

Poly de révision

Nombre de sujets nécessaires, puissance statistique et annales résolues

Table des matières

Poly de révision	3
Nombre de sujets nécessaires, puissance statistique et annales résolues	3
Comment utiliser ce poly	4
1. Carte rapide des questions couvertes	4
2. Réflexe n°1 : identifier le bon type de question	5
2.1. Vrai NSN prospectif	5
2.2. Puissance a posteriori	5
2.3. Critique méthodologique autour du NSN	5
3. Les briques de base à connaître	6
3.1. Les paramètres indispensables	6
3.2. Les valeurs de z à mémoriser	6
3.3. Formule indispensable pour remonter d'un IC95% vers un SE	7
3.3.1. Cas général : estimateur exprimé sur une échelle linéaire	7
3.3.2. Cas très fréquent des RR, OR et HR	7
3.3.3. Comment on s'en sert à l'examen	8
3.4. Les trois grandes formules	11
3.4.1. Deux moyennes	11
3.4.2. Deux proportions	11
3.4.3. Essais de survie / log-rank	11
4. Les ajustements qu'il ne faut pas oublier	11
4.1. Dropout et perdus de vue	11
4.2. Essai en cluster	12
4.3. Multiplicité	12
4.4. Co-primary endpoints	12
4.5. Surdispersion	12
4.6. Non-infériorité	13
5. Comment écrire une bonne réponse d'examen	13
5.1. Trame universelle	13
5.2. Si le calcul exact est impossible	13
5.3. Si l'étude est observationnelle	13
6. Annales résolues avec calculs détaillés	13
6.1. 2011 EPI	14
Article et contexte	14
Ce qu'il faut aller chercher dans l'article	14
Ce que l'examineur veut vérifier	14
Comment passer des chiffres de l'article au calcul	15
Ce qu'il faut comprendre	19

6.2. 2011 RC	20
Article et contexte	20
Ce qu'il faut aller chercher dans l'article	20
Ce que l'examineur veut vérifier	21
Résolution guidée	21
6.3. 2012 EPI	21
Article et contexte	21
Ce qu'il faut aller chercher dans l'article	22
Ce que l'examineur veut vérifier	22
Ce qu'il faut comprendre	23
6.4. 2012 RC	24
Article et contexte	24
Ce qu'il faut aller chercher dans l'article	24
Ce que l'examineur veut vérifier	24
Interprétation	26
6.5. 2013 EPI	26
Article et contexte	26
Ce qu'il faut aller chercher dans l'article	26
Ce que l'examineur veut vérifier	27
Ce qu'il faut comprendre	27
6.6. 2014 EPI	28
Article et contexte	28
Ce qu'il faut aller chercher dans l'article	28
Ce que l'examineur veut vérifier	28
6.7. 2016 EPI	30
Article et contexte	30
Ce qu'il faut aller chercher dans l'article	30
Ce que l'examineur veut vérifier	30
Ce qu'il faut écrire	30
6.8. 2019 EPI	31
Article et contexte	31
Ce qu'il faut aller chercher dans l'article	31
Ce que l'examineur veut vérifier	32
Ce qu'il faut comprendre	32
6.9. 2020 RC	32
Article et contexte	32
Ce qu'il faut aller chercher dans l'article	33
Ce que l'examineur veut vérifier	33
Pourquoi ce calcul naïf ne suffit pas	33
Résolution guidée	34
6.10. 2021 RC	34
Article et contexte	34
Ce qu'il faut aller chercher dans l'article	35
Ce que l'examineur veut vérifier	35
Point pédagogique important	36
6.11. 2022 RC	36
Article et contexte	36
Ce qu'il faut aller chercher dans l'article	36
Ce que l'examineur veut vérifier	37
Ce qu'il faut écrire	37
Pourquoi le binomial négatif est central	37

6.12. 2023 RC	37
Article et contexte	37
Ce qu'il faut aller chercher dans l'article	38
Ce que l'examineur veut vérifier	38
Étape 2 : revenir au nombre total de sujets dans l'essai	39
Multiplicité	39
6.13. 2024 RC	39
Article et contexte	39
Ce qu'il faut aller chercher dans l'article	40
Ce que le corrigé de l'annale fait réellement	40
Le message méthodologique à faire apparaître	41
6.14. 2025 RC	41
Article et contexte	41
Ce qu'il faut aller chercher dans l'article	42
Ce que l'examineur veut vérifier	42
Conclusion chiffrée	42
7. Réponses prêtes à l'emploi par grand type de question	42
7.1. Si l'on vous demande un vrai NSN	42
7.2. Si l'on vous demande une puissance a posteriori	42
7.3. Si l'article ne donne pas assez d'éléments	43
7.4. Si l'étude est une grande cohorte	43
8. Checklist finale avant de rendre la copie	43
9. Conclusion	43

Poly de révision

Nombre de sujets nécessaires, puissance statistique et annales résolues

Document de travail orienté examen : méthodes, pièges, calculs détaillés et réponses exploitables directement dans les annales.

Ce document doit permettre de répondre entièrement aux questions d'annales portant sur le nombre de sujets nécessaires, la puissance statistique et les cas méthodologiques voisins. Il progresse des principes généraux vers les annales résolues, avec des calculs détaillés dès que les données de l'article le permettent.

Ce poly a un seul objectif : **permettre de répondre entièrement aux questions d'annales portant sur le calcul du nombre de sujets nécessaires, la puissance, et les cas voisins.**

Il suit une progression volontairement simple :

1. reconnaître immédiatement le type de question ;
2. mobiliser la bonne formule ou le bon raisonnement ;
3. écrire une réponse d'examen complète, même si le calcul exact n'est pas reconstructible ;
4. retrouver année par année ce qu'il fallait faire.

didactique progressif détaillé orienté annales HTML A4 + PDF

! Idée directrice

Le mot **NSN** ne signifie pas toujours la même chose selon le design.

Dans un essai randomisé, on peut souvent faire un vrai calcul prospectif.

Dans une cohorte ou une étude transversale, il faut souvent raisonner plutôt en **nombre d'événements**.

ments, en précision, ou en interprétation des intervalles de confiance.

Comment utiliser ce poly

Lecture conseillée :

1. lire les sections 1 à 4 pour acquérir la mécanique générale ;
2. passer ensuite directement à la section “Annales résolues” ;
3. revenir à la fin sur les réponses-type et la checklist.

Si vous êtes en révision rapide, les éléments à mémoriser en priorité sont :

- les paramètres indispensables d’un NSN ;
- les trois grandes formules de base ;
- les ajustements majeurs : dropout, cluster, multiplicité, surdispersion, co-primary endpoints, non-infériorité ;
- la façon de répondre quand le calcul exact est impossible.

1. Carte rapide des questions couvertes

Année	Épreuve	Nature de la question	Mouvement attendu
2011	EPI	puissance a posteriori	discuter l’effet brut vs l’effet net et la précision
2011	RC	ITT, perdus de vue, inflation du NSN	distinguer arrêt de traitement et vraie perte de suivi
2012	EPI	faible nombre d’événements, puissance	montrer que peu d’événements = faible puissance
2012	RC	retrouver le NSN	raisonner “event-driven” à partir du nombre d’événements
2013	EPI	pas de calcul a priori, puissance rudimentaire	remplacer le NSN par un raisonnement de précision
2014	EPI	même situation, très grande cohorte	conclure à une puissance très élevée, limite = biais
2016	EPI	comment discuter la puissance d’une cohorte	raisonner par événements, sous-groupes, IC
2019	EPI	cohorte immense	dire que la puissance n’est pas le vrai problème
2020	RC	co-primary endpoints	partir d’un calcul simple puis ajouter la corrélation

Année	Épreuve	Nature de la question	Mouvement attendu
2021	RC	essai individuel au lieu de cluster	faire un NSN standard à deux proportions
2022	RC	modèle binomial négatif	énumérer les paramètres et expliquer ce qui manque
2023	RC	essai individuel, trois bras, majoration	calcul continu simple puis inflation
2024	RC	non-infériorité	retrouver le calcul simplifié du corrigé
2025	RC	log-rank, hypothèses explicites	reprendre directement les hypothèses de l'article

2. Réflexe n°1 : identifier le bon type de question

2.1. Vrai NSN prospectif

Vous êtes dans ce cas si l'énoncé demande :

- “combien de sujets faut-il inclure ?”
- “retrouvez le nombre de sujets nécessaires”
- “si l'essai avait randomisé des sujets et non des clusters”
- “essai de non-infériorité”
- “test du log-rank”

Le bon réflexe est alors :

- identifier le type de critère principal ;
- récupérer α , β , l'effet attendu, et la variabilité ;
- appliquer la formule ou la logique pertinente ;
- corriger pour les éléments de design.

2.2. Puissance a posteriori

Vous êtes dans ce cas si l'étude est déjà réalisée et que la question dit :

- “calculez une puissance a posteriori”
- “le faible nombre d'événements a limité la puissance”
- “comment discuter la puissance dans une telle étude ?”

Le bon réflexe est alors :

- dire d'abord que la puissance observée est un raisonnement secondaire ;
- utiliser plutôt l'estimation d'effet et l'IC95% ;
- si on vous force à chiffrer, proposer un calcul grossier et l'interpréter avec prudence.

2.3. Critique méthodologique autour du NSN

Vous êtes dans ce cas si l'énoncé mêle NSN et :

- ITT ;
- perdus de vue ;

- cluster ;
- co-primary endpoints ;
- multiplicité ;
- surdispersion.

Le bon réflexe est alors :

- ne pas foncer sur une formule ;
- commencer par expliquer **pourquoi** le calcul simple est insuffisant ;
- expliciter l'ajustement qui manque.

