

Power Calculation | Calcul de Puissance

Horace Gninafon

June 29, 2023

Key points for this lecture | *Points clés du cours*

- ▶ Power is the ability of our experiment to detect statistically significant treatment effects, if they in fact exist
- ▶ Power matters: for practical reasons and for interpretation
- ▶ Increase power by strengthening intervention, reducing noise, and increasing sample size
- ▶ La puissance est la capacité de notre expérience à détecter des effets de traitement statistiquement significatifs, s'ils existent réellement
- ▶ La puissance est importante : pour des raisons pratiques et pour l'interprétation
- ▶ Augmenter la puissance en renforçant l'intervention, en réduisant le bruit et en augmentant la taille de l'échantillon

Review | *Révision*

Review | *Révision* (1)

- ▶ Estimand (the truth) vs estimate of the average treatment effect (ATE)
- ▶ Sampling distribution: distribution of estimates of the ATE from all possible randomizations
- ▶ La paramètre (la vérité) vs l'estimation ponctuelle (la devinette) d'ATE
- ▶ Distribution d'échantillonnage: La distribution des effets moyens du traitement estimés pour toutes les affectations de traitement possibles.

Review | Révision (2)

- ▶ Standard deviation of a distribution
- ▶ Standard error: standard deviation of the sampling distribution
- ▶ Un écart-type d'une distribution
- ▶ Standard error: L'écart-type d'une distribution d'échantillonnage

$$SD_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

Power | *La puissance*

- ▶ The ability of the experiment to detect statistically significant treatment effects, if the effect really exists
- ▶ Experiment's ability to avoid making a Type II error (incorrect failure to reject the null hypothesis of no effect).
- ▶ La capacité de notre expérience à détecter des effets de traitement statistiquement significatifs, s'ils existent réellement.
- ▶ La capacité de l'expérience à éviter de commettre une erreur de type II (incapacité à rejeter l'hypothèse nulle d'absence d'effet).

Power Analysis | *L'analyse de la puissance*

- ▶ Power analysis is something we do before we run a study.
- ▶ Goal: To discover whether our planned design has enough power to detect effects if they exist.
- ▶ L'analyse de puissance est quelque chose que nous faisons avant de mener une étude.
- ▶ Objectif : Déterminer si le design de l'étude a suffisamment de puissance pour détecter les effets s'ils existent.

Power Analysis | *L'analyse de la puissance*

- ▶ We usually state a hypothesis about the effect-size of a treatment and compare this against the null hypothesis of no effect.
- ▶ Nous formulons généralement une hypothèse sur l'ampleur de l'effet d'un traitement et la comparons à l'hypothèse nulle d'absence d'effet.

What are the three main inputs into statistical power? |

Quels sont les trois principaux éléments de la puissance statistique ?

- ▶ Sample size;
- ▶ Noisiness of the outcome variable (σ);
- ▶ Treatment-effect size.
- ▶ Taille de l'échantillon
- ▶ La dispersion de la variable de résultat (i.e. la variable dépendante) (σ)
- ▶ Taille de l'effet du traitement.

The Power Formula | *The Power Formula (1)*

$$Power = \Phi\left(\frac{|\tau|\sqrt{N}}{2\sigma}\right) - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \quad (1)$$

- ▶ Power is a number between 0 and 1; higher is better;
- ▶ Φ is the cumulative density function of the normal distribution: FIXED;
- ▶ τ is the effect size.
- ▶ La puissance est un nombre compris entre 0 et 1 ; plus elle est élevée, mieux c'est...
- ▶ Φ est la fonction de densité cumulative de la distribution normale: FIXE
- ▶ τ est la taille de l'effet.

The Power Formula | *The Power Formula (2)*

- ▶ N is the sample size
- ▶ σ is the standard deviation of the outcome
- ▶ α is the significance level: FIXED (by convention)
- ▶ N est la taille de l'échantillon
- ▶ σ est l'écart-type du résultat
- ▶ α est le niveau de signification: FIXE (par convention)

Three main inputs into statistical power 1: Sample size |

Trois éléments principaux de la puissance statistique 1 : La taille de l'échantillon

- ▶ More observations → more power
- ▶ Add observations!
- ▶ Problems?
- ▶ Plus d'observations → plus de puissance
- ▶ Ajoutez des observations !
- ▶ Problèmes éventuels ?

