



Présentation des Logiciels

RobPower

Calcul du risque de première espèce et de la puissance d'un protocole avec des animaux apparentés pour différentes méthodes d'analyse d'association

Muller

Calcul des seuils pour les tests multiples



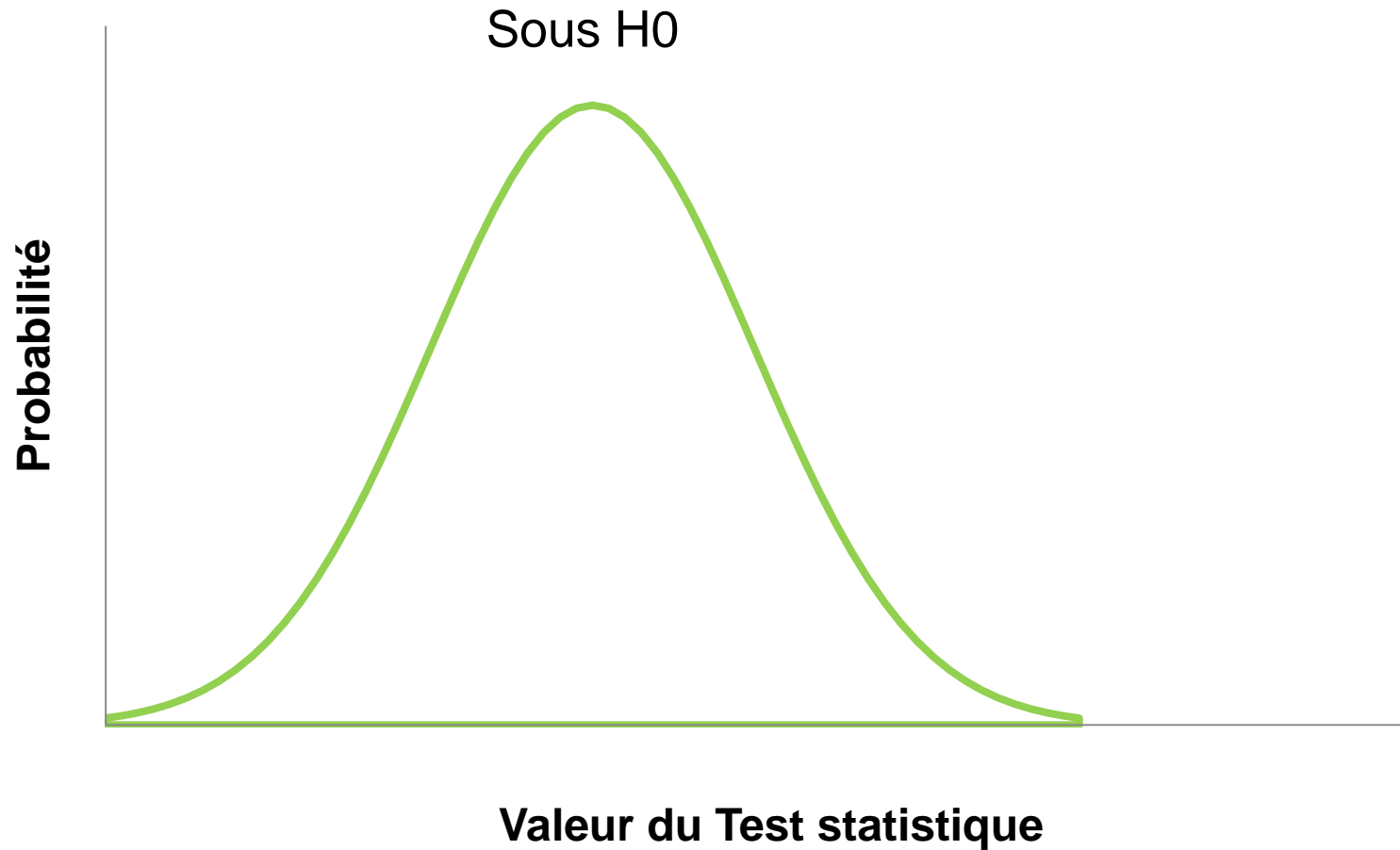
RobPower : Objectif

- Déterminer algébriquement le risque de première espèce et la puissance des principales méthodes utilisées en GWAS, et notamment des méthodes simplifiées, en présence d'apparentement.
- En corolaire, déterminer les effectifs nécessaires pour obtenir une puissance donnée en autorisant un risque de première espèce choisi, selon le degré d'apparentement des animaux d'un protocole.

Teyssedre, S., J. M. Elsen, and A. Ricard. 2012. Statistical distributions of test statistics used for quantitative trait association mapping in structured populations. *Genet. Sel. Evol.* 44.