3. Les briques de base à connaître

3.1. Les paramètres indispensables

Paramètre	Symbole	Ce qu'il faut dire
risque de première espèce	α	probabilité de faux positif acceptée
risque de deuxième espèce	beta	probabilité de faux négatif acceptée
puissance	$1 - \beta$	probabilité de détecter l'effet s'il existe
effet minimal pertinent	delta, HR, $p_1 - p_2$	effet qu'on ne veut pas manquer
variabilité / risque de base	sigma, p , taux d'événements	composante indispensable du calcul
allocation	r, p	1 :1, 2 :1, trois bras, cluster
suivi / recrutement	durée	crucial en survie et en event-driven
pertes prévues	dropout	inflation éventuelle de l'effectif
design complexe	ICC, rho, surdispersion	cluster, co-primary, comptages

i Phrase à écrire presque à chaque fois

Pour calculer un NSN, il faut au minimum un α , une puissance, une taille d'effet cliniquement pertinente, et une mesure de la variabilité ou du risque de base. Ensuite on ajuste si nécessaire pour les pertes, le cluster, la multiplicité ou d'autres contraintes du design.

3.2. Les valeurs de z à mémoriser

Hypothèse	Valeur utile
$\alpha = 0,05$ bilatéral	1,96
$\alpha = 0,05$ unilatéral	1,645
$\alpha = 0,025$ unilatéral	1,96
puissance 80%	0,84
puissance 85%	1,04 environ
puissance 90%	1,28

3.3. Formule indispensable pour remonter d'un IC95% vers un SE

Cette formule est capitale dans les **questions de puissance a posteriori**, parce que les articles donnent souvent une estimation et son IC95%, mais pas directement l'écart-type standard (SE).

3.3.1. Cas général : estimateur exprimé sur une échelle linéaire

Si l'intervalle de confiance est écrit sous la forme :

$$IC95\% = \hat{\theta} \pm 1,96 \times SE$$

alors :

$$SE = \frac{\text{borne sup} - \text{borne inf}}{2 \times 1,96} = \frac{\text{largeur de l'IC}}{3,92}$$

Autrement dit :

- vous prenez la **largeur totale** de l'intervalle ;
- vous la divisez par 3,92 ;
- vous obtenez un SE approché.

Exemple simple : si une réduction relative est estimée à 10% avec IC95% -4% à 24%, alors

$$SE \approx \frac{0,24 - (-0,04)}{3,92} = \frac{0,28}{3,92} \approx 0,071$$

3.3.2. Cas très fréquent des RR, OR et HR

Pour les risques relatifs, odds ratios et hazard ratios, il faut presque toujours travailler sur l'échelle **logarithmique**, car c'est sur cette échelle que l'IC est approximativement symétrique.

Dans tout ce poly, \log désigne le **logarithme népérien**, c'est-à-dire \ln .

Si l'article donne :

$$RR \text{ avec } IC95\%[L; U]$$

alors on utilise :

$$SE(\log RR) \approx \frac{\log(U) - \log(L)}{3,92}$$

et de même :

$$SE(\log OR) \approx \frac{\log(U) - \log(L)}{3,92}$$

$$SE(\log HR) \approx \frac{\log(U) - \log(L)}{3,92}$$

Exemple : si RR = 0,72 avec IC95% 0,63 à 0,81,

$$SE(\log RR) \approx \frac{\log(0,81) - \log(0,63)}{3,92} \approx 0,064$$

avec :

$$\log(0,81) \approx -0,21, \quad \log(0,63) \approx -0,46$$

$$\log(0,81) - \log(0,63) \approx 0,25$$

$$\frac{0,25}{3,92} \approx 0,064$$

Table utile de $\ln(x)$ sans calculatrice :

x	$\ln(x)$
0,1	-2,303
0,2	-1,609
0,3	-1,204
0,4	-0,916
0,5	-0,693
0,6	-0,511
0,7	-0,357
0,8	-0,223
0,9	-0,105
1,0	0,000
1,1	0,095
1,2	0,182
1,3	0,262
1,4	0,336
1,5	0,405
1,6	0,470
1,7	0,531
1,8	0,588
1,9	0,642
2,0	0,693

3.3.3. Comment on s'en sert à l'examen

Le chemin de calcul est presque toujours le même :

1. repérer dans l'article l'**estimateur** et son IC95%;
2. choisir la bonne échelle :
 - échelle brute si l'effet est donné comme une différence ou une réduction exprimée directement;
 - échelle logarithmique pour un RR, OR ou HR;
3. calculer le SE avec la formule largeur / 3,92;
4. calculer un score de Wald grossier :

$$z \approx \frac{\hat{\theta}}{SE}$$

ou, pour un RR / OR / HR :

$$z \approx \frac{|\log(\widehat{RR})|}{SE(\log RR)}$$

5. conclure qualitativement :

- z très supérieur à 1,96 : puissance observée très élevée ;
- z proche de 1,96 : puissance limitée ou modérée ;
- z inférieur à 1,96 : résultat imprécis, puissance faible.

Pour un chiffrage grossier de puissance observée, on peut utiliser :

$$\text{puissance} \approx \Phi(z - 1,96)$$

en bilatéral au seuil 5%, avec toutes les précautions d'usage.

Ici, Φ désigne la **fonction de répartition de la loi normale centrée réduite** :

$$\Phi(a) = P(Z \leq a) \quad \text{avec } Z \sim \mathcal{N}(0, 1)$$

Concrètement, pour passer de $z - 1,96$ à une puissance :

- soit vous utilisez une **table de la loi normale** ;
- soit vous utilisez des **valeurs repères** ;
- soit vous raisonnez par **ordre de grandeur**.

Table de $\Phi(x)$ de -2 à 2 par pas de $0,25$:

x	$\Phi(x)$
-2	0,0228
-1,75	0,0401
-1,5	0,0668
-1,25	0,1056
-1	0,1587
-0,75	0,2266
-0,5	0,3085
-0,25	0,4013
0	0,5000
0,25	0,5987
0,5	0,6915
0,75	0,7734
1	0,8413
1,25	0,8944
1,5	0,9332
1,75	0,9599
2	0,9772

Règle de symétrie très utile :

$$\Phi(-x) = 1 - \Phi(x)$$

Par exemple :

$$\Phi(-0,5) \approx 1 - 0,69 = 0,31$$

$$\Phi(-0,75) \approx 1 - 0,77 = 0,23$$

$$\Phi(-1) \approx 1 - 0,84 = 0,16$$

i Piège classique

Dans cette partie du poly, on n'utilise **pas** directement $\Phi(z)$.

On utilise :

$$\text{puissance} \approx \Phi(z - 1,96)$$

Donc :

- si $z = 1,96$, alors $z - 1,96 = 0$ et la puissance est autour de 50%;
- si $z = 2,96$, alors $z - 1,96 = 1$ et la puissance est autour de 84%;
- si $z = 3,96$, alors $z - 1,96 = 2$ et la puissance est autour de 98%.

Tableau mental très pratique :

z observé	$z - 1,96$	puissance approximative
1,0	-0,96	0,17
1,5	-0,46	0,32
1,96	0	0,50
2,5	0,54	0,70
2,96	1,00	0,84
3,5	1,54	0,94
3,96	2,00	0,98
5,0	3,04	$\approx 1,00$

Lecture pratique :

- si $z - 1,96$ est proche de 0, la puissance est autour de 50%;
- si $z - 1,96$ vaut environ 1, la puissance est déjà élevée;
- si $z - 1,96$ vaut 2 ou plus, la puissance est très élevée;
- si $z - 1,96$ vaut 3 ou plus, elle est pratiquement égale à 1.

⚠ Important

Cette astuce sert surtout à produire un **ordre de grandeur** quand l'examen impose une discussion de puissance a posteriori. En pratique, l'IC95% est souvent plus informatif que la puissance observée elle-même.

3.4. Les trois grandes formules

3.4.1. Deux moyennes

$$n_{\text{par bras}} = \frac{2(z_{1-\alpha/2} + z_{1-\beta})^2 \sigma^2}{\delta^2}$$

À utiliser quand le critère principal est continu et qu'on compare deux groupes indépendants.

3.4.2. Deux proportions

$$n_{\text{par bras}} = \frac{\left[z_{1-\alpha/2} \sqrt{2\bar{p}(1-\bar{p})} + z_{1-\beta} \sqrt{p_1(1-p_1) + p_2(1-p_2)} \right]^2}{(p_1 - p_2)^2}$$

avec $\bar{p} = (p_1 + p_2)/2$.

À utiliser quand le critère principal est binaire.

3.4.3. Essais de survie / log-rank

Ce qu'on dimensionne souvent d'abord, c'est le **nombre d'événements requis**.

Forme de base de Schoenfeld :

$$D = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{p(1-p) [\log(HR)]^2}$$

où :

- D = nombre d'événements nécessaires ;
- p = proportion allouée à un groupe, souvent 0,5 ;
- HR = hazard ratio visé.

Ensuite seulement on convertit ce besoin en un **nombre de sujets** à partir :

- du risque attendu ;
- de la durée de recrutement ;
- de la durée de suivi ;
- des pertes de vue.

4. Les ajustements qu'il ne faut pas oublier

4.1. Dropout et perdus de vue

Si l'on prévoit un taux de perte q :

$$n_{\text{ajusté}} = \frac{n_{\text{calculé}}}{1 - q}$$

Exemple :

- si $n = 100$ par bras et $q = 15\%$, alors $100/0,85 = 117,6$, donc 118 par bras.

⚠ Piège majeur

Un **arrêt du traitement** n'est pas forcément une **perte de suivi**.

