

# Le poolage

Nicolas Brouard

Mai 1992

## 1 Facteur de pool optimal en fonction de la prévalence

Soit  $p$  la prévalence et  $P$  le facteur de poolage envisagé et  $Pop$  la taille de la population à tester. Alors, le nombre de tests en pool à effectuer est  $N = \frac{Pop}{P}$ . Le nombre des pools positifs parmi l'ensemble des  $N$  pools dépend de la probabilité de trouver un positif dans la population initiale, c'est à dire de la prévalence  $p$ . La distribution aléatoire d'un pool suivant le nombre  $0, 1, \dots, P$  de sérums positifs qu'il peut contenir est une loi binomiale  $B(P, p)$ . Autrement dit la probabilité de former un pool de  $P$  sérums contenant  $i$  sérums positifs est  $q_i = C_P^i p^i (1-p)^{P-i}$ . La probabilité de tirer un pool positif (c'est à dire contenant au moins un sérum malade) ou prévalence apparente des pools est donc  $(1 - (1-p)^P)$ . Le nombre moyen de pools positifs à tester en individuel est donc  $Pop \cdot (1 - (1-p)^P)$ .

On vérifie que le nombre moyen de sérums positifs dans un pool est  $pP = \sum_{i=0}^P i q_i$ . Les courbes  $Pop(\frac{1}{p} + (1 - (1-p)^P))$  sont dessinées sur la figure 1 pour différentes valeurs de la prévalence.

En faisant varier  $p$  et  $P$  on obtient la surface de la figure 2 qui une fois projetée dans le plan  $(P, p)$  en courbe de niveaux donne la figure 3. On remarque qu'avec une prévalence de 1 % il est assez équivalent d'utiliser un facteur de poolage de 4 ou de 8. Si la prévalence est comprise entre 0 et 10 % mais sans précision supplémentaire, il faut utiliser un pool par 4, car les courbes sont très espacées.

## 2 Calcul des sensibilités et spécificités avec poolage

On rappelle que la sensibilité d'un test est la probabilité d'obtenir un test négatif pour un vrai porteur de la maladie et que la spécificité est la probabilité d'obtenir un faux positif. La sensibilité est généralement de l'ordre de 100 % alors que la spécificité est souvent moins bonne, 98 % par exemple. Ainsi lorsque la prévalence est faible, il ne faut pas être étonné de trouver beaucoup plus de faux positifs (près de 2 % de la population) que de vrais ! Ceci explique en partie les prévalences fantaisistes observées parfois et montre l'extrême nécessité d'un test de contrôle, Blot dans le cas du sida, pourtant coûteux. Le coût d'un dépistage dans une population faiblement prévalente dépend du nombre des Blots à effectuer et donc de la spécificité du réactif.

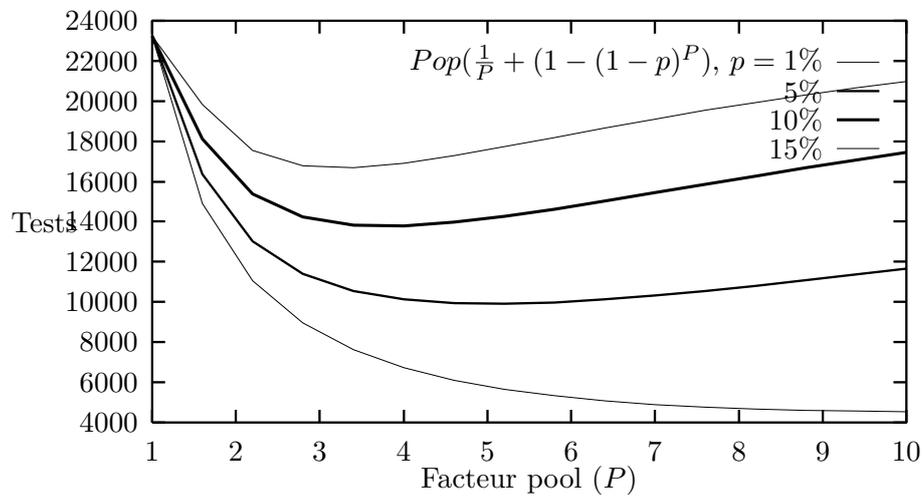


Figure 1: Nombre de tests totaux à effectuer en fonction du facteur de pool pour différents niveaux de prévalence et une population de 23240 individus.