Three main inputs into statistical power 2: Noisiness of outcome measure | *Les trois principaux facteurs de la puissance statistique 2 : La dispersion de la variable dépendante*

- ▶ Less noise → more power
 - ▶ Reduce noise. How?
 - ▶ Blocking — conduct experiments among subjects that look more similar;
 - ▶ Collect baseline covariates — background information about experimental units;
 - ▶ Problems?
- ▶ Faible dispersion → plus de puissance
 - ▶ Comment réduire la dispersion ?
 - ▶ Stratifier - mener des expériences avec des sujets qui se ressemblent le plus.
 - ▶ Recueillir des covariables de base - informations de base sur les unités expérimentales;
 - ▶ Problèmes éventuels ?

Three Main Inputs into Statistical Power 3: Size of Treatment Effect | *Trois éléments principaux de la puissance statistique 3 : Taille de l'effet du traitement*

- ▶ Bigger effect → more power;
 - ▶ Boost dosage / avoid very weak treatments;
 - ▶ Problems?
- ▶ Plus d'effet → plus de puissance;
 - ▶ Augmenter le dosage / éviter les traitements très faibles;
 - ▶ Problèmes éventuels ?

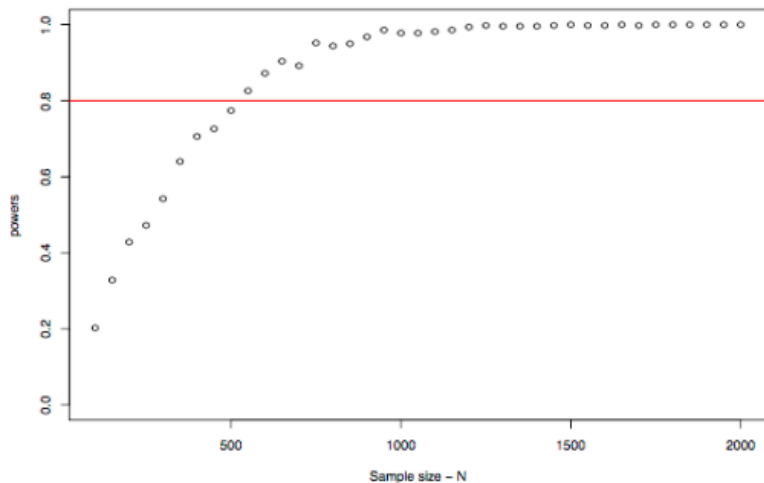
Power is the Art of Tweaking! | La puissance est l'art du réglage !

We tweak different parts of our design up front to make sure that our experiment has enough power to detect effects (assuming they exist)

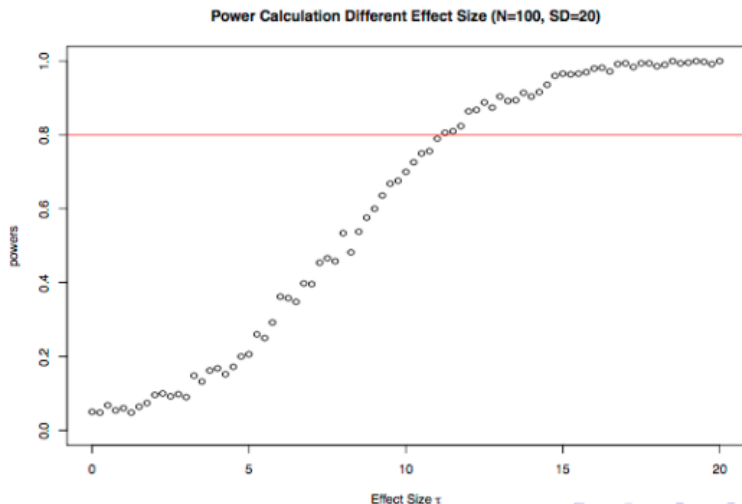
Nous modifions les différentes parties de notre conception dès le départ afin de nous assurer que notre expérience a suffisamment de puissance pour détecter les effets (en supposant qu'ils existent).