En ITT, un patient qui arrête le traitement reste analysable si l'on continue à recueillir le critère principal.

4.2. Essai en cluster

Effet de design :

$$DE = 1 + (m - 1) \times ICC$$

Puis :

$$n_{\text{cluster-ajusté}} = n_{\text{individuel}} \times DE$$

où :

- m = taille moyenne des clusters;
- ICC = corrélation intra-cluster.

4.3. Multiplicité

Correction la plus simple :

$$\alpha_{\text{ajusté}} = \frac{\alpha}{m}$$

si l'on fait m comparaisons indépendantes et que l'on veut un contrôle global simple de l'erreur de type I.

4.4. Co-primary endpoints

Quand le critère principal comporte plusieurs composantes qui doivent toutes être positives :

- le NSN augmente;
- la corrélation entre critères devient importante;
- une simple formule “deux moyennes” ou “deux proportions” n'est plus suffisante.

4.5. Surdispersion

Pour les comptages, si la variance dépasse la moyenne :

- un modèle de Poisson simple devient trop optimiste;
- il faut un modèle binomial négatif ou une correction équivalente;
- le NSN est plus grand que dans un calcul naïf.

4.6. Non-infériorité

Toujours vérifier :

- la **marge de non-infériorité** ;
- le sens de l'hypothèse ;
- le caractère souvent **unilatéral** du test ;
- la justification clinique de la marge.

5. Comment écrire une bonne réponse d'examen

5.1. Trame universelle

1. identifier le design ;
2. identifier le type de critère principal ;
3. lister les paramètres nécessaires ;
4. préciser lesquels sont présents dans l'article ;
5. faire le calcul ou le raisonnement détaillé ;
6. conclure par un chiffre ou une interprétation claire.

5.2. Si le calcul exact est impossible

Une excellente copie peut écrire :

Le calcul exact n'est pas reconstituable car il manque [paramètre], mais la logique du dimensionnement est claire : il faut [liste des paramètres], et l'ordre de grandeur attendu est de [chiffre ou intervalle plausible].

5.3. Si l'étude est observationnelle

Une excellente copie peut écrire :

Ici, la bonne discussion n'est pas vraiment celle d'un NSN a priori. Il faut plutôt raisonner en nombre d'événements, en largeur des IC95%, et en limites de validité causale.

6. Annales résolues avec calculs détaillés

Mode d'emploi de cette section

Pour vous en servir comme d'un vrai cahier d'exercices :

1. lisez d'abord **l'article et le contexte** ;
2. reformulez la **question d'examen** avec vos propres mots ;
3. relevez dans l'article les **hypothèses utiles** avant de regarder le calcul ;
4. comparez ensuite votre raisonnement à la **résolution guidée** et à la **réponse type**.

6.1. 2011 EPI

Article et contexte

Article support : **Kalager et al., NEJM 2010**, sur l'effet de la **mammographie de dépistage** sur la mortalité par cancer du sein en Norvège.

- programme national commencé en 1996, avec extension géographique progressive pendant 9 ans;
- femmes âgées de 50 à 69 ans, mammographie tous les 2 ans;
- 40 075 femmes atteintes d'un cancer du sein incluses dans l'analyse;
- comparaison en **quatre groupes** :
 - groupe actuel avec dépistage;
 - groupe actuel sans dépistage;
 - groupe historique miroir du premier;
 - groupe historique miroir du second.

Ce point est crucial : l'article n'oppose pas simplement "dépistage" contre "pas de dépistage". Il essaie aussi de **séparer l'effet du dépistage** de l'amélioration générale de la prise en charge du cancer du sein au fil du temps.

Exercice contextualisé

L'énoncé vous demande une **puissance a posteriori**, mais le vrai travail préalable est de comprendre **sur quel effet** vous allez la calculer.

Dans ce papier, on peut raisonner sur deux effets différents :

1. l'effet **brut** du groupe avec dépistage par rapport à son groupe historique;
2. l'effet **net attribuable spécifiquement au dépistage**, après prise en compte du fait que la mortalité baisse aussi dans le groupe non dépisté grâce aux progrès thérapeutiques et diagnostiques.

Si vous ne faites pas cette distinction, vous ne répondez pas à la vraie difficulté de la question.

Ce qu'il faut aller chercher dans l'article

- baisse de mortalité dans le groupe avec dépistage par rapport à son groupe historique :
 - 7,2 décès en moins pour 100 000 personnes-années;
 - RR = 0,72, IC95% 0,63 à 0,81;
- baisse de mortalité dans le groupe sans dépistage par rapport à son groupe historique :
 - 4,8 décès en moins pour 100 000 personnes-années;
 - RR = 0,82, IC95% 0,71 à 0,93;
- effet attribuable au dépistage seul :
 - différence absolue $7,2 - 4,8 = 2,4$ décès pour 100 000 personnes-années;
 - soit environ **un tiers** de la baisse totale;
 - les auteurs traduisent cela par une réduction relative d'environ 10%, avec IC95% -4% à 24%.

Ce que l'examineur veut vérifier

- savez-vous repérer qu'il s'agit d'une **quasi-expérience** et non d'un essai randomisé classique;
- savez-vous distinguer **effet brut** et **effet net du dépistage**;
- savez-vous comprendre qu'une puissance post hoc calculée sur l'effet brut et sur l'effet net ne raconte **pas la même histoire**.

Comment passer des chiffres de l'article au calcul

La difficulté de cette annale n'est pas seulement de réciter une formule. Il faut montrer **quel nombre on prend dans l'article et pourquoi**.

Étape 1. Décider quel effet on veut quantifier

Deux options sont possibles :

- **effet brut du programme avec dépistage** : $RR = 0,72$, IC95% $0,63$ à $0,81$;
- **effet net attribuable au dépistage seul** après soustraction de l'amélioration générale : environ 10%, IC95% -4% à 24% .

La copie doit dire explicitement quel effet est choisi. Si on ne le dit pas, le calcul est ambigu.

Étape 2. Choisir la bonne échelle

Pour le $RR = 0,72$, on travaille sur l'échelle **logarithmique**, car un risque relatif ne se traite pas directement sur l'échelle brute.

On applique donc la formule générale rappelée en section 3.3 :

$$SE(\log RR) \approx \frac{\log(\text{borne sup}) - \log(\text{borne inf})}{3,92}$$

Pour l'effet net 10% avec IC95% -4% à 24% , l'effet est déjà présenté sur une échelle linéaire centrée autour de 0, donc on peut utiliser directement :

$$SE \approx \frac{\text{largeur de l'IC}}{3,92}$$

Étape 3. Refaire le calcul sur l'effet brut

Rappel de formule

Pour un risque relatif, on travaille sur l'échelle logarithmique et on utilise :

$$SE(\log RR) \approx \frac{\log(\text{borne sup}) - \log(\text{borne inf})}{3,92}$$

Ici, l'article donne $RR = 0,72$ et IC95% $0,63$ à $0,81$. On remplace donc dans la formule :

$$SE(\log RR) \approx \frac{\log(0,81) - \log(0,63)}{3,92} \approx 0,064$$

Pour voir d'où vient ce $0,064$, on peut expliciter :

$$\log(0,81) \approx -0,21$$

$$\log(0,63) \approx -0,46$$

$$\log(0,81) - \log(0,63) \approx -0,21 - (-0,46) = 0,25$$

$$\frac{0,25}{3,92} \approx 0,064$$

Rappel de formule

Une fois le SE obtenu, on calcule un score de Wald grossier par :

$$z \approx \frac{|\log(\widehat{RR})|}{SE(\log RR)}$$

Ici :

$$z \approx \frac{|\log(0,72)|}{0,064} \approx \frac{0,329}{0,064} \approx 5,12$$

avec :

$$\log(0,72) \approx -0,33$$

$$\frac{|\log(0,72)|}{0,064} \approx \frac{0,33}{0,064} \approx 5,12$$

Si l'on veut traduire ce z en puissance observée grossière, on utilise :

$$\text{puissance} \approx \Phi(z - 1,96)$$

donc :

$$\Phi(5,12 - 1,96) = \Phi(3,16)$$

Comme $\Phi(3)$ vaut déjà environ 0,9987, $\Phi(3,16)$ est encore un peu au-dessus et donc **très proche de 1**.

Interprétation :

- z est très au-dessus de 1,96 ;
- donc le signal statistique est très fort ;
- si l'on force une puissance a posteriori sur cet effet brut, elle sera **quasi égale à 100%**.

Étape 4. Refaire le calcul sur l'effet net

Ici, il faut être très explicite sur l'origine des nombres :

- 10% est l'**estimation ponctuelle** rapportée par l'article pour l'effet net du dépistage ;
- -4% et 24% sont les **bornes de l'IC95%** données par l'article ;
- en revanche, SE et z ne sont **pas** donnés par l'article : on les **reconstruit**.

On travaille ici sur une échelle linéaire, donc on utilise directement :

$$SE \approx \frac{\text{borne sup} - \text{borne inf}}{3,92}$$

En remplaçant par les valeurs de l'article :

$$SE \approx \frac{0,24 - (-0,04)}{3,92} = \frac{0,28}{3,92} \approx 0,071$$

Ce calcul veut simplement dire :

- largeur de l'IC = $0,24 - (-0,04) = 0,28$;
- puis $0,28 / 3,92 \approx 0,071$.

On a donc reconstruit un SE d'environ $0,071$.