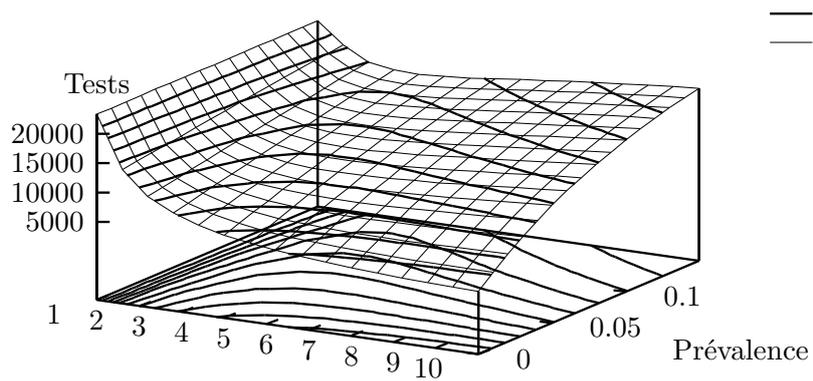


Figure 2: Nombre de tests totaux à effectuer en fonction de la prévalence et du facteur de pool.

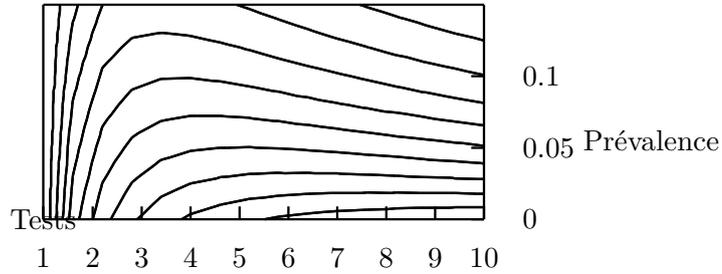


Figure 3: Equivalence entre la prévalence et le facteur de pool pour un même nombre total de tests.

La sensibilité doit être forte pour ne pas avoir de faux négatifs, la spécificité doit être également forte pour ne pas gréver les budgets. En fait, un réactif est dosé pour donner une forte sensibilité au détriment de la spécificité. Un réactif trop sensible réagit souvent sur d'autres anticorps que celui recherché et n'est donc pas spécifique de l'anticorps recherché.

On note désormais,  $s_i$  la sensibilité d'un test effectué sur un sérum individuel  $S_i$  la spécificité.

Le tableau 1 donne la répartition d'un sérum suivant les 4 cas, vrais+ (T+M+),

Table 1:

	M+	M-	
T+	$s_i p$	$(1 - S_i)(1 - p)$	
T-	$(1 - s_i)p$	$S_i(1 - p)$	
	$p$	$(1 - p)$	1

faux+ (T+M-), faux- (T-M+) et vrais- (T-M-).

Il faut qu'en effectuant un pool  $P$ , le tableau final ne se dégrade pas trop, c'est à dire que les faux+ et faux- restent dans des proportions raisonnables.

Dans le cas d'un pool, chaque sérum est dilué dans les  $P - 1$  autres sérums. Ainsi si le réactif étaloné pour une certaine dilution ne supporte pas une dilution plus forte, la "sensibilité" sera détériorée. D'après les biologistes, la sensibilité pourrait décroître de 1 % lorsque le pool est de 10 individus.

Réciproquement la spécificité doit aussi être détériorée, mais là l'enjeu financier est important. En fait, il ne semble pas qu'il y ait beaucoup de théorie en la matière.

Les calculs théoriques exposés ci-dessous dépendent de l'appréciation que

l'ont se fait de la réalité des faux+ et faux-. En fait pour un biologiste, spécificité et sensibilité dépendent beaucoup de la manipulation (temps d'incubation plus ou moins bien respectée, non respect des dilutions, mise à l'abri de la lumière non faite, éluat trop peu agité, etc) et non pas seulement des propriétés des réactifs face à des sérums. C'est bien cette ambiguïté qui nous interdit de faire un calcul facile.

Une première approche consiste à supposer qu'un pool par  $P$  a une sensibilité  $s_p$  (par exemple plus faible de 1 % lorsque le pool est de 10, soit de 99 à 98 %), et une spécificité également plus faible (par exemple de 3 %, soit 95 à 92 %).

Table 2:

	M+	M-	
T+	$s_p(1 - (1 - p)^P)$	$(1 - S_p)(1 - p)^P$	
T-	$(1 - s_p)(1 - (1 - p)^P)$	$S_p(1 - p)^P$	
	$1 - (1 - p)^P$	$(1 - p)^P$	1

En notant M+ les pools qui contiennent au moins un sérum vraiment porteur du virus, on obtient la répartition du tableau 2 pour un pool.

Intéressons maintenant à la désagrégation des pools.