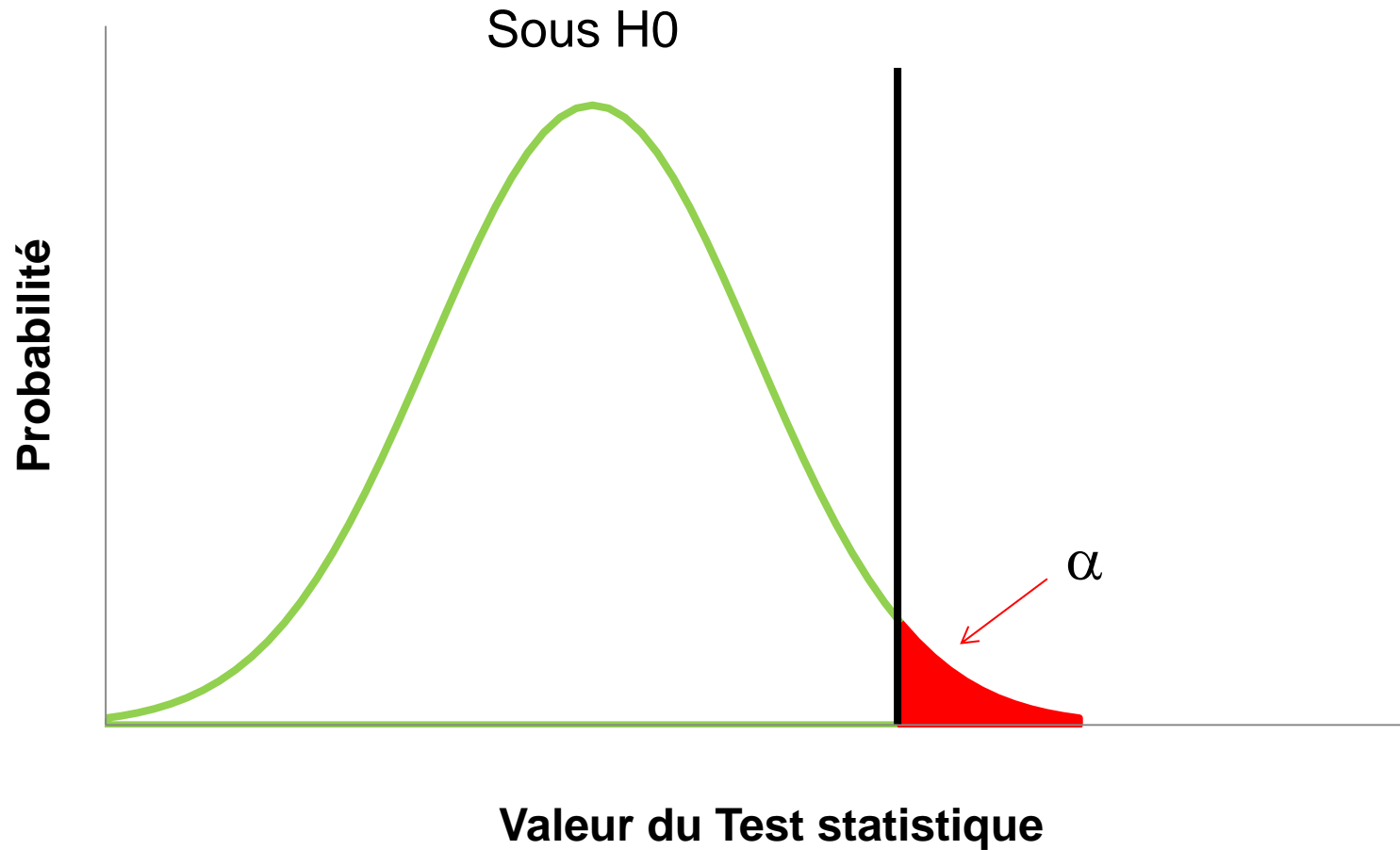
Risque de première espèce α

Puissance



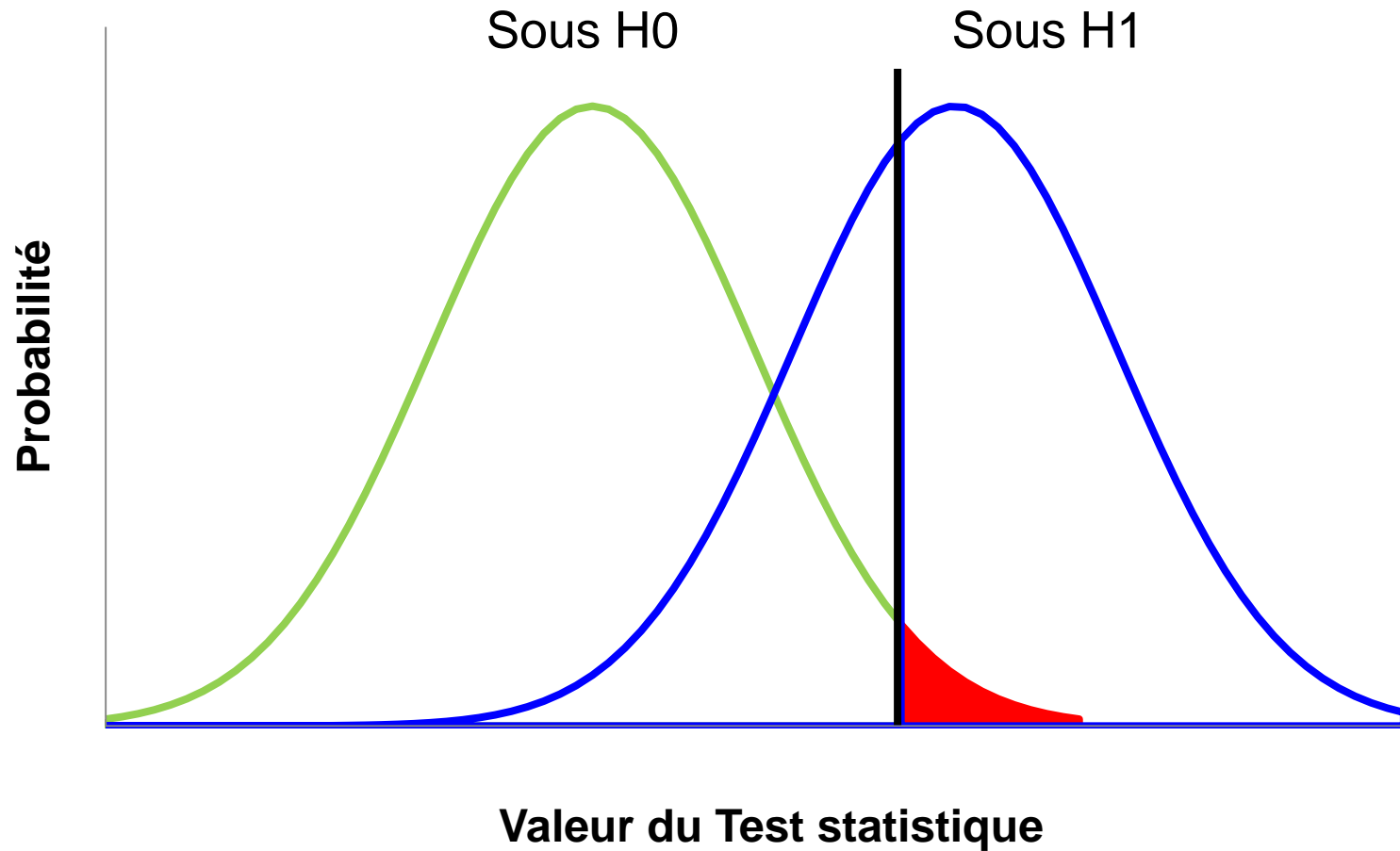
Risque de première espèce α

Puissance



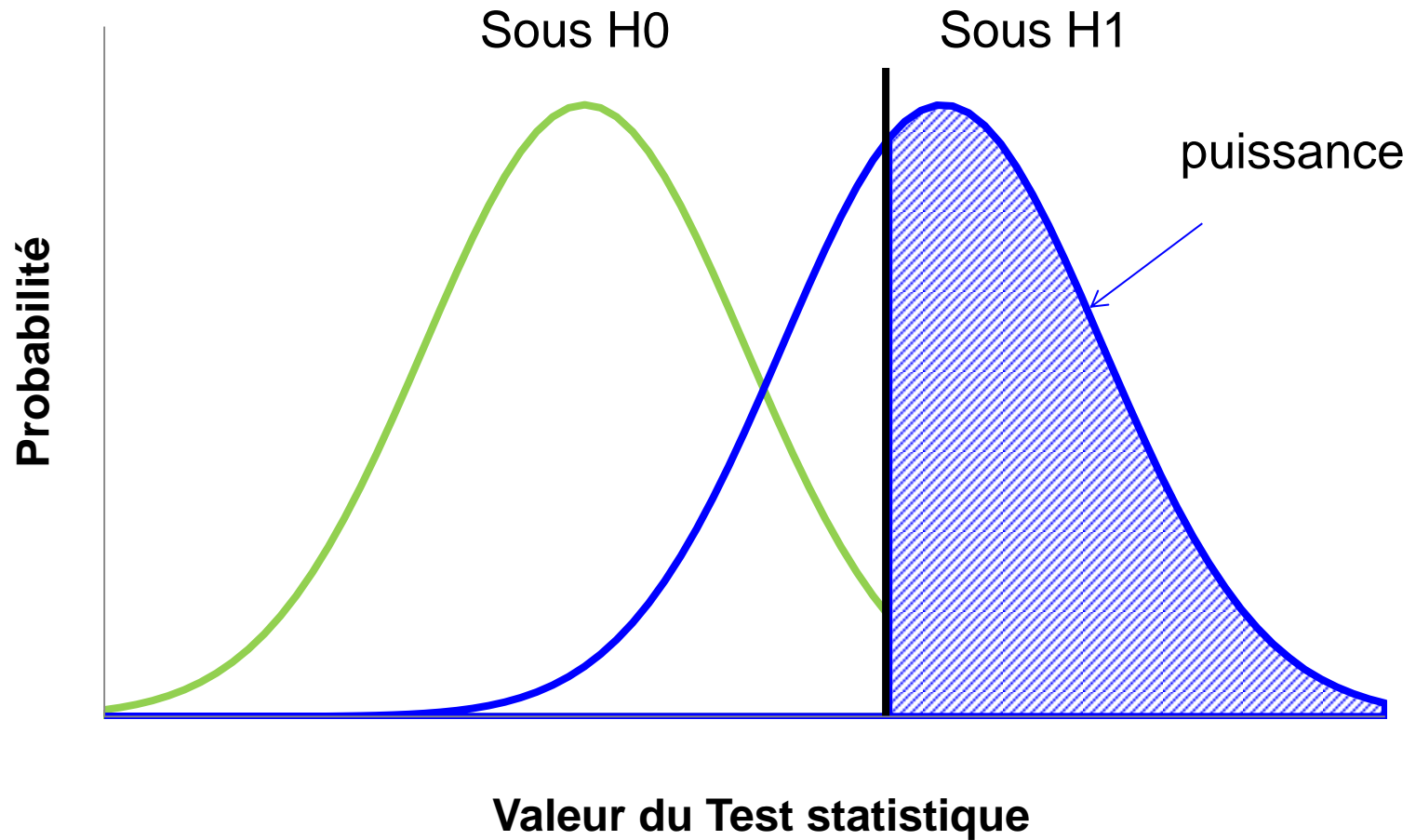
Risque de première espèce α

Puissance



Risque de première espèce α

Puissance



Modèles analysés uniSNP

- *Vrai model*

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{x}\beta + \mathbf{u} + \mathbf{e}$$

$$V(\mathbf{u}) = \mathbf{A}\sigma_u^2$$

Method : models used

- *Vrai modèle*

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{x}\boldsymbol{\beta} + \mathbf{u} + \mathbf{e}$$

$$V(\mathbf{u}) = \mathbf{A}\sigma_u^2$$

- Regression
- Grammar
- QTDT
- FASTA

Method : models used

- *Vrai modèle*

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{x}\beta + \mathbf{u} + \mathbf{e}$$