Rappel de formule

Sur une échelle linéaire, le score de Wald grossier s'écrit :

$$z \approx \frac{\hat{\theta}}{SE}$$

Ici, l'effet estimé vaut $0,10$, donc :

$$z \approx \frac{0,10}{0,071} \approx 1,40$$

Autrement dit :

- effet estimé = $0,10$;
- erreur standard reconstruite = $0,071$;
- donc $z \approx 0,10 / 0,071 \approx 1,40$.

Interprétation :

- z est inférieur à $1,96$;
- l'IC95% recouvre 0 ;
- la puissance observée grossière n'est que d'environ 30%.

Si l'on veut expliciter le passage vers 30%, on utilise :

$$\text{puissance} \approx \Phi(z - 1,96)$$

donc :

$$\Phi(1,40 - 1,96) = \Phi(-0,56)$$

Comme $\Phi(0) = 0,50$, une valeur négative comme $-0,56$ donne forcément quelque chose **en dessous de 0,50**.

Avec une table de la loi normale, on trouve environ $0,29$, soit **29%**, que l'on peut arrondir à **30%**.

Étape 5. Conclure correctement

La conclusion attendue n'est pas seulement numérique :

- sur l'effet **brut**, l'étude paraît très puissante ;
- sur l'effet **spécifique** du dépistage, elle est beaucoup moins convaincante ;
- l'article est donc surtout limité par la **difficulté d'isoler l'effet causal propre du dépistage**, pas par un manque global de femmes incluses.

Calcul grossier si l'on force une puissance observée sur le résultat brut

Si l'on prend la comparaison la plus simple, celle du groupe "screening" contre son groupe historique :

Formule de départ :

$$SE(\log RR) \approx \frac{\log(\text{borne sup}) - \log(\text{borne inf})}{3,92}$$

Application numérique :

$$SE(\log RR) \approx \frac{\log(0,81) - \log(0,63)}{3,92} \approx 0,064$$

Décomposition mentale possible :

$$\log(0,81) \approx -0,21, \quad \log(0,63) \approx -0,46$$

$$\log(0,81) - \log(0,63) \approx 0,25$$

$$\frac{0,25}{3,92} \approx 0,064$$

Puis, avec :

$$z \approx \frac{|\log(\widehat{RR})|}{SE(\log RR)}$$

$$z \approx \frac{|\log(0,72)|}{0,064} \approx 5,12$$

avec :

$$\log(0,72) \approx -0,33$$

$$\frac{|\log(0,72)|}{0,064} \approx \frac{0,33}{0,064} \approx 5,12$$

Pour passer de ce z à une puissance observée grossière, on utilise :

$$\text{puissance} \approx \Phi(z - 1,96)$$

donc :

$$\Phi(5,12 - 1,96) = \Phi(3,16)$$

Or $\Phi(3,16)$ signifie : probabilité qu'une variable normale standard soit inférieure à 3,16.

En pratique :

- avec une **table de la loi normale**, on trouve une valeur très proche de 1 ;
- sans table, il suffit de savoir que $\Phi(3) \approx 0,9987$;
- donc $\Phi(3,16)$ est encore un peu au-dessus.

Conclusion : la puissance observée est **pratiquement égale à 1**, c'est-à-dire **quasi 100%**.

Calcul grossier sur l'effet net réellement attribué au dépistage

Si l'on essaie maintenant de raisonner sur l'effet propre du dépistage, estimé à environ 10% avec IC95% -4% à 24% :

Formule de départ :

$$SE \approx \frac{\text{borne sup} - \text{borne inf}}{3,92}$$

Application numérique :

$$SE \approx \frac{0,24 - (-0,04)}{3,92} \approx 0,071$$

Puis, avec :

$$z \approx \frac{\hat{\theta}}{SE}$$

$$z \approx \frac{0,10}{0,071} \approx 1,40$$

Puis :

$$\text{puissance} \approx \Phi(1,40 - 1,96) = \Phi(-0,56) \approx 0,29$$

Donc la puissance observée approximative n'est que d'environ **30%**.

Ce qu'il faut comprendre

La section "puissance" de la copie doit faire apparaître ceci :

- si vous regardez l'effet brut, l'étude semble **très puissante** ;
- si vous regardez le bénéfice **spécifique** du dépistage, la puissance est **nettement plus faible** ;
- la question de fond n'est donc pas "y avait-il assez de femmes ?", mais "l'étude pouvait-elle isoler proprement un effet du dépistage au-delà des progrès généraux de prise en charge ?".

Réponse type

L'article est une étude quasi-expérimentale fondée sur le déploiement progressif du dépistage mammographique en Norvège. Il faut donc d'abord distinguer l'effet brut du groupe dépisté par rapport à son groupe historique et l'effet net attribuable au dépistage seul. Sur l'effet brut, la baisse de mortalité est marquée (RR = 0,72, IC95% 0,63 à 0,81) et une puissance a posteriori serait très élevée. Mais l'article montre aussi qu'une partie importante de la baisse de mortalité existe dans le groupe non dépisté, ce qui réduit l'effet propre du dépistage à environ 10% (IC95% -4% à 24%). Si l'on force un calcul sur cet effet net, la puissance devient faible, autour de 30%. La bonne conclusion est donc que l'étude est grande, mais qu'elle distingue imparfaitement le bénéfice spécifique du dépistage de l'amélioration globale des traitements.

6.2. 2011 RC

Article et contexte

Article support : **James et al., NEJM 2010**, essai **SCOUT** sur la **sibutramine** chez des sujets en surpoids ou obèses à haut risque cardiovasculaire.

- 10 744 sujets inclus au départ ;
- après une phase de run-in simple aveugle de 6 semaines, 9804 sujets randomisés ;
- participants âgés d'au moins 55 ans, avec maladie cardiovasculaire préexistante, diabète de type 2, ou les deux ;
- critère principal : temps jusqu'au premier événement du composite :
 - infarctus du myocarde non fatal ;
 - AVC non fatal ;
 - arrêt cardiaque récupéré ;
 - décès cardiovasculaire.

L'article est donc un **essai de survie / temps jusqu'à événement**, pas un essai à critère binaire mesuré une seule fois.

Exercice contextualisé

La question d'examen ne vous demande pas de refaire tout le calcul du NSN.

Elle vous demande de juger la phrase suivante :

L'analyse est réalisée en ITT et le nombre de sujets nécessaire est calculé en tenant compte d'un nombre de perdus de vue possible. Qu'en pensez-vous ?

Pour répondre, vous devez lire l'article avec trois idées en tête :

1. l'essai est en **ITT** ;
2. le critère principal est un **événement cardiovasculaire** que l'on peut continuer à observer même si le traitement est arrêté ;
3. il faut distinguer **arrêt du traitement** et **vraie perte de suivi**.

Ce qu'il faut aller chercher dans l'article

- planification :
 - environ 9000 sujets ;
 - 2160 événements confirmés ;
 - puissance 80% ;
 - hypothèse de taux annuel d'événements placebo de 7% ;
 - hypothèse de discontinuation d'environ 30% ;
- après randomisation :
 - 4906 dans le bras sibutramine ;
 - 4898 dans le bras placebo ;
- suivi moyen du traitement : 3,4 ans ;
- résultat principal :
 - 11,4% d'événements dans le bras sibutramine ;
 - 10,0% dans le bras placebo ;
- pertes de vue réelles après randomisation :
 - 47 dans le bras sibutramine ;
 - 58 dans le bras placebo ;
- arrêts permanents du traitement beaucoup plus fréquents :
 - environ 4046 au total, soit un peu plus de 40%.

Ce que l'examineur veut vérifier

- savez-vous reconnaître qu'en ITT, un sujet reste analysable **même s'il arrête le traitement**, tant que l'on continue à recueillir les événements ;
- savez-vous dire qu'une inflation du NSN pour "perdus de vue" n'a de sens que s'il y a risque de **perdre l'information sur le critère principal** ;
- savez-vous montrer, avec les données de l'article, que la vraie perte d'information est faible alors que la non-observance est forte.

Résolution guidée

Le point clé est le suivant :

- l'essai a beaucoup d'**arrêts de traitement** ;
- mais très peu de **vraies pertes de vue** ;
- or l'ITT tolère les premiers, mais pas les secondes.

On peut chiffrer cela simplement :

- pertes de vue réelles :

$$\frac{47 + 58}{9804} \approx 1,1\%$$

- discontinuations de traitement :

$$\frac{4046}{9804} \approx 41,3\%$$

Donc l'article mélange potentiellement deux réalités très différentes si l'on n'est pas précis dans le vocabulaire.

Réponse type

Le papier SCOUT est un essai ITT à critère principal de type temps jusqu'à événement cardiovasculaire. Il faut donc distinguer les **arrêts de traitement**, fréquents, des **vraies pertes de suivi**, rares. En ITT, un patient qui arrête la sibutramine reste analysable si les événements cardiovasculaires continuent à être recueillis. Or c'est justement ce qui était prévu dans l'étude. Les chiffres de l'article montrent que les pertes de vue réelles sont faibles (105/9804, soit environ 1,1%), alors que les arrêts de traitement sont très nombreux (4046/9804, soit plus de 40%). Il est donc légitime de majorer un NSN pour anticiper des **données manquantes sur le critère principal**, mais il ne faut pas confondre cela avec la simple non-observance. La critique méthodologique attendue est là.

6.3. 2012 EPI

Article et contexte

Article support : **Cooper et al., NEJM 2011**, étude de cohorte rétrospective sur les **médicaments du TDAH** et les événements cardiovasculaires graves chez l'enfant et l'adulte jeune.

- 1 200 438 enfants et jeunes adultes de 2 à 24 ans ;
- 2 579 104 personnes-années de suivi ;
- 373 667 personnes-années d'utilisation actuelle de médicaments du TDAH ;
- issue principale : événements cardiovasculaires graves :
 - mort subite d'origine cardiaque ;

- infarctus du myocarde ;
- AVC ;
- seulement 81 événements au total dans toute la cohorte, soit 3,1 pour 100 000 personnes-années.