Tweak Sample Size: How Does Power Respond? | *Modifier la taille de l'échantillon : comment Power réagit-il ?*

Power Calculation Different Sample Size ($\tau = 5$, $SD = 20$)

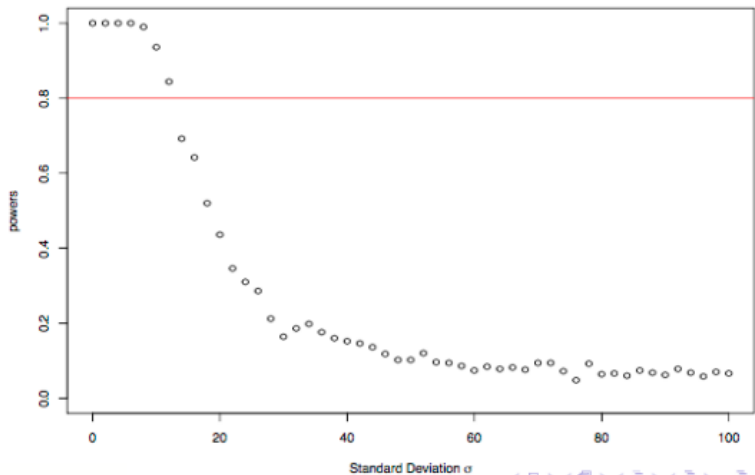


Tweak Effect Size: How Does Power Respond? | *Modifier la taille de l'effet: comment Power réagit-il ?*



Tweak SD of Outcome: How Does Power Respond? | *Modifier l'Ecart type de la variable dépendante: comment la puissance réagit-elle ?*

Power Calculation Different Noise Size (N=200, $\tau = 5$)



Your Turn! | *A vous de jouer ! (1)*

- ▶ Go to <http://egap.org/>
 - ▶ Tools -> Apps -> EGAP Tool: Power Calculator
 - ▶ Set Significance Level at Alpha equal to 0.05
 - ▶ Set Power Target at 0.8
 - ▶ Set Maximum Number of Subjects at 1,000.
- ▶ Aller sur <http://egap.org/>
 - ▶ Outils -> Applications -> Outil EGAP : Calculateur de puissance
 - ▶ Fixer le niveau de signification à Alpha égal à 0,05
 - ▶ Fixer l'objectif de puissance à 0,8
 - ▶ Fixer le nombre maximum de sujets à 1000.

Your Turn! | *A vous de jouer ! (2)*

- ▶ Fix Standard Deviation of Outcome Variable at 10.
How many subjects do I need if my Treatment Effect Size is 2 in order for my experiment to have 80% power? What about Treatment Effect Size 5? Treatment Effect Size 10?
- ▶ Fixez l'écart-type de la variable de résultat à 10.
De combien de sujets ai-je besoin si l'ampleur de l'effet de traitement est de 2 pour que mon expérience ait une puissance de 80 % ? Qu'en est-il si l'ampleur de l'effet de traitement est de 5 ou de 10?

Example | *Exemple*

- ▶ Caleb runs an experiment to see whether giving women cash reduce Intimate Partner Violence (IPV) their experience.
 - ▶ He finds no statistically significant difference between the groups.
 - ▶ What does this mean?
- ▶ Caleb a mené une expérience pour déterminer si le fait de donner de l'argent liquide aux femmes a réduit les violences qu'elles subissent.
 - ▶ Il ne constate pas de différence statistiquement significative entre les deux groupes.
 - ▶ Qu'est-ce que cela signifie ?

- ▶ With a study that does not have enough power, we can't tell whether a null finding (no statistically significant effect) means there is no effect or we just can't detect an effect that exists.
- ▶ Dans le cas d'une étude dont la puissance est insuffisante, il est impossible de dire si un résultat nul (pas d'effet statistiquement significatif) signifie qu'il n'y a pas d'effet ou si nous ne pouvons tout simplement pas détecter un effet existant.

Under-powered studies | *Les études de faible puissance*

- ▶ We don't know whether we should adopt this treatment because we don't know whether it's effective.
- ▶ Perhaps running this study is not a good use of our resources.
- ▶ Nous ne savons pas si nous devons adopter ce traitement car nous ne savons pas s'il est efficace.
- ▶ Peut-être que la réalisation de cette étude n'est pas une bonne utilisation de nos ressources.