$$V(\mathbf{u}) = \mathbf{A}\sigma_u^2$$

- Regression

$$\mathbf{y} = \mathbf{1}\mu_1 + \mathbf{x}\beta + \mathbf{e}_1$$

$$V(\mathbf{e}_1) = \mathbf{I}\sigma_{e_1}^2$$

- Grammar

$$\mathbf{y} = \mathbf{1}\mu_2 + \mathbf{u}_2 + \mathbf{e}_2$$

$$V(\mathbf{u}_2) = \mathbf{A}\sigma_{u_2}^2$$

$$\mathbf{e}_2 = \mathbf{1}\mu_3 + \mathbf{x}\beta + \mathbf{e}_3$$

$$V(\mathbf{e}_3) = \mathbf{I}\sigma_{e_3}^2$$

- QTDT

$$\mathbf{y} = \mathbf{1}\mu_4 + \frac{\mathbf{x}_{sire} + \mathbf{x}_{dam}}{2} \beta_b + \left(\mathbf{x} - \frac{\mathbf{x}_{sire} + \mathbf{x}_{dam}}{2}\right) \beta_w + \mathbf{e}_4$$

$$V(\mathbf{e}_4) = \mathbf{I}\sigma_{e_4}^2$$

- FASTA

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{x}\beta + \mathbf{u} + \mathbf{e}$$

$$V(\mathbf{u}) = \mathbf{A}\sigma_u^2$$

$$\sigma_u^2 = \sigma_{u_2}^2$$

Méthode

- Pour chaque modèle, le risque de première espèce α et la puissance sont calculés en supposant le modèle juste.
- Connaissant le vrai modèle, nous calculons l'espérance et la variance de l'estimation de β et de la statistique de test utilisés pour les différents modèles simplifiés.
- Ainsi, nous en déduisons le vrai risque de première espèce et la vraie puissance en fonction du risque supposé α , de l'héritabilité et des relations de parenté.

Exemple : régression

- Regression Model

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{x}\beta + \mathbf{e}_1$$

$$V(\mathbf{e}_1) = \mathbf{I}\sigma_{e_1}^2$$

Exemple : régression

- Regression Model

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{x}\beta + \mathbf{e}_1 \quad V(\mathbf{e}_1) = \mathbf{I}\sigma_{e_1}^2$$

- Estimate of SNP effect

$$\hat{\beta} = (\mathbf{x}'\mathbf{x})^{-1} \mathbf{x}'\mathbf{y}$$

Exemple : régression

- Regression Model

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{x}\beta + \mathbf{e}_1 \quad V(\mathbf{e}_1) = \mathbf{I}\sigma_{e_1}^2$$

- Estimate of SNP effect
- Student test

$$\hat{\beta} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}$$
$$test = \hat{\beta} \sqrt{\frac{\mathbf{x}'\mathbf{x}(n-2)}{\hat{\mathbf{e}}_1'\hat{\mathbf{e}}_1}} = \left(\frac{\hat{\beta}}{\sqrt{V(\hat{\beta})}} \sqrt{\frac{E(\hat{\mathbf{e}}_1'\hat{\mathbf{e}}_1)}{\hat{\mathbf{e}}_1'\hat{\mathbf{e}}_1}} \right)$$

Exemple : régression

- Regression Model

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{x}\beta + \mathbf{e}_1 \quad V(\mathbf{e}_1) = \mathbf{I}\sigma_{e_1}^2$$

- Estimate of SNP effect

$$\hat{\beta} = (\mathbf{x}'\mathbf{x})^{-1} \mathbf{x}'\mathbf{y}$$

- Student test

$$test = \hat{\beta} \sqrt{\frac{\mathbf{x}'\mathbf{x}(n-2)}{\hat{\mathbf{e}}_1'\hat{\mathbf{e}}_1}} = \left(\frac{\hat{\beta}}{\sqrt{V(\hat{\beta})}} \sqrt{\frac{E(\hat{\mathbf{e}}_1'\hat{\mathbf{e}}_1)}{\hat{\mathbf{e}}_1'\hat{\mathbf{e}}_1}} \right)$$

- Then :

$$E(\hat{\beta}) = \beta$$

$$V(\hat{\beta}) = (\mathbf{x}'\mathbf{x})^{-1} \sigma_{e_1}^2$$

$$E(\hat{\mathbf{e}}_1'\hat{\mathbf{e}}_1) = (n-2)\sigma_{e_1}^2$$

Exemple : régression

- Regression Model $\mathbf{y} = \mathbf{1}\mu + \mathbf{x}\beta + \mathbf{e}_1 \quad V(\mathbf{e}_1) = \mathbf{I}\sigma_{e_1}^2$

- Estimate of SNP effect

$$\hat{\beta} = (\mathbf{x}'\mathbf{x})^{-1} \mathbf{x}'\mathbf{y}$$

- Student test

$$test = \hat{\beta} \sqrt{\frac{\mathbf{x}'\mathbf{x}(n-2)}{\hat{\mathbf{e}}_1'\hat{\mathbf{e}}_1}} = \left(\frac{\hat{\beta}}{\sqrt{V(\hat{\beta})}} \sqrt{\frac{E(\hat{\mathbf{e}}_1'\hat{\mathbf{e}}_1)}{\hat{\mathbf{e}}_1'\hat{\mathbf{e}}_1}} \right)$$

- Then :

$$E(\hat{\beta}) = \beta$$

$$V(\hat{\beta}) = (\mathbf{x}'\mathbf{x})^{-1} \sigma_{e_1}^2$$

$$E(\hat{\mathbf{e}}_1'\hat{\mathbf{e}}_1) = (n-2)\sigma_{e_1}^2$$

- The test follows student \rightarrow normal distribution with :

$$test \xrightarrow[n \rightarrow \infty]{} \begin{cases} H_0 (\beta = 0) & N(0,1) \\ H_1 (\beta \neq 0) & N\left(\frac{E(\hat{\beta})}{\sqrt{V(\hat{\beta})}} = \frac{\beta}{\sqrt{(\mathbf{x}'\mathbf{x})^{-1} \sigma_{e_1}^2}}, 1\right) \end{cases}$$