Exercice contextualisé

Le piège de cet article est classique :

- la cohorte est **énorme** ;
- mais les événements intéressants sont **très rares**.

Quand l'énoncé vous demande de retrouver approximativement la puissance, il veut que vous voyiez que la taille globale de la base de données ne suffit pas. Ce qui compte ici, c'est le **nombre d'événements informatifs**.

Ce qu'il faut aller chercher dans l'article

- événements chez les utilisateurs actuels : 7 ;
- événements chez les non-utilisateurs : 49 ;
- HR ajusté = 0,75 ;
- IC95% 0,31 à 1,85.

Il faut aussi relever la phrase importante de la conclusion :

- l'étude ne montre pas de sur-risque significatif ;
- mais **un doublement du risque ne peut pas être exclu**.

Ce que l'examineur veut vérifier

- savez-vous expliquer pourquoi une cohorte gigantesque peut malgré tout être **peu puissante** pour un événement rare ;
- savez-vous utiliser l'intervalle de confiance pour produire un ordre de grandeur de puissance post hoc ;
- savez-vous relier la faible puissance à la largeur de l'IC95%.

Approximation par l'intervalle de confiance

Formule de départ pour un hazard ratio :

$$SE(\log HR) \approx \frac{\log(\text{borne sup}) - \log(\text{borne inf})}{3,92}$$

Application numérique :

$$SE(\log HR) \approx \frac{\log(1,85) - \log(0,31)}{3,92} \approx 0,456$$

Sans sauter d'étapes :

$$\log(1,85) \approx 0,62, \quad \log(0,31) \approx -1,17$$

$$\log(1,85) - \log(0,31) \approx 0,62 - (-1,17) = 1,79$$

$$\frac{1,79}{3,92} \approx 0,456$$

Puis, avec le score de Wald :

$$z \approx \frac{|\log(\widehat{HR})|}{SE(\log HR)}$$

$$z \approx \frac{|\log(0,75)|}{0,456} \approx 0,63$$

avec :

$$\log(0,75) \approx -0,29$$

$$\frac{|\log(0,75)|}{0,456} \approx \frac{0,29}{0,456} \approx 0,63$$

Puissance observée approximative :

$$\text{puissance} \approx \Phi(z - 1,96)$$

$$\Phi(0,63 - 1,96) \approx 0,09$$

car :

$$0,63 - 1,96 = -1,33$$

$$\Phi(-1,33) \approx 0,09$$

soit environ **10%**.

Ce qu'il faut comprendre

La copie doit faire apparaître explicitement que :

- l'effectif total est très grand ;
- mais le nombre d'événements dans le groupe vraiment informatif est minuscule ;
- c'est pour cela que l'IC95% reste très large ;
- donc l'étude rassure contre un effet massif, mais n'exclut pas proprement un sur-risque modéré à important.

Réponse type

L'article inclut plus de 1,2 million de sujets, mais il ne faut pas se laisser impressionner par la taille globale de la cohorte : il n'y a que 81 événements cardiovasculaires graves au total et seulement 7 chez les utilisateurs actuels. C'est donc une situation typique où la cohorte est grande mais l'information statistique utile reste faible. À partir du HR = 0,75 et de son IC95% 0,31 à 1,85, on retrouve une puissance post hoc très grossière d'environ 10%. La bonne conclusion est que l'étude ne met pas en évidence de sur-risque, mais qu'elle manque de puissance pour exclure proprement un effet modéré, ce que montre déjà la largeur de l'intervalle de confiance.

6.4. 2012 RC

Article et contexte

Article support : **AIM-HIGH Investigators, NEJM 2011**, essai comparant **niacine à libération prolongée** versus placebo chez des patients ayant une **maladie cardiovasculaire athéroscléreuse**, un **HDL bas** et déjà une prise en charge lipidique intensive.

- 3414 patients randomisés :
 - 1718 niacine;
 - 1696 placebo;
- traitement de fond par simvastatine, avec ézétimibe si besoin;
- critère principal : premier événement du composite cardiovasculaire;
- essai interrompu après environ 3 ans de suivi moyen pour absence d'efficacité;
- taux observés du critère principal :
 - 16,4% dans le groupe niacine;
 - 16,2% dans le groupe placebo.

Exercice contextualisé

Cette annale est typique d'un essai **event-driven**.

Vous ne devez pas raisonner comme si les auteurs avaient choisi un effectif fixe à partir d'une simple formule "deux proportions". En réalité, le papier dit explicitement qu'il attend :

- un **nombre d'événements** donné;
- sur une **durée moyenne de suivi** donnée;
- pour détecter une **réduction relative** donnée.

Votre travail consiste donc à montrer que le besoin était de l'ordre de **plusieurs milliers de patients**.

Ce qu'il faut aller chercher dans l'article

- cible de dimensionnement : 800 événements primaires adjudiqués;
- suivi moyen prévu : 4,6 ans;
- puissance 85%;
- alpha unilatéral 0,025;
- effet visé : réduction relative de 25% du critère principal;
- effectif finalement randomisé : 3414 sujets.

Ce que l'examineur veut vérifier

- savez-vous reconnaître un essai **piloté par le nombre d'événements**;
- savez-vous exploiter un risque observé pour revenir à un **ordre de grandeur d'effectif**;
- savez-vous dire que le calcul exact n'est pas parfaitement reconstituable sans toutes les hypothèses de recrutement et de censure.

Raisonnement de coin de table

L'article observé donne environ 16,2% d'événements à 3 ans dans le groupe placebo.

On transforme cela en un taux annuel sous hypothèse exponentielle :

Formule de départ :

$$\lambda \approx -\frac{\log(1 - R(t))}{t}$$

Application numérique :

$$\lambda_0 \approx -\frac{\log(1 - 0,162)}{3} \approx 0,0589 \text{ par an}$$

En détaillant :

$$1 - 0,162 = 0,838$$

$$\log(0,838) \approx -0,177$$

$$-\frac{\log(0,838)}{3} \approx \frac{0,177}{3} \approx 0,0589$$

Risque placebo sur 4,6 ans :

Formule de départ :

$$R(t) = 1 - e^{-\lambda t}$$

Application numérique :

$$R_0(4,6) = 1 - e^{-0,0589 \times 4,6} \approx 0,237$$

Sous un HR = 0,75, risque traité :

Comme $\lambda_1 = \text{HR} \times \lambda_0$, on obtient :

$$R_1(4,6) \approx 1 - e^{-0,0589 \times 0,75 \times 4,6} \approx 0,184$$

Risque moyen d'événement :

Formule de départ :

$$\bar{R} \approx \frac{R_0 + R_1}{2}$$

$$\bar{R} \approx \frac{0,237 + 0,184}{2} \approx 0,211$$

Pour obtenir 800 événements :

Formule de départ :

$$N \approx \frac{D}{\bar{R}}$$

$$N \approx \frac{800}{0,211} \approx 3798$$

soit **environ 3800 sujets au total**, donc **autour de 1900 par bras**.

Interprétation

Ce calcul ne prétend pas reconstituer exactement le protocole, mais il montre bien que :

- l'ordre de grandeur attendu est celui de **quelques milliers** de patients ;
- l'effectif de 3414 randomisés est cohérent avec un essai event-driven de cette nature.

Réponse type

L'essai AIM-HIGH n'est pas dimensionné comme un essai simple à deux proportions, mais comme un essai **event-driven**. Les auteurs annoncent vouloir observer 800 événements primaires, avec une puissance de 85%, un alpha unilatéral de 2,5%, et une réduction relative attendue de 25%. En utilisant le risque observé dans le groupe placebo pour retrouver un ordre de grandeur du risque cumulé sur la durée de suivi, on retombe sur un besoin d'environ 3500 à 3800 sujets randomisés, soit autour de 1700 à 1900 par bras. L'effectif réellement inclus (3414) est donc cohérent avec cette logique.

6.5. 2013 EPI

Article et contexte

Article support : **Lasalvia et al., Lancet 2013**, enquête transversale sur la **discrimination vécue et anticipée** chez les personnes ayant un **trouble dépressif majeur**.

- enquête **transversale**, pas essai interventionnel ;
- 39 sites dans 35 pays ;
- 1082 participants adultes avec diagnostic de dépression majeure ;
- outil principal : DISC-12 ;
- résultat descriptif central :
 - 855 / 1082, soit 79%, rapportent une discrimination vécue dans au moins un domaine de vie.

Exercice contextualisé

La question d'examen est volontairement piégeuse :

Dans l'article on ne trouve pas de justification de la taille de l'échantillon. A posteriori, il n'est pas facile de calculer une puissance statistique. Essayez tout de même de proposer un tel calcul, même très rudimentaire.

Or le papier :

- n'est pas un essai ;
- ne teste pas une hypothèse principale unique ;
- rapporte surtout des **estimations descriptives** et des associations multivariées.

La bonne réponse consiste donc à dire pourquoi un calcul de puissance classique est artificiel, puis à proposer un raisonnement de **précision**.

Ce qu'il faut aller chercher dans l'article

- type de dessin : **cross-sectional survey** ;
- taille de l'échantillon : 1082 ;
- résultat principal facilement exploitable :
 - 855 participants sur 1082 ont rapporté une discrimination vécue ;
- autres chiffres qui montrent le caractère très descriptif de l'étude :
 - 405 (37%) ont renoncé à initier une relation proche ;

- 271 (25%) ont renoncé à postuler à un emploi ;
- 218 (20%) ont renoncé à une formation.