1. Réglons le cas de T-M+ ; on peut faire l'hypothèse que tous les sérums positifs se retrouvent avec  $P - 1$  sérums négatifs.
2. Lorsqu'on dépole T+M+, on trouve une proportion  $\frac{p}{1-(1-p)^P}$  de sérums vraiment positifs et  $1 - \frac{p}{1-(1-p)^P}$  de sérums vraiment négatifs. Par contre lors du test, une fraction  $s_i$  des sérums vraiment positifs sera reconnue comme telle et de même une fraction  $1 - S_i$  des sérums vraiment négatifs seront des faux positifs.
3. Les T-M- se décomposeront en  $P$  sérums individuels T-M-.
4. Pour les T+M-, le problème est plus complexe. En effet il s'agit de pools d'individus tous négatifs mais pour lequel le pool réagit positivement. Si on suppose que la réaction du pool provient d'un sérum particulier parmi les  $P$ , il paraît assez probable qu'en testant chaque élément du pool avec ce même réactif, on trouve une positivité pour ce sérum particulier. Ces pools ne sont pas tirés au hasard mais sélectionnés, il n'est donc pas possible de leurs appliquer une spécificité  $S_i$  ; par contre on peut calculer le nombre moyen de faux+ dans ces pools et d'en déduire le nombre des faux+ individuels. Si on note  $p^+$  la prévalence estimée des faux plus dans la population générale, on trouve un nombre moyen de faux+ parmi les pools faussement + égal à  $P \frac{p^+}{1-(1-p^+)^P}$ . On en déduit ainsi la proportion d'individus qui seront testés à tort positivement ( $P \frac{p^+}{1-(1-p^+)^P} \times \text{T+M-}$ ) et donc la proportion d'individus testés, à juste titre, négativement.

Si, on utilise un réactif différent pour les tests individuels, on peut espérer obtenir une proportion,  $1 - Sp$ , beaucoup plus faible de faux+ et ainsi réduire les coûts des tests de confirmation par Western Blots.

Lorsque on “dépoole” et en admettant qu’on change de réactif pour les tests individuel, on obtient Le tableau 3 donne un récapitulatif de la désagrégation lorsqu’on utilise un autre réactif pour les tests individuels.

Table 3:

		M+ pool		M- pool	
		M+ ind.	M- ind.		
T+p	T+i	$s_p(1 - (1 - p)^P) \frac{p}{1 - (1 - p)^P} s_i P$	$s_p(1 - (1 - p)^P) [1 - \frac{p}{1 - (1 - p)^P}] (1 - S_i) P$	$(1 - S_p)(1 - p)^P (1 - S_i) P$	
	T-i	$s_p(1 - (1 - p)^P) \frac{p}{1 - (1 - p)^P} (1 - s_i) P$	$s_p(1 - (1 - p)^P) [1 - \frac{p}{1 - (1 - p)^P}] S_i P$	$(1 - S_p)(1 - p)^P (1 - S_i) P$	
T-p		$(1 - s_p)(1 - (1 - p)^P)$	$(P - 1)(1 - s_p)(1 - (1 - p)^P)$	$S_p(1 - p)^P P$	
		$1 - (1 - p)^P P$		$(1 - p)^P P$	
					$P$

**Discussions et autres approches** Dans le cas d’une prévalence de 2 % et avec de telles hypothèses, la sensibilité finale se dégrade légèrement et passe à 97,4 % au lieu de 99 % en test individuel, par contre la spécificité s’améliore nettement, même dans le cas où on conserve le même réactif pour les tests individuels et passe à 98,39 au lieu de 95 % en individuel.

- La dégradation de la sensibilité qui conduit à 13 faux négatifs dans le cas de 23240 tests, contre 8 en individuel, résulte de la manière d’interpréter les faux négatifs. Si on considère que les faux négatifs sont déterminés avant la formation des pools, c’est à dire qu’ils correspondent par exemple à des séroconversions (donc faiblement positives), alors chacune de ses séroconversions se retrouvera dès le début dans des pools faussement négatifs et on ne les retrouvera plus lorsqu’on dépoolera les pools positifs contenant des vrais positifs.
- Il peut paraître étonnant qu’avec de telles hypothèses on améliore la spécificité en poolant. Il pourrait bien s’agir d’un artefact. En effet, si on fait l’hypothèse que les faux- sont déterminés avant la formation des pools, alors les pools faussements positifs sont dûs à la répartition des individus faux+ dans les pools. Au lieu de  $(1 - S_p)(1 - p)^P P$  on trouve  $(1 - p) S_i \frac{1}{P \frac{p^+}{1 - (1 - p^+)^P}}$ , soit un nombre beaucoup plus grand.

Ces différentes interprétations conduisent à des variations budgétaires assez importantes, si bien que si la technique de poolage est souvent utilisée en biologie, il n’existe pas beaucoup de modèle théorique tel que celui présenté ici pour évaluer les coûts.