An Alternative Perspective: Minimum
Detectable Effect | *Une autre perspective :*
L'effet minimal détectable

Difficulty with Power Analysis |

- ▶ Hardest part of power analysis is plugging in treatment effect : how can we possibly know before experiment has been run?
- ▶ La partie la plus difficile de l'analyse de puissance consiste à introduire l'effet du traitement - comment pouvons-nous le savoir avant que l'expérience n'ait été réalisée?

Ask two questions: | *Posez deux questions :*

1. For a give set of inputs, what's the smallest effect that my study would be able to detect?
 2. Would this effect-size be "satisfactory"?
 - ▶ Cost-effectiveness
 - ▶ Disciplinary rules of thumb (e.g. 0.2 SD effects in education research)
 - ▶ Other studies which had similar goals to yours
1. Pour un ensemble donné de données, quel est le plus petit effet que mon étude pourrait détecter ?
 2. Cette taille d'effet serait-elle "satisfaisante" ?
 - ▶ Coût-efficacité
 - ▶ Règles empiriques disciplinaires (par exemple, effets de 0,2 SD dans la recherche sur l'éducation)
 - ▶ Autres études ayant des objectifs similaires aux vôtres

Minimum Detectable Effect | *L'effet minimal détectable*

- ▶ After we answer these two questions, we design the study that will allow us to detect an effect with that size.



Special Case: Cluster-Randomized Designs |
*Cas particulier : Designs d'une randomisation
par grappe*

Special Case: Clustered-Randomized Designs | *Cas particulier : Designs d'une randomisation par grappe*

- ▶ We often randomize treatment at the level of groups, but only have the ability to sample a few people within those groups.
 - ▶ Also used if intervention has to function at the cluster level.
 - ▶ Examples?
- ▶ Nous randomisons le traitement au niveau des groupes, mais nous n'avons la possibilité d'échantillonner que quelques personnes au sein de ces groupes.
 - ▶ Utiliser aussi cela si l'intervention doit s'effectuer au niveau de la grappe.
 - ▶ Exemples ?

Special issues for power | *Challenges particuliers liés à la puissance*

- ▶ Number of individuals sampled per cluster
- ▶ Intra-cluster correlation: how similar are people that are in the same cluster
- ▶ Nombre d'individus échantillonnés par grappe
- ▶ Correlation intra-grappe: le degré de similarité des individus qui sont dans la même grappe

Example | *Exemple*

- ▶ 4,000 bureaucrats, divided into 200 municipalities of 20 workers each; 100 municipalities in treatment and 100 municipalities in control.
 - ▶ When the intracluster correlation is 0, bureaucrats within the same municipalities are not similar;
 - ▶ It's like assigning 4,000 bureaucrats to treatment or control!
- ▶ 4000 bureaucrats, répartis en 200 communes de 20 bureaucrates par communes ; 100 communes dans le groupe de traitement et 100 communes dans le groupe de contrôle.
 - ▶ Lorsque la corrélation intracluster est égale à 0, les bureaucrates d'une même municipalités ne sont pas semblables.
 - ▶ Cela revient à assigner 4000 bureaucrates qui seront répartis dans le groupe de traitement et dans le groupe de contrôle!

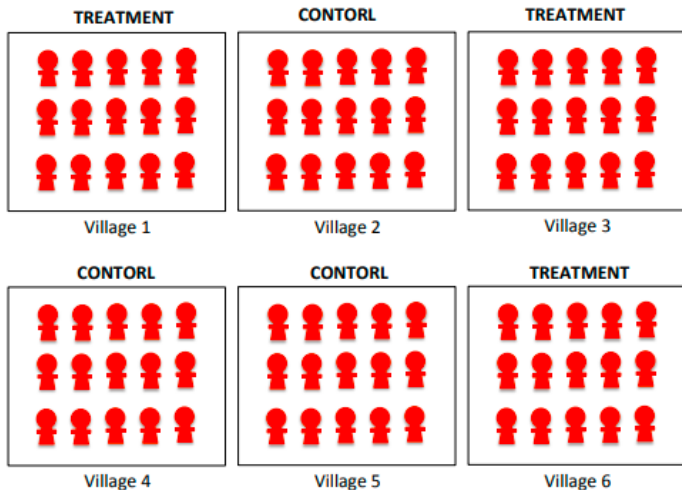
Example | *Exemple*

- ▶ When the intracluster correlation is 1, everyone within a municipality acts the same, and so you effectively have 200 independent observations
- ▶ Implications for power?
- ▶ Lorsque la corrélation intra grappe est de 1, tous les individus d'une même municipalité agissent de la même manière et vous disposez donc de 200 observations indépendantes.
- ▶ Implications pour la puissance ?