Ce que l'examinateur veut vérifier

- savez-vous dire qu'il n'y a **pas de vrai NSN prospectif** ici ;
- savez-vous remplacer l'idée de puissance par l'idée de **précision d'estimation** ;
- savez-vous montrer que 1082 sujets donnent déjà une estimation assez stable de la proportion principale.

Calcul de précision simple

Formule de départ pour la proportion observée :

$$p = \frac{x}{n}$$

$$p = \frac{855}{1082} \approx 0,790$$

Puis l'erreur standard d'une proportion :

$$SE(p) = \sqrt{\frac{p(1-p)}{n}}$$

$$SE(p) = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{0,79 \times 0,21}{1082}} \approx 0,0124$$

Enfin l'intervalle de confiance approximatif :

$$IC95\% \approx p \pm 1,96 \times SE(p)$$

$$IC95\% \approx 0,79 \pm 1,96 \times 0,0124$$

soit environ :

$$[0,766; 0,814]$$

c'est-à-dire **76,6% à 81,4%**.

Ce qu'il faut comprendre

Avec cette annale, l'examinateur ne veut pas un faux calcul sophistiqué. Il veut que vous disiez :

- que le papier n'est pas construit autour d'un unique test principal ;
- que le meilleur substitut d'une discussion de puissance est la **largeur de l'IC95%** autour de la proportion centrale ;
- que l'étude est assez grande pour décrire ce phénomène avec une précision raisonnable.

Réponse type

L'article est une enquête transversale internationale et ne repose pas sur un vrai calcul prospectif de nombre de sujets pour tester une hypothèse unique. Il est donc assez artificiel de parler de puissance

au sens classique. En revanche, on peut raisonner en précision : la proportion principale observée, 855/1082, soit environ 79%, est estimée avec un IC95% assez étroit, autour de 76,6% à 81,4%. Cela montre que l'effectif est globalement confortable pour décrire la fréquence de la discrimination vécue, même si le papier n'explicite pas un calcul de taille d'échantillon a priori.

6.6. 2014 EPI

Article et contexte

Article support : **Bao et al., NEJM 2013**, étude de cohorte sur la **consommation de noix** et la mortalité totale et cause-spécifique.

- deux grandes cohortes prospectives :
 - 76 464 femmes ;
 - 42 498 hommes ;
- participants exempts de cancer, cardiopathie et AVC au départ ;
- consommation de noix réévaluée tous les 2 à 4 ans ;
- 3 038 853 personnes-années de suivi ;
- 27 429 décès au total.

Exercice contextualisé

Ici encore, la question parle de “puissance a posteriori”, mais l'article ne manque clairement pas de taille.

Ce qu'il faut montrer dans votre copie, c'est que :

- la cohorte est gigantesque ;
- le nombre d'événements est gigantesque ;
- l'estimation principale est **très précise** ;
- donc la discussion méthodologique doit porter ailleurs que sur la puissance.

Ce qu'il faut aller chercher dans l'article

- résultat principal pour la consommation la plus élevée de noix (≥ 7 fois/semaine) versus jamais :

$$HR \approx 0,80, \quad IC95\% 0,73 \text{ à } 0,86$$

- effet dose-réponse monotone avec plusieurs catégories intermédiaires ;
- événements très nombreux, ce qui explique la précision de l'estimation.

Ce que l'examineur veut vérifier

- savez-vous voir qu'un papier observationnel peut être **surabondant en information** ;
- savez-vous utiliser la finesse de l'IC95% comme indicateur indirect de précision ;
- savez-vous conclure que la faiblesse du papier éventuel serait la **confusion résiduelle**, pas un manque de puissance.

Approximation par l'IC95%

Formule de départ pour un hazard ratio :

$$SE(\log HR) \approx \frac{\log(\text{borne sup}) - \log(\text{borne inf})}{3,92}$$

$$SE(\log HR) \approx \frac{\log(0,86) - \log(0,73)}{3,92} \approx 0,0418$$

En détaillant :

$$\log(0,86) \approx -0,151, \quad \log(0,73) \approx -0,315$$

$$\log(0,86) - \log(0,73) \approx -0,151 - (-0,315) = 0,164$$

$$\frac{0,164}{3,92} \approx 0,0418$$

Puis :

$$z \approx \frac{|\log(\widehat{HR})|}{SE(\log HR)}$$

$$z \approx \frac{|\log(0,80)|}{0,0418} \approx 5,34$$

avec :

$$\log(0,80) \approx -0,223$$

$$\frac{|\log(0,80)|}{0,0418} \approx \frac{0,223}{0,0418} \approx 5,34$$

Enfin :

$$\text{puissance} \approx \Phi(z - 1,96)$$

donc ici :

$$\Phi(5,34 - 1,96) = \Phi(3,38)$$

Comme $\Phi(3)$ vaut déjà environ 0,9987, une telle valeur correspond à une puissance observée pratiquement égale à 1.

Réponse type

Le papier repose sur deux très grandes cohortes prospectives totalisant plus de 118 000 participants, plus de 3 millions de personnes-années et plus de 27 000 décès. Le résultat principal pour les plus gros consommateurs de noix (HR \approx 0,80, IC95% 0,73 à 0,86) est estimé avec un intervalle très étroit, ce qui traduit une grande précision. Si l'on forçait un raisonnement en puissance observée, celle-ci serait quasi maximale. La bonne conclusion n'est donc pas "l'étude manque de puissance", mais plutôt "l'étude est largement informative, et les vraies limites éventuelles tiennent à la confusion et à l'interprétation causale".

6.7. 2016 EPI

Article et contexte

Article support : **Schnabel et al., Lancet 2015**, analyse des **tendances sur 50 ans** de la fibrillation atriale dans la cohorte de Framingham.

- 9511 participants;
- inclusion entre 1958 et 2007;
- 202 417 personnes-années;
- 1544 cas incidents de fibrillation atriale;
- analyses par périodes de 10 ans et séparées selon le sexe.

Exercice contextualisé

La question :

Comment aborder la question de la puissance statistique dans une telle étude ?

ne vise pas un calcul chiffré unique.

Elle vous demande de comprendre qu'un papier de **tendances temporelles observationnelles** ne se juge pas comme un essai randomisé à critère principal unique.

Il faut donc discuter la puissance à plusieurs niveaux :

1. sur l'ensemble de la cohorte;
2. à l'intérieur des décennies;
3. dans les analyses plus fines ou plus rares.

Ce qu'il faut aller chercher dans l'article

- ampleur globale :
 - 1544 cas incidents;
 - donc beaucoup d'information pour les analyses générales;
- structure des analyses :
 - comparaison sur cinq décennies;
 - stratification par sexe;
- remarque explicite des auteurs :
 - dans les premières décennies, le nombre de cas est plus faible;
 - les erreurs standard sont plus grandes;
 - la puissance pour détecter certaines **tendances dans le temps** est donc limitée.

Ce que l'examineur veut vérifier

- savez-vous renoncer à chercher un **NSN fictif** quand il n'y en a pas;
- savez-vous basculer vers une discussion en **nombre d'événements** et **précision**;
- savez-vous dire que la puissance n'est pas uniforme dans toutes les sous-analyses.

Ce qu'il faut écrire

La bonne réponse se structure ainsi :

1. **au niveau global**, l'étude est riche en information (1544 cas);

2. **au niveau des décennies anciennes**, les nombres sont plus petits, donc les comparaisons de tendance sont moins robustes ;
3. **au niveau des sous-groupes**, certains résultats restent fragiles malgré la taille totale de la cohorte.

Réponse type

Dans cette étude de cohorte sur 50 ans, la puissance ne doit pas être discutée comme dans un essai randomisé avec un nombre de sujets nécessaire a priori. Il faut raisonner à partir du **nombre d'événements** et de la **structure des analyses**. Globalement, l'étude est très informative, avec 9511 participants, plus de 202 000 personnes-années et 1544 cas incidents de fibrillation atriale. En revanche, lorsqu'on découpe ces données par décennies et par sexe, certaines périodes anciennes reposent sur des effectifs de cas plus faibles ; les auteurs signalent eux-mêmes des erreurs standard plus larges et une puissance limitée pour détecter certaines tendances. La bonne discussion porte donc sur la précision des estimations selon les sous-analyses, pas sur un NSN classique.

6.8. 2019 EPI

Article et contexte

Article support : **Shan et al., BMJ 2018**, étude de cohorte sur le **travail posté de nuit**, le **mode de vie** et le risque de **diabète de type 2** chez des infirmières américaines.

- deux grandes cohortes de nurses ;
- 143 410 femmes sans diabète, maladie cardiovasculaire ni cancer au départ ;
- suivi sur environ 22 à 24 ans ;
- 10 915 cas incidents de diabète de type 2 ;
- résultats principaux :
 - HR = 1,31 par tranche de 5 ans de travail de nuit tournant ;
 - HR = 2,30 par facteur de mode de vie défavorable.

Exercice contextualisé

L'énoncé vous demande :

Dans une cohorte de cette taille, comment se discute la question de la puissance statistique ?

Le but n'est pas de refaire un calcul.

Le but est de montrer que, quand une cohorte cumule :

- beaucoup de sujets ;
- beaucoup d'événements ;
- une exposition fréquente ;

la significativité statistique cesse d'être la question principale.

Ce qu'il faut aller chercher dans l'article

- taille de la population : 143 410 participantes ;
- nombre de cas incidents : 10 915 ;
- exposition d'intérêt courante dans cette population ;
- effets estimés avec précision.