Exemple : régression

- Regression Model $\mathbf{y} = \mathbf{1}\mu + \mathbf{x}\beta + \mathbf{e}_1 \quad V(\mathbf{e}_1) = \mathbf{I}\sigma_{e_1}^2$
- but knowing the true model* $\mathbf{y} = \mathbf{1}\mu + \mathbf{x}\beta + \mathbf{u} + \mathbf{e} \quad V(\mathbf{u}) = \mathbf{A}\sigma_u^2$
- Estimate of SNP effect $\hat{\beta} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}$
- Student test $test = \hat{\beta} \sqrt{\frac{\mathbf{x}'\mathbf{x}(n-2)}{\hat{\mathbf{e}}_1'\hat{\mathbf{e}}_1}}$
- we have now* : $E(\hat{\beta}) = \beta$
 $V(\hat{\beta}) = (\mathbf{x}'\mathbf{x})^{-1}\sigma_e^2 + (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{A}\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\sigma_u^2$
 $E(\hat{\mathbf{e}}_1'\hat{\mathbf{e}}_1) = (n-2)\sigma_e^2 + (tr(\mathbf{A}) - (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{A}\mathbf{x} - \frac{1}{n}\mathbf{1}'\mathbf{A}\mathbf{1})\sigma_u^2$
- The test follows student \rightarrow normal distribution with :

$$test \xrightarrow[n \rightarrow \infty]{} \begin{cases} H_0 (\beta = 0) & N(0, V) \\ H_1 (\beta \neq 0) & N(E, V) \end{cases}$$

Nice formula for the test

Genotypes

$$H_0(\beta = 0)$$

$$N\left(0, \frac{\mathbf{x}'\mathbf{A}\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}h^2 + (1-h^2)}{(1-h^2) + (tr(\mathbf{A}) - (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{A}\mathbf{x} - \frac{1}{n}\mathbf{1}'\mathbf{A}\mathbf{1})\frac{h^2}{n-2}}\right)$$

heritability $h^2 = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$

$$H_1(\beta \neq 0)$$

$$N\left(\frac{\sqrt{\mathbf{x}'\mathbf{x}}\beta / \sigma_y}{\sqrt{(1-h^2) + (tr(\mathbf{A}) - (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{A}\mathbf{x} - \frac{1}{n}\mathbf{1}'\mathbf{A}\mathbf{1})\frac{h^2}{n-2}}}, \frac{\mathbf{x}'\mathbf{A}\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}h^2 + (1-h^2)}{(1-h^2) + (tr(\mathbf{A}) - (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{A}\mathbf{x} - \frac{1}{n}\mathbf{1}'\mathbf{A}\mathbf{1})\frac{h^2}{n-2}}\right)$$

Relationship matrix

Nice formula for the test

$$E_x(\mathbf{x}'\mathbf{A}\mathbf{x}) = \text{tr}(\mathbf{A}'\mathbf{A}) - \frac{2}{n}\mathbf{1}'\mathbf{A}'\mathbf{A}\mathbf{1} + \frac{1}{n^2}\mathbf{1}'\mathbf{A}\mathbf{1}\mathbf{1}'\mathbf{A}\mathbf{1}$$

Expectation of genotypes

$H_0(\beta = 0)$

$$E_x(\mathbf{x}'\mathbf{x}) = \text{tr}(\mathbf{A}) - \frac{1}{n}\mathbf{1}'\mathbf{A}\mathbf{1}$$

$$N\left(0, \frac{\mathbf{x}'\mathbf{A}\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}h^2 + (1-h^2)}{(1-h^2) + (\text{tr}(\mathbf{A}) - (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{A}\mathbf{x} - \frac{1}{n}\mathbf{1}'\mathbf{A}\mathbf{1})\frac{h^2}{n-2}}\right)$$

heritability $h^2 = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$

$H_1(\beta \neq 0)$

$$N\left(\frac{\sqrt{\mathbf{x}'\mathbf{x}}\beta / \sigma_y}{\sqrt{(1-h^2) + (\text{tr}(\mathbf{A}) - (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{A}\mathbf{x} - \frac{1}{n}\mathbf{1}'\mathbf{A}\mathbf{1})\frac{h^2}{n-2}}}, \frac{\mathbf{x}'\mathbf{A}\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}h^2 + (1-h^2)}{(1-h^2) + (\text{tr}(\mathbf{A}) - (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{A}\mathbf{x} - \frac{1}{n}\mathbf{1}'\mathbf{A}\mathbf{1})\frac{h^2}{n-2}}\right)$$

Relationship matrix

Results



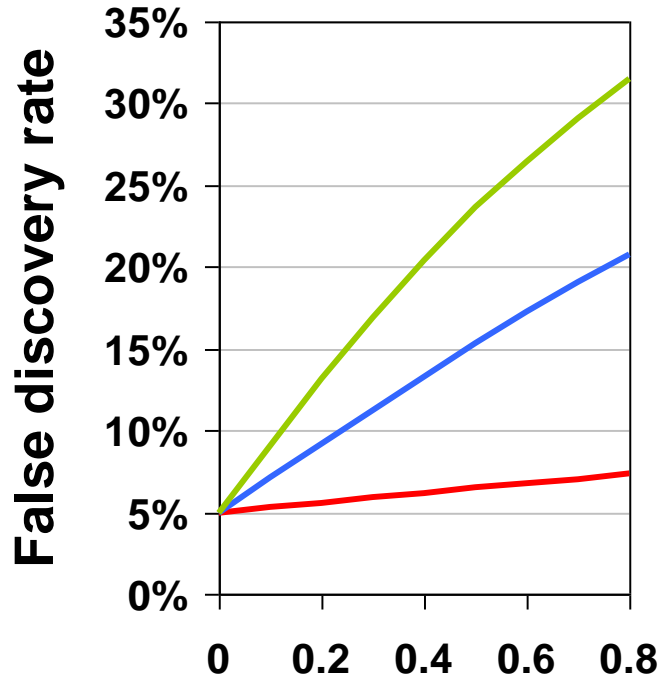
ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA

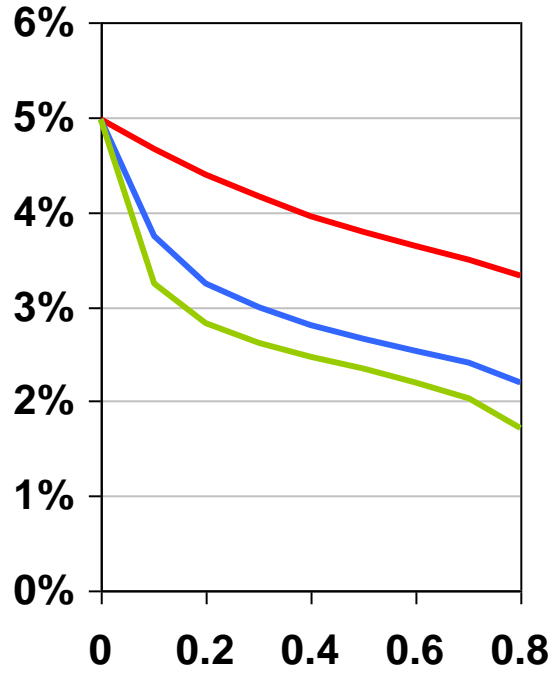
Results : Robustness

($n=600$, threshold supposed by user 5%)

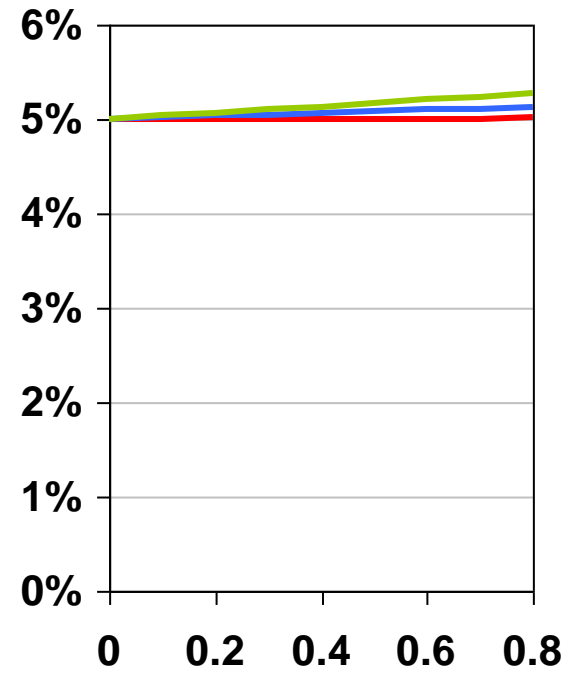
Regression



Grammar



QTD



Number of
progeny/sire

— 5

— 30

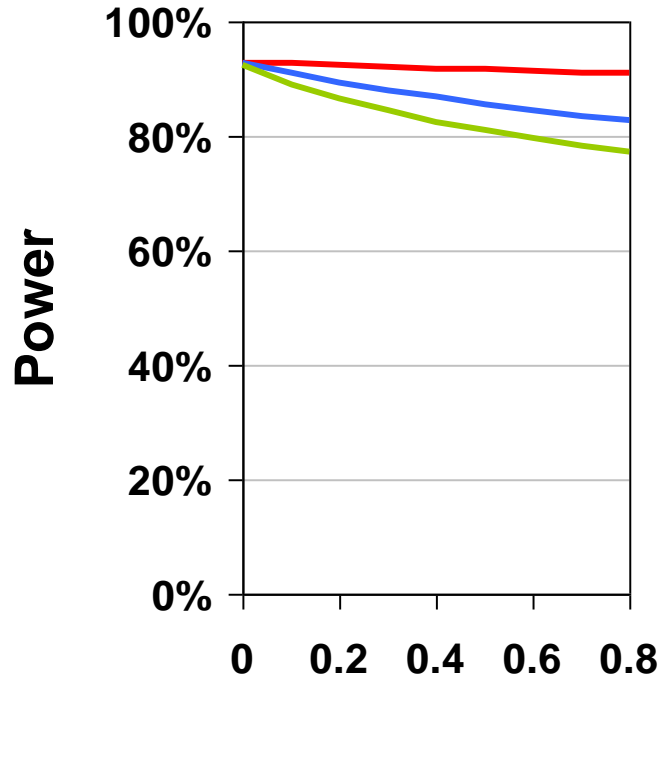
— 60

Heritability

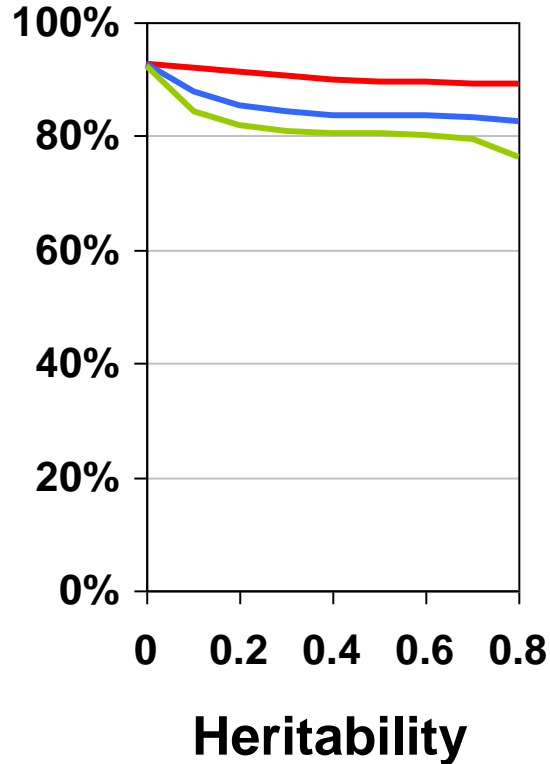
Results : Power

($n=600$, threshold 5%, SNP effect 0.20, Frequency 50%)