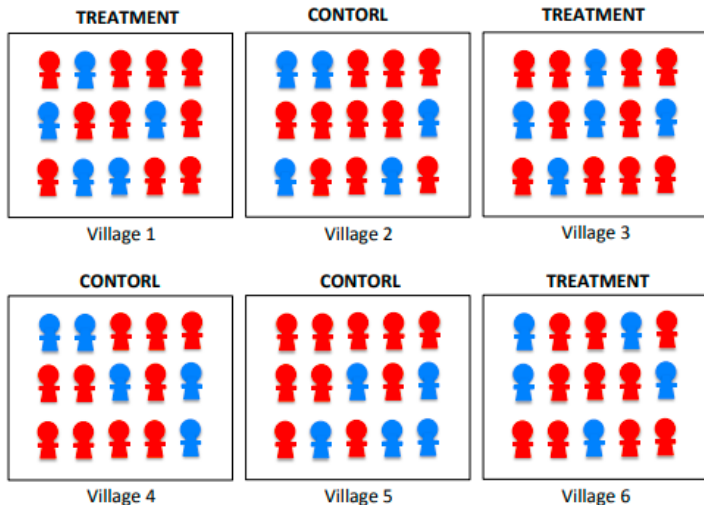
Special Case: Clustered-Randomized Designs | Cas particulier : Plans d'échantillonnage aléatoire en grappes



Special Case: Clustered-Randomized Designs | Cas particulier : Plans d'échantillonnage aléatoire en grappes

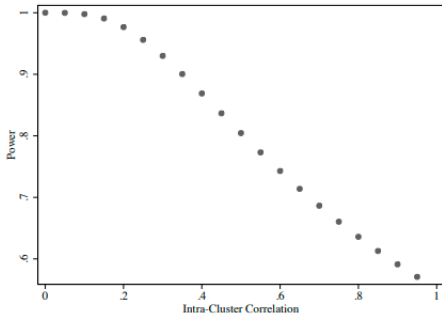


Special Case: Clustered-Randomized Designs | Cas particulier : Plans d'échantillonnage aléatoire en grappes



Tweak Intra-Cluster Correlation: How Does Power Respond? | *Modifier la corrélation intra grappe : Comment la puissance réagit-elle ? (1)*

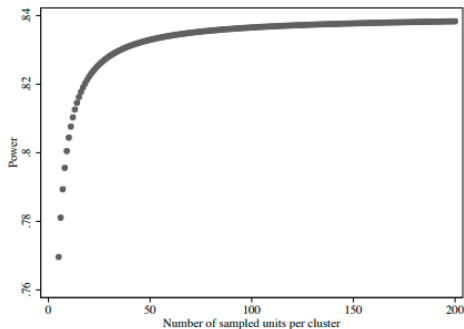
- ▶ Number of clusters = 140;
10 sampled per cluster
- ▶ Nombre de grappes = 140 ;
10 échantillonnées par grappe



Tweak Number of Units Per Cluster: How Does Power Respond? | *Modifier le nombre d'unités par grappe: Comment la puissance réagit-elle ? (2)*

- ▶ Another choice we have to make in cluster designs is how many units within clusters to sample
- ▶ Surely we want to sample as many as possible, right?
- ▶ Hmm..
- ▶ Un autre choix à faire dans les conceptions par grappe est celui du nombre d'unités à échantillonner à l'intérieur des grappes
- ▶ Nous voulons certainement en échantillonner le plus possible, n'est-ce pas ?
- ▶ Hmm...

Tweak Number of Units Per Cluster: How Does Power Respond? | Modifier le nombre d'unités par grappe : Comment la puissance réagit-elle ?



Golden Rule of Cluster-Randomized Designs | *Règle d'or des plans randomisés en grappes*

- ▶ Unless intra-cluster correlation is very small, it's always better to add more clusters than to sample more people within the clusters
- ▶ A moins que la corrélation intra-grappe ne soit très faible, il est toujours préférable d'ajouter des grappes plutôt que d'échantillonner davantage de personnes à l'intérieur des grappes.

