 3.

Les trois équations déterminent les valeurs et la taille des échantillons en supposant le rapport d'allocation des sujets du premier groupe et le rapport d'allocation des sujets malades et non-malades sont tous les deux connus.

La fonction puissance augmente dans l'équation précédente n'est pas monotone augmentant avec la sensibilité quand la spécificité est constant et vice versa.

Cela signifie que pour certaines tailles d'échantillons, la puissance au point  $(P_2^D, P_2^{\bar{D}})$  est plus grande que celle au point  $(P_1^D, P_1^{\bar{D}})$ , ou  $P_2^D > P_1^D$  et  $P_2^{\bar{D}} > P_1^{\bar{D}}$ .

Donc, la puissance est seulement calculée au niveau spécifié de sensibilité et de spécificité sous l'espace de paramètre alternatif. La caractéristique non-monotonique de la fonction puissance présentée selon les modèles à plusieurs stades complique le compte des valeurs critiques. Dans le stade 2, le taux d'erreurs n'a pas besoin d'être contrôlé plus que  $\alpha - \alpha_1$  pour tout  $(P^D, P^{\bar{D}}) \in \theta_0$ .

### I.5.2) Calcul des valeurs critiques

Puisqu'il n'ya aucune forme fermée pour les valeurs critiques en utilisant une distribution exacte de probabilité binomiale, les valeurs critiques et les tailles d'échantillons doivent être déterminées numériquement.

On supposera que les tailles totales des échantillons,  $n_2^D$  et  $n_2^{\bar{D}}$ , pour les sujets malades et les sujets non-malade sont données dans l'étude, et que pour simplifier

nous considérerons une égalité d'allocation a chaque stade c'est-à-dire  $n_1^D = n_1^{\bar{D}}$  et  $n_2^D = n_2^{\bar{D}}$ ; ce qui nous permettra de modifier facilement la procédure numérique proposée ci-dessous à l'allocation inégale. Dorénavant, nous supposons que le rapport de taille d'échantillon de promotion du stade 1 au stade 2 est connu.

Alors, en spécifiant, l'erreur de passage au stade 1,  $\alpha_1$  va être connue. Par la suite, les valeurs critiques pour le stade 1 peuvent être résolues selon la procédure semblable à celui utilisé dans un modèle de stade simple. Pour chaque paire de valeurs des tailles d'échantillons totales  $n_2^D$  et de  $n_2^{\bar{D}}$ , les valeurs critiques  $b_1^D$  et  $b_1^{\bar{D}}$  sont déterminés en résolvant :

$$P_{P_0^D}(S_1^D > b_1^D) = \alpha_1, \quad P_{P_0^{\bar{D}}}(S_1^{\bar{D}} > b_1^{\bar{D}}) = \alpha_1$$

Dans un stade deux, pour chaque valeurs de nombre entier de  $b_2^D$  dans la gamme  $(b_1^D, n_2^D)$  et  $b_2^{\bar{D}}$  dans celui de  $(b_1^{\bar{D}}, n_2^{\bar{D}})$ , nous pouvons calculer le taux d'erreur maximum de type 1 et la puissance totale à un point spécifique  $\theta_1 = (P_1^D, P_1^{\bar{D}})$ .

Les valeurs critiques  $b_1^D$  et  $b_1^{\bar{D}}$  sont choisies pour satisfaire les contraintes suivantes :

$$\sup_{\theta \in \theta_0} P(\{S_1^D > b_1^D, \text{ or } S_1^{\bar{D}} > b_1^{\bar{D}}\}, S_2^D > b_2^D, S_2^{\bar{D}} > b_2^{\bar{D}}) < \alpha - \alpha_1 \quad (10)$$

$$P_{\theta_1}(S_1^D > b_1^D, S_1^{\bar{D}} > b_1^{\bar{D}}) + P_{\theta_1}(\{S_1^D > b_1^D, \text{ or } S_1^{\bar{D}} > b_1^{\bar{D}}\}, S_2^D > b_2^D, S_2^{\bar{D}} > b_2^{\bar{D}}) \geq 1 - \beta \quad (11)$$

A cause de la discontinuité de la fonction de probabilité binomiale, les valeurs exactes de  $\alpha_1$ ,  $\alpha$  et  $1 - \beta$  ne peuvent être connus, et donc les valeurs critiques sont choisies pour garantir que l'erreur globale de type 1 n'est pas plus grande que  $\alpha$  et que la puissance n'est pas moins que  $1 - \beta$ , avec un minimum de différence.

La valeur qui est à gauche de l'équation (10) peut toujours être évaluée sur une grille de valeurs pour  $(P^D, P^{\bar{D}}) \in \theta_0$  et alors la taille exacte de l'échantillon peut être trouvée.

Finalement, le modèle avec la taille d'échantillon la plus petite et garantissant à la fois l'erreur de type et les contraintes de puissances est choisi.