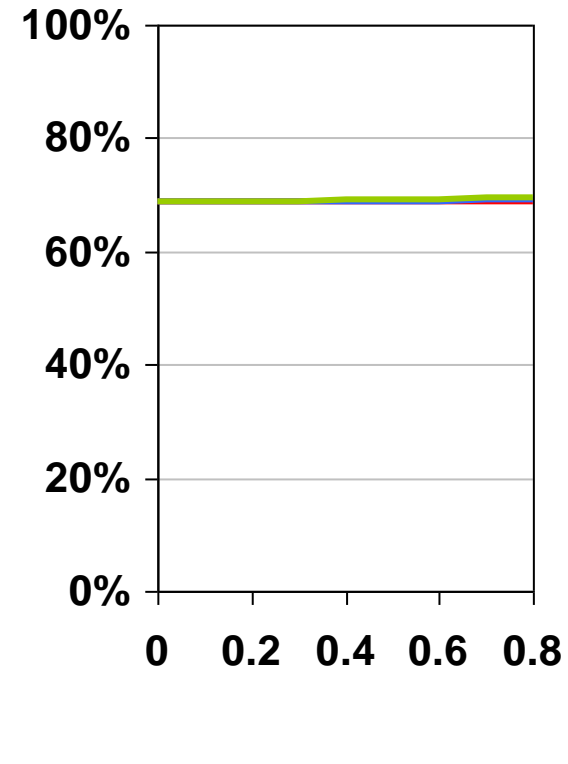
Regression



Grammar



QTDT



Number of
progeny/sire

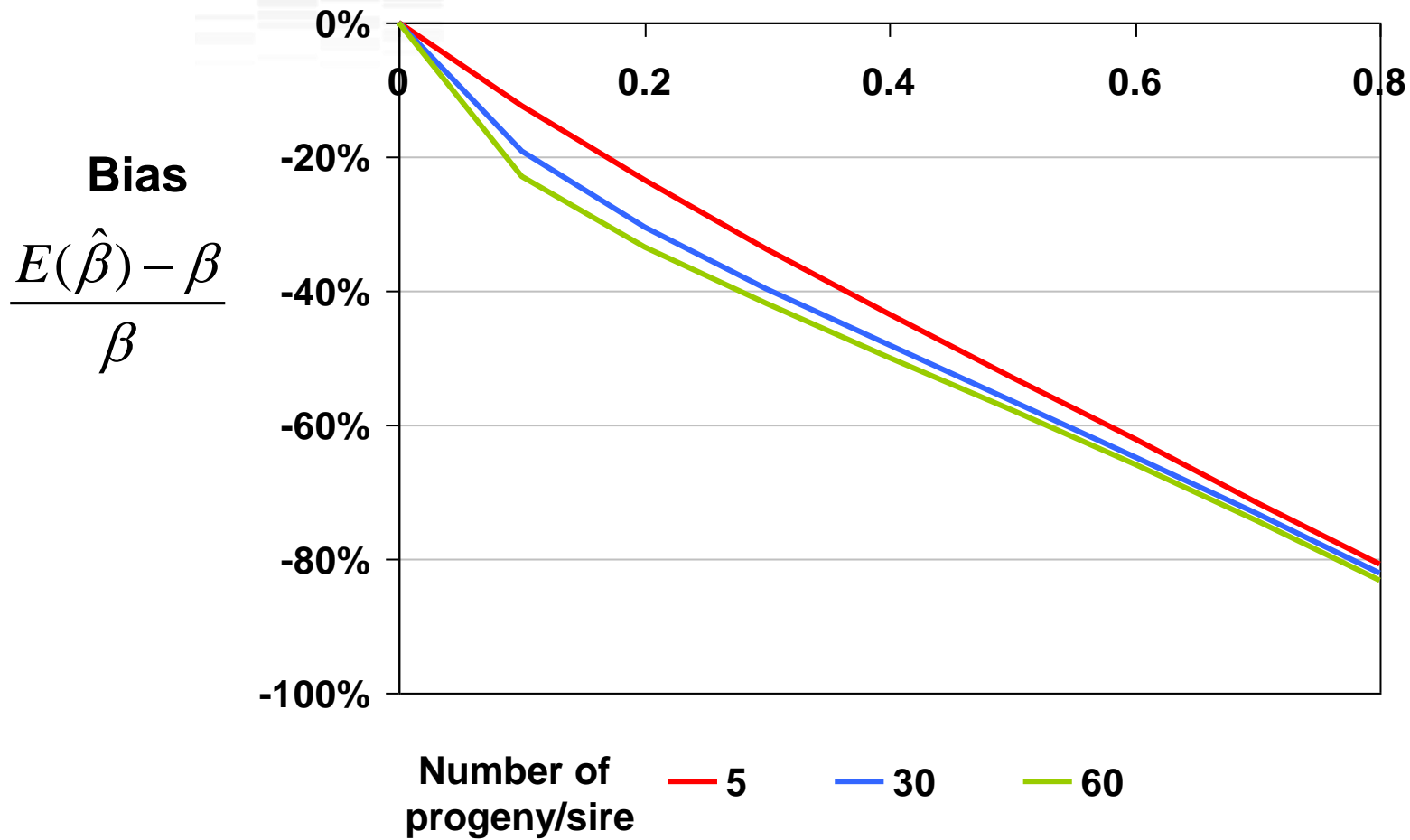
— 5

— 30

— 60

Results : Bias Grammar (n=600)