Ce que l'examineur veut vérifier

- savez-vous reconnaître une situation où la puissance n'est **pas** le vrai problème ;
- savez-vous déplacer la discussion vers la **taille d'effet**, la **pertinence clinique** et les IC95% ;
- savez-vous éviter une réponse automatique de type "il faudrait calculer la puissance".

Ce qu'il faut comprendre

Dans une cohorte immense :

- il est facile d'obtenir des p-values petites ;
- ce n'est pas cela qui fait l'intérêt scientifique principal ;
- il faut surtout regarder :
 - l'ampleur des HR ;
 - leurs IC95% ;
 - la plausibilité causale et les biais résiduels.

Réponse type

Dans cette cohorte, la puissance n'est pas le sujet principal. L'étude suit plus de 143 000 infirmières pendant plus de vingt ans et observe plus de 10 000 cas incidents de diabète de type 2. Avec une exposition assez fréquente et un aussi grand nombre d'événements, l'étude est largement capable de détecter des effets modestes. La discussion doit donc porter surtout sur la **taille des associations**, leurs IC95%, et la validité causale, bien plus que sur un éventuel manque de puissance.

6.9. 2020 RC

Article et contexte

Article support : **Schnitzer et al., JAMA 2019**, essai randomisé en double aveugle sur le **tanezumab** dans l'arthrose de hanche ou de genou.

- 698 patients randomisés ;
- 3 bras : placebo, tanezumab 2,5 mg, tanezumab 2,5/5 mg ;
- critère principal = **3 co-primary endpoints** à 16 semaines :
 - WOMAC Pain ;
 - WOMAC Physical Function ;
 - PGA-OA.

L'article dit explicitement que l'effectif d'environ 230 patients par bras est choisi pour obtenir 90% de puissance **across all 3 co-primary end points**.

Exercice contextualisé

Vous lisez un essai à trois bras chez des patients ayant une arthrose symptomatique. Les auteurs annoncent un dimensionnement à 230 patients par bras.

Votre travail consiste à raisonner en deux temps :

1. refaire un calcul naïf **comme si chaque endpoint était isolé** ;
2. expliquer pourquoi ce calcul est insuffisant dès lors que la réussite de l'essai exige que **les trois endpoints** soient positifs.

Ce qu'il faut aller chercher dans l'article

- puissance : 90%;
- niveau bilatéral global : 5%;
- différences attendues :
 - delta = 1 pour WOMAC Pain;
 - delta = 1 pour WOMAC Physical Function;
 - delta = 0,32 pour PGA-0A;
- écarts-types :
 - SD = 2,73;
 - SD = 2,58;
 - SD = 0,92;
- corrélations entre co-critères : 0,93, 0,68, 0,68.

Ce que l'examineur veut vérifier

- savez-vous reconnaître un **vrai piège de co-primary endpoints** ;
- savez-vous faire un calcul simple à deux moyennes pour obtenir un **ordre de grandeur de départ** ;
- savez-vous expliquer pourquoi l'effectif réel doit être **plus grand** que ce calcul naïf.

Étape 1 : calcul naïf si chaque endpoint était analysé séparément

On applique la formule "deux moyennes" avec $z_{\alpha} = 1,96$ et $z_{\beta} = 1,28$.

Formule de départ :

$$n_{\text{par bras}} = \frac{2(z_{1-\alpha/2} + z_{1-\beta})^2 \sigma^2}{\delta^2}$$

Critère 1 :

$$n \approx \frac{2(1,96 + 1,28)^2 \times 2,73^2}{1^2} \approx 156$$

Critère 2 :

$$n \approx \frac{2(1,96 + 1,28)^2 \times 2,58^2}{1^2} \approx 140$$

Critère 3 :

$$n \approx \frac{2(1,96 + 1,28)^2 \times 0,92^2}{0,32^2} \approx 174$$

Si l'on oubliait complètement la multiplicité du critère principal, on conclurait donc à **140 à 174 sujets par bras**, soit **environ 165 par bras**.

Pourquoi ce calcul naïf ne suffit pas

Ici, l'essai n'est pas positif si **un seul** endpoint marche.

Il n'est positif que si **les trois co-primary endpoints** sont tous significatifs.

Autrement dit :

- il ne faut pas seulement alimenter un test sur une variable ;
- il faut maintenir une bonne puissance pour un **critère principal composite logique** ;
- le coût en sujets dépend alors de la **corrélation** entre les trois mesures.

Le corrigé officiel s'appuie justement sur cette idée et utilise une méthode dédiée aux **multiple co-primary endpoints**.

Résolution guidée

Le raisonnement attendu est le suivant :

1. partir du calcul simple à deux moyennes pour montrer que l'ordre de grandeur de base est autour de 165 par bras ;
2. rappeler que l'article n'a pas un critère principal unique mais **trois co-critères simultanés** ;
3. conclure qu'il faut majorer l'effectif en fonction de la structure de corrélation.

Le corrigé retrouve alors environ **195 sujets par bras**, soit :

Formule de départ :

$$N_{\text{total}} = n_{\text{par bras}} \times \text{nombre de bras}$$

$$195 \times 3 = 585 \text{ sujets au total}$$

Les auteurs ont finalement choisi :

$$230 \times 3 = 690 \text{ sujets}$$

donc un choix un peu plus conservateur que le minimum reconstruit.

Réponse type

L'article correspond à un essai à trois bras avec trois co-primary endpoints continus. Si chaque critère était dimensionné isolément, un calcul standard à deux moyennes donnerait un besoin d'environ 140 à 174 sujets par bras, soit autour de 165 par bras. Mais cette valeur est trop faible, car l'essai n'est concluant que si les **trois** endpoints sont positifs. Il faut donc intégrer la corrélation entre co-critères et le fait que la puissance doit porter sur l'ensemble du critère principal multiple. Le corrigé retrouve alors environ 195 sujets par bras, soit 585 au total, ce qui est cohérent avec le choix des auteurs de retenir 230 sujets par bras.

6.10. 2021 RC

Article et contexte

Article support : **Roy et al., PLoS Medicine 2020**, essai **cluster randomisé apparié par paires** dans des cliniques VIH en Zambie, évaluant les **adherence clubs**.

- 10 cliniques randomisées (5 intervention, 5 contrôle) ;
- critère principal : **temps jusqu'au premier retrait tardif du traitement** (>7 jours de retard) ;
- hypothèse des auteurs : environ 65% des patients du groupe contrôle auront au moins un retard important, avec une réduction relative attendue de 50% sous intervention ;
- calcul de l'article fondé sur une formule de **matched-pair cluster randomized trial**, avec 5 paires de cliniques et 120 patients par clinique.

Exercice contextualisé

L'article original dimensionne un **essai en cluster apparié**, avec une formule assez spécialisée.

Mais la question d'examen vous demande autre chose :

Si le design avait été celui d'un essai randomisé standard, combien de patients aurait-il fallu inclure ?

Vous devez donc volontairement **simplifier le papier** :

- oublier l'appariement des cliniques ;
- oublier l'effet cluster ;
- oublier le temps jusqu'au retard ;
- ramener le problème à un **essai individuel à deux proportions**.

Ce qu'il faut aller chercher dans l'article

- le risque de base dans le groupe contrôle est de l'ordre de 60% à 65% ;
- l'effet attendu est une **réduction relative de 50%** ;
- donc, en version simplifiée de cours, on peut prendre :
 - $P_1 = 0,60$ dans le bras contrôle ;
 - $P_2 = 0,30$ dans le bras intervention.

Ce que l'examineur veut vérifier

- savez-vous distinguer le **design réel du papier** du **design fictif demandé par l'énoncé** ;
- savez-vous enlever mentalement tout ce qui tient au cluster pour revenir à une formule standard ;
- savez-vous donner un chiffre cohérent avec le corrigé : **un peu moins de 50 par bras**.

Calcul à deux proportions pour l'essai individuel fictif

Avec une allocation 1 :1, un alpha bilatéral à 5% et une puissance de 80% :

Formule de départ :

$$\bar{p} = \frac{p_1 + p_2}{2}$$

$$\bar{p} = \frac{0,60 + 0,30}{2} = 0,45$$

Puis on applique la formule standard à deux proportions :

$$n_{\text{par bras}} = \frac{\left[z_{1-\alpha/2} \sqrt{2\bar{p}(1-\bar{p})} + z_{1-\beta} \sqrt{p_1(1-p_1) + p_2(1-p_2)} \right]^2}{(p_1 - p_2)^2}$$

$$n \approx \frac{\left[1,96 \sqrt{2 \times 0,45 \times 0,55} + 0,84 \sqrt{0,60 \times 0,40 + 0,30 \times 0,70} \right]^2}{(0,60 - 0,30)^2}$$

$$n \approx 41,9$$

On retient donc environ **42 patients par bras**, soit en pratique **45 à 50 par bras** après arrondi prudent.

Point pédagogique important

Le calcul de l'article et celui de l'examen ne répondent **pas à la même question** :

- l'article : essai en **clusters appariés**, avec coefficient de variation et nombre de cliniques ;
- l'examen : **essai individuel standard** pour vérifier que vous savez refaire la base "deux proportions".

Réponse type

Le papier original est un essai en clusters appariés, mais la question demande explicitement ce qu'aurait donné un **essai randomisé individuel standard**. Il faut donc simplifier le problème en un calcul à deux proportions. En reprenant un risque de base autour de 60% dans le bras contrôle et une réduction relative de 50%, on obtient $P1 = 0,60$ et $P2 = 0,30$. Le calcul standard donne environ 42 patients par bras pour $\alpha = 5\%$ et une puissance de 80%, soit **un peu moins de 50 patients par bras** après arrondi, ce qui est cohérent avec le corrigé.