### 1.5.3) Exemple

Dans le tableau ci-dessous, nous présentons un modèle à deux stades pour une variété de modèle paramétrique. Nous choisissons  $n_2^D = n_2^{\bar{D}} = n_2$  et  $n_1^D = \lfloor n_2^D/2 \rfloor$ ,  $n_1^{\bar{D}} = \lceil n_2^{\bar{D}}/2 \rceil$  et la fonction d'erreur  $\alpha_1 = \alpha t$ , où  $t \in (0,1)$  est le ratio de la taille provisoire de promotion et la taille globale de promotion si une analyse provisoire est effectuée. (Pour un modèle à deux stades considérer ici, nous choisissons  $t = 1/2$  au premier stade, donc l'erreur de type 1 déterminée au premier stade est  $\alpha_1 = \alpha/2$ .) Les Valeurs possibles pour le modèle paramétrique considérer sont  $\alpha = 0.025, 0.05, 0.1$ , et  $\beta = 0.1, 0.2$ . Les tailles de l'échantillon du modèle à un stade simple sont calculées par la méthode d'écrit plus haut, et suppose une égalité d'allocation c'est-à-dire  $n_2^D = n_2^{\bar{D}} = n_f$ . On donne  $N_i = n_i^D + n_i^{\bar{D}}$  la taille totale de l'échantillon jusqu'au stade  $i$ .

Nous introduisons la taille de promotion attendue  $ASN_i$  sous l'hypothèse

$H_i$  ( $i = 0, 1$ ):

$$ASN_i = N_1 P_{H_i}(\text{l'étude arrêtée au stade 1}) + N_2 P_{H_i}(\text{l'étude arrêtée au stade 2}) \quad (12)$$

Le tableau en annexe donne les valeurs obtenues avec l'allocation égale conformément aux deux hypothèses et compare avec la taille de promotion par groupe pour un modèle à un stade simple. En faisant une approche pour l'erreur, la probabilité que l'étude s'arrête au stade 1 est égale à  $P(S_1^D > b_1^D, S_1^{\bar{D}} > b_1^{\bar{D}})$ .

Les résultats présentés sous forme de tableau incluent aussi les probabilités théoriques de l'erreur de types ( $\alpha^*$ ) et la vraie puissance ( $1 - \beta^*$ ) basé sur les valeurs critiques dans le deux stades.

Il est noté qu'un modèle à deux stades exige une plus grande taille d'échantillon qu'un modèle à un stade simple pour accomplir le même niveau de signification et de puissance. Les tailles moyennes d'échantillons sous  $H_1$ ,  $ASN_1$ , dans la table 1, apparaissent plus petites que la taille totale d'échantillon pour le modèle fixé,  $2 \times n_f$  la plupart du temps.

### 1.6) Modèle optimal à deux stades :

Nous avons présenté et avons adopté les méthodes pour les modèles à deux stades et avons proposé l'utilisation du Lan et l'erreur DeMets en passant par l'approche de fonction de déterminer les valeurs critiques. Une des inquiétudes dans une mise à l'essai médicale du diagnostique est que les tailles de l'échantillon des études sont désirées pour être minimisées en raison de la restriction de ressource et du coût de l'étude. À cette fin, l'erreur en passant par l'approche de fonction n'est pas une

méthode optimale du point de vue de la minimisation de la taille de l'échantillon conformément aux hypothèses. La table I montre que les tailles attendues de l'échantillon sous  $H_0$  sont plus grandes que ceux des modèles de tailles fixés

Simon a proposé un critère optimal de deux stades destiné à évaluer l'efficacité d'un médicament thérapeutique simple. Ses critères optimaux doivent minimiser la taille attendue de promotion conformément à l'hypothèse nulle. Nous proposons ici des modèles à deux stades qui minimisent le nombre total attendu de sujets malades et les sujets non-malades quand l'épreuve médicale est montrée pour avoir une basse sensibilité ou spécificité. Une première résiliation de ces épreuves diagnostiques pauvrement exécutées est une stratégie efficace pour sauver des ressources.

Souvent dans les épreuves diagnostiques médicales, les niveaux de tolérance de la sensibilité et de spécificité dépendent sur lequel d'entre eux, on est plus critique. Par exemple : l'ensemble

$$\theta_0 = \{P^D \leq P_0^D \text{ ou } P^{\bar{D}} \leq P_0^{\bar{D}}\}$$

Peut être partitionné en trois régions :

$$A_1 = \{P^D \leq P_0^D, P^{\bar{D}} \leq P_0^{\bar{D}}\}$$

$$A_2 = \{P^D \leq P_0^D, P^{\bar{D}} > P_0^{\bar{D}}\}$$

$$A_3 = \{P^D > P_0^D, P^{\bar{D}} \leq P_0^{\bar{D}}\}$$

Chaque région a une interprétation expressive du point de vue de la médecine diagnostique et doit donc être contrôlée individuellement.