Heritability



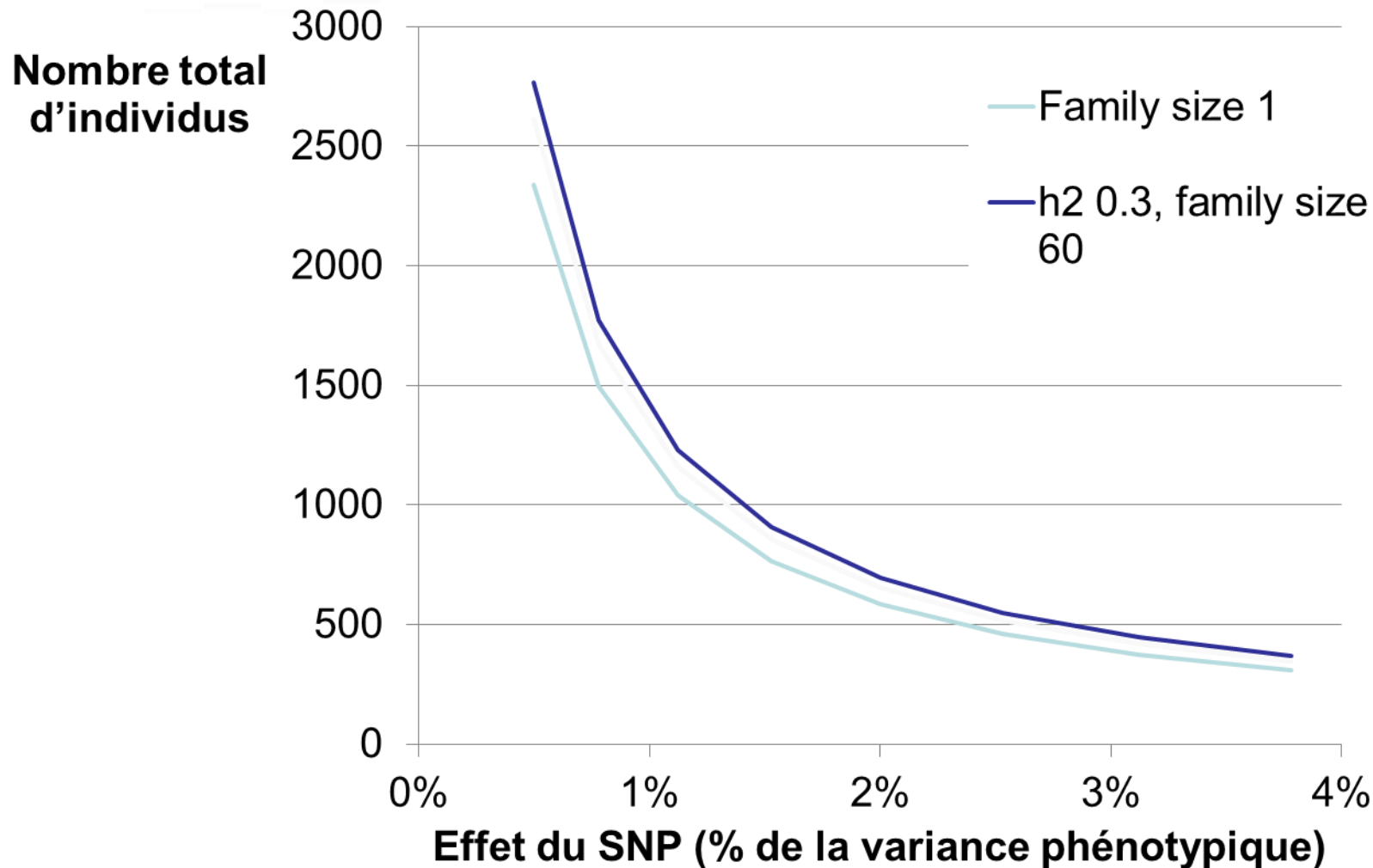
Quelques résultats amusants

- Pour la régression, l'accroissement du risque de première espèce n'entraîne pas une augmentation de la puissance mais une diminution de celle-ci
- C'est la variabilité des relations de parenté qui augmente le risque de première espèce et non la parenté elle-même (par exemple une seule grande famille n'augmente pas ce risque)
- GRAMMAR n'est pas sensible à la structure familiale
- Le biais de GRAMMAR est très important même si cela n'affecte pas la puissance car l'augmentation du biais est parallèle à la décroissance de la variance de la performance corrigée utilisée pour le test.
- A cause de l'utilisation de l'analyse de liaison le QTDT n'est pas affecté par les relations de parenté entre animaux même si la matrice de parenté n'est pas introduite dans le modèle

Comparaison des tests

- Cette comparaison est possible pour toute configuration de parenté, pas besoin de simulations
- L'utilisation de la régression "simple" donne de beaucoup de SNP significatifs sachant que beaucoup seront de faux positifs (ex: le test multiplie par 4 le nombre de SNP significatifs pour $h^2=0.50$ et des tailles de famille de 1/2 frères de 30..)
- La puissance de la **Regression** et **GRAMMAR** est affecté par la structure familiale dans le même ordre de grandeur mais modérément (perte de 7-9% of puissance pour $h^2=0.50$ et des tailles de famille de 1/2 frères de 30)
- **QTD** n'est pas affecté par la structure familiale mais a une puissance faible (-15 to 20% !)

Population size required to reach a 80% power with a 1% first type error



Population required to reach a 80% power with a 1% first type error, with the SNP effect explaining 2% of phenotypic variance.