Your Turn! | *À vous de jouer !*

- ▶ Go to <http://egap.org/>
- ▶ Tools > Apps > EGAP
Tool: Power Calculator
- ▶ Click box which says
“Clustered Design?”
- ▶ Set Significance Level at
Alpha = 0.05
- ▶ Set Treatment Effect Size
at 5
- ▶ Standard Deviation of
Outcome Variable at 10
- ▶ Set Power Target at 0.8
- ▶ Set Maximum Number of
Subjects at 2000
- ▶ Aller sur <http://egap.org/>
- ▶ Outils > Applications >
Outil EGAP : Calculateur
de puissance
- ▶ Cliquez sur la case qui dit
“Clustered Design ?”
- ▶ Fixez le niveau de
signification à Alpha = 0,05
- ▶ Fixez la taille de l'effet de
traitement à 5
- ▶ Écart-type de la variable de
résultat à 10
- ▶ Fixer l'objectif de puissance
à 0,8
- ▶ Fixer le nombre maximum
de sujets à 2000

Your Turn! | *À vous de jouer !* (1)

1. Fix Number of Clusters per Arm at 40. How many subjects do I need if my Intra-cluster Correlation is 0.6 in order for my experiment to have 80% power? What about Intra-cluster Correlation of 0.4? 0.1? 0?
1. Fixer le nombre de grappes par traitement à 40. De combien de sujets ai-je besoin si ma corrélation intra-groupe est de 0,6 pour que mon expérience ait une puissance de 80 % ? Qu'en est-il d'une corrélation intra-groupe de 0,4 ? 0.1 ? 0 ?

Your Turn! | À vous de jouer ! (2)

2. Fix Intra-cluster Correlation at 0.5. How many subjects do I need if the Number of Clusters per Arm is 100 in order for my experiment to have 80% power? What if the Number of Clusters per Arm is 50? 35? 20?

2. Fixer la corrélation intra-grappe à 0,5. De combien de sujets ai-je besoin si le nombre de grappes par bras est de 100 pour que mon expérience ait une puissance de 80 % ? Quel serait le nombre d'individus par grappe à échantillonner si le nombre de grappe par traitement est de 50 ? 35 ? 20 ?

Recap: What Have you Learned? | *Récap : Qu'avez-vous appris ? (1)*

- ▶ Power is the ability of our experiment to detect statistically significant treatment effects, if they in fact exist
- ▶ Power matters: for practical reasons and for interpretation
- ▶ Increase power by strengthening intervention, reducing noise, and increasing sample size
- ▶ La puissance est la capacité de notre expérience à détecter des effets de traitement statistiquement significatifs, s'ils existent réellement
- ▶ La puissance est importante : pour des raisons pratiques et pour l'interprétation
- ▶ Augmenter la puissance en renforçant l'intervention, en réduisant le bruit et en augmentant la taille de l'échantillon

Recap: What Have you Learned? | *Récap : Qu'avez-vous appris ? (2)*

- ▶ In cluster-randomized designs, almost always better to add more clusters rather than interview more people within clusters
- ▶ Always run a power analysis before committing to a final design
- ▶ But beware that it involves some guesswork; be skeptical and vary assumptions
- ▶ Dans les expériences randomisées par grappe, il est presque toujours préférable d'ajouter des grappes plutôt que d'interroger davantage de personnes au sein des grappes.
- ▶ Effectuez toujours une analyse de puissance avant de vous engager dans une conception finale.
- ▶ Mais attention, cela implique des suppositions ; soyez sceptiques et variez les hypothèses.

- ▶ Note: Most of these slides are not original. Material is borrowed from several sources:
 - ▶ Gareth Nellis, Learning Days slides (on power analysis)
 - ▶ Cyrus Samii, NYU slides (on minimum detectable effects)
 - ▶ Tara Slough, Columbia slides (graphs on the sensitivity of effects)
- ▶ Note: La plupart de ces slides ne sont pas originales. Le matériel est extrait de plusieurs sources:
 - ▶ Gareth Nellis, Slides des Journées d'Apprentissages de EGAP (Analyse de puissance)
 - ▶ Cyrus Samii, NYU diapositives (sur les effets minimaux détectables)
 - ▶ Tara Slough, Columbia diapositives (graphiques sur la sensibilité des effets)