La Région  $A_1$  indique une épreuve médicale non prometteuse tant avec la sensibilité que spécificité au-dessous des niveaux très légèrement acceptables. Les Régions  $A_2$  et  $A_3$  indiquent aussi des épreuves non prometteuses, mais avec des résultats favorables pour la spécificité ou dans la sensibilité. Selon lequel des deux la mesure est plus critique, les niveaux de signification de ces régions peuvent être montés à de différents niveaux.

Le modèle optimal à deux stades que nous proposons ici arrête l'évaluation si la première évidence montre que l'épreuve médicale dans l'enquête ne rencontre pas les critères minimaux pour poursuivre l'étude.

Dans le stade 1, quand les deux statistiques d'essais excèdent les valeurs critiques  $(a_1^D, a_1^{\bar{D}})$ , c'est-à-dire  $S_1^D > a_1^D$  et  $S_1^{\bar{D}} > a_1^{\bar{D}}$  alors le processus d'évaluation continue au stade 2; autrement, l'échantillonnage est arrêté et l'épreuve médicale tester est rejetée pour plus de considération.

Dans le stade 2, quand les deux statistique d'essai excède les valeurs critiques  $(a_2^D, a_2^{\bar{D}})$ , c'est-à-dire  $S_2^D > a_2^D$  et  $S_2^{\bar{D}} > a_2^{\bar{D}}$  alors les investigateurs rejettent  $H_0$  et acceptent l'épreuve pour poursuivre l'étude.

Comparé à la procédure utilisant l'erreur en la fonction dans la Section 3.1, la différence clé avec le modèle optimal de deux stades est la décision de le faire dans le stade 1. Quand l'évidence dans le stade 1 montre la performance favorable, au lieu de rejeter  $H_0$ , comme précédemment, nous recruteront plus de sujets dans l'étude pour obtenir la puissance désirée; autrement, l'étude doit être terminée. Dorénavant, l'hypothèse nulle est rejetée seulement au deuxième stade quand

$$\{S_1^D > a_1^D, S_1^{\bar{D}} > a_1^{\bar{D}} \text{ et } S_2^D > a_2^D, S_2^{\bar{D}} > a_2^{\bar{D}}\}.$$

À cause de l'indépendance des cas et des contrôles, la fonction puissance suivante est le produit de deux fonctions, chacun impliquant seulement un des deux paramètres :

$$\beta(P^D, P^{\bar{D}}) = P\{S_1^D > a_1^D, S_1^{\bar{D}} > a_1^{\bar{D}}, S_2^D > a_2^D, S_2^{\bar{D}} > a_2^{\bar{D}}\}$$

## II RESULTAT

Les méthodes séquentielles sont devenues standard aux essais cliniques, mais n'ont pas été bien développées dans le domaine de la médecine diagnostique. Comme démontré dans ce papier, de tels développements sont nécessaires en raison des considérations éthiques et économiques et de plus, ont besoin des dérivations analytiques spécifiques différentes de ceux dans les applications d'essai cliniques.

Les deux méthodes que nous avons proposées ont des différences fondamentales dans la motivation et dans la définition des règles d'arrêt. L'erreur en passant par l'approche permet la première résiliation d'épreuves de diagnostics apparemment prometteuses avec le contrôle flexible sur l'erreur dans le stade 1. Pourtant, cette procédure ne fournit pas de modèles optimaux dans le sens de la minimisation la grandeur moyenne de promotion.

D'autre part, la deuxième approche qui termine l'étude tôt à cause de l'apparente exactitude, fournit des modèles qui sont optimaux concernant les tailles attendues de l'échantillon conformément à l'hypothèse nulle. Cette deuxième approche utilise la partition de l'espace de paramètre conformément à l'hypothèse nulle et permet d'obtenir des plus hauts niveaux de signification dans la partie de l'espace de paramètre, qui provient d'une plus petite taille d'échantillon. Puisque le modèle optimal à deux stades est plus efficace, il peut être appliqué dans les études préliminaires quand les ressources sont limitées.

### **III Bibliographie**

1. Pocock S. Group sequential methods in the design and analysis of clinical trials. *Biometrika* 1977; 64:191–200.
2. O'Brien P, Fleming T. A multiple testing procedure for clinical trials. *Biometrics* 1979; 35:549–556.
3. Lan K, DeMets D. Discrete sequential boundaries for clinical trials. *Biometrika* 1983; 70:659–663.
4. Simon R. Optimal two-stage designs for phase II clinical trials. *Controlled Clinical Trials* 1989; 10:1–10.
5. Todd S, Whitehead J. Confidence interval calculation for a sequential clinical trial of binary response. *Biometrika* 1997; 84:737–743.
6. Casella G, Berger R. *Statistical Inference*. Chapman & Hall/CRC: New York, 2001.