6.11. 2022 RC

Article et contexte

Article support : **Bacharier et al., NEJM 2021**, essai de phase 3, randomisé, en double aveugle, chez des enfants de 6 à 11 ans ayant un **asthme modéré à sévère non contrôlé**.

- 52 semaines de suivi ;
- randomisation 2 : 1 dupilumab versus placebo ;
- critère principal : **annualized rate of severe asthma exacerbations** ;
- analyse des exacerbations par **modèle binomial négatif**.

Exercice contextualisé

Ici, l'examinateur ne vous demande pas un petit calcul de cours "deux proportions" ou "deux moyennes".

Il veut vérifier que vous savez **lire une section de dimensionnement plus avancée**.

La bonne démarche est :

1. repérer dans l'article quels paramètres servent au calcul ;
2. expliquer pourquoi l'issue est un **comptage récurrent** et non un simple oui/non ;
3. montrer pourquoi le **binomial négatif** est utilisé à la place d'un modèle de Poisson simple.

Ce qu'il faut aller chercher dans l'article

La section statistique du papier donne beaucoup plus d'indices que ce que laisse penser le corrigé court :

- deux populations principales d'efficacité ;
- échantillons visés d'environ 345 et 255 patients ;
- puissance d'environ 94% et 96% ;
- réductions relatives visées de 54% et 60% ;
- taux annualisés d'exacerbation dans le groupe placebo de 0,7 et 0,8 ;
- randomisation 2 : 1 ;
- durée moyenne d'exposition de 0,9 an ;
- **dispersion = 1,5** pour le modèle binomial négatif.

Ce que l'examineur veut vérifier

- savez-vous reconnaître un **critère de comptage récurrent** ;
- savez-vous dire qu'il faut un **taux de base**, une **durée d'observation**, un **effet relatif visé**, et non seulement une proportion ;
- savez-vous expliquer le rôle de la **surdispersion**.

Ce qu'il faut écrire

Pour reconstruire le calcul, il faut au minimum :

- un α ;
- une puissance ;
- le **taux annualisé d'exacerbations** dans le groupe contrôle ;
- l'effet minimal à détecter, ici un **rate ratio** ou une réduction relative attendue ;
- la durée moyenne de suivi ou d'exposition ;
- le ratio d'allocation (2 : 1) ;
- un paramètre de **surdispersion**, ici 1, 5.

Pourquoi le binomial négatif est central

Le nombre d'exacerbations par patient n'est pas un simple résultat binaire :

- certains enfants ont 0 exacerbation ;
- d'autres en ont plusieurs ;
- les comptages varient plus que ce que prévoirait un modèle de Poisson idéal.

Le binomial négatif permet précisément de modéliser cette **variabilité excédentaire**.

Si on faisait semblant d'utiliser un modèle de Poisson simple, on risquerait de **sous-estimer le NSN**.

Réponse type

Le papier correspond à un essai où le critère principal est un **nombre annualisé d'exacerbations sévères**, analysé en **binomial négatif**. Le calcul du nombre de sujets ne repose donc pas seulement sur α , β et une différence binaire, mais aussi sur un **taux de base d'exacerbations**, une **durée moyenne d'exposition**, un **ratio d'allocation 2 : 1**, un **effet relatif attendu**, et un **paramètre de dispersion**. L'article mentionne justement des tailles visées d'environ 345 et 255 patients selon la population d'efficacité, avec une dispersion supposée de 1, 5. La réponse attendue est donc surtout méthodologique : on doit expliquer que le NSN dépend d'un modèle de comptage avec surdispersion, et qu'un calcul naïf de type Poisson ou deux proportions serait inadapté.

6.12. 2023 RC

Article et contexte

Article support : **Edwardson et al., BMJ 2022**, essai **cluster randomisé à trois bras** chez des employés de bureau du Royaume-Uni.

- 78 clusters et 756 salariés ;
- 3 bras : SWAL, SWAL + desk, contrôle ;
- critère principal : **temps assis quotidien à 12 mois** ;
- hypothèses de dimensionnement de l'article :
 - différence pertinente $\Delta = 60$ minutes ;

- écart-type sigma = 90 minutes;
- ICC = 0,05;
- taille moyenne de cluster 20;
- inflation supplémentaire de 30%.

Exercice contextualisé

L'article réel dimensionne un **essai en cluster** avec ICC, taille de cluster et inflation pour pertes.

La question d'examen vous demande un exercice de simplification :

Si l'essai avait randomisé des sujets et non des clusters, quel aurait été le nombre de sujets à inclure ?

Vous devez donc :

- enlever tout ce qui relève du **cluster**;
- garder les hypothèses de base sur delta et sigma;
- tenir compte du fait qu'il y a **3 bras** et une **majoration de 30%**.

Ce qu'il faut aller chercher dans l'article

- différence cliniquement pertinente : 60 minutes de temps assis;
- écart-type : 90 minutes;
- puissance : >90%;
- alpha bilatéral : 5%;
- article à 3 bras;
- inflation prévue de 30% pour pertes et non-compliance.

Ce que l'examineur veut vérifier

- savez-vous transformer un essai cluster en un **calcul individuel simple**;
- savez-vous ne **pas** réinjecter l'ICC alors que l'énoncé vous demande explicitement de l'oublier;
- savez-vous ensuite ré-élargir au vrai essai en tenant compte du **nombre de bras** et de l'attrition.

Étape 1 : calcul individuel de base

Le corrigé retient :

Formule de départ :

$$n_{\text{par bras}} = \frac{2(z_{1-\alpha/2} + z_{1-\beta})^2 \sigma^2}{\delta^2}$$

Comme $z_{1-\alpha/2} = 1,96$ et $z_{1-\beta} = 1,28$, on approxime :

$$(z_{\alpha} + z_{\beta})^2 \approx 10,5$$

Application numérique :

$$n_{\text{individuel}} \approx \frac{10,5 \times 2 \times 90^2}{60^2}$$

$$n_{\text{individuel}} \approx 47,25$$

soit environ **48**, arrondi à **50**.

Étape 2 : revenir au nombre total de sujets dans l'essai

Comme l'essai comporte 3 bras :

Formule de départ :

$$N_{\text{total}} = n_{\text{par bras}} \times 3$$

$$50 \times 3 = 150$$

Puis on applique la majoration de 30% annoncée dans le papier :

Formule de départ :

$$N_{\text{majoré}} = N_{\text{base}} \times 1,30$$

$$150 \times 1,30 = 195$$

Donc on retient **environ 200 sujets au total**.

Multiplicité

Le corrigé précise qu'une correction simple de multiplicité de type Bonferroni augmenterait l'effectif, mais pas de façon spectaculaire ici.

Le cœur de la question reste surtout :

- enlever l'effet cluster;
- garder de Δ et σ ;
- réintégrer ensuite le **nombre de bras** et la **marge de sécurité de 30%**.

Réponse type

L'article réel est un essai en cluster à trois bras, mais la question demande de raisonner comme s'il s'agissait d'un **essai individuel**. On enlève donc l'ICC et tout l'habillage cluster, puis on fait un calcul classique à deux moyennes avec $\sigma = 90$ et $\Delta = 60$, ce qui donne environ 48 sujets, soit 50 par groupe après arrondi. Comme l'essai comporte 3 bras, cela fait 150 sujets au total, puis 195 après majoration de 30%. On peut donc répondre qu'un essai individuel aurait nécessité **environ 200 sujets au total**.

6.13. 2024 RC

Article et contexte

Article support : **Moncrieff et al., Lancet Psychiatry 2023**, essai randomisé comparant **réduction progressive d'antipsychotique** versus **maintien** chez des patients ayant une psychose récurrente.

- essai ouvert, en groupes parallèles;
- suivi de 24 mois;
- critère principal : **social functioning (SFS)**;
- critère secondaire majeur : **severe relapse**;
- l'article dit qu'il fallait :
 - 206 patients pour détecter 4 points de différence sur le SFS;

- 372 patients pour un raisonnement de **non-infériorité** excluant une différence de 10% sur les rechutes sévères.

Exercice contextualisé

Cette annale est subtile, car le papier réel et la formule de cours ne parlent pas exactement la même langue.

Le papier :

- a un **critère principal continu** (SFS);
- ajoute un raisonnement séparé de **non-infériorité** sur les rechutes sévères.

L'examen vous demande :

Essayez, dans la mesure du possible, de retrouver le nombre de sujets nécessaires pour le test de l'hypothèse de non-infériorité.

Il faut donc reconnaître que l'exercice vise surtout un **ordre de grandeur pédagogique** autour de la marge de non-infériorité.

Ce qu'il faut aller chercher dans l'article

- puissance : 90%;
- marge de non-infériorité : 10 points;
- ordre de grandeur annoncé par l'article : 372 sujets;
- issue clinique concernée : **severe relapse**.

Ce que le corrigé de l'annale fait réellement

Le corrigé transforme le problème en une version très simplifiée de cours, du type **deux proportions en non-infériorité**, avec :

- un facteur global $(z_\alpha + z_\beta)^2 \approx 10$;
- un risque de base pris à 10%;
- une différence admissible de 10 points.

Autrement dit, on ne reconstruit pas toute la stratégie statistique du papier ; on essaie plutôt de **retomber sur le bon ordre de grandeur**.

Calcul simplifié du corrigé

Formule de départ, dans la version très simplifiée utilisée par le corrigé :

$$N \approx \frac{2(z_\alpha + z_\beta)^2 [p_C(1 - p_C) + p_T(1 - p_T)]}{\Delta^2}$$

Ici, le corrigé prend approximativement :

- $(z_\alpha + z_\beta)^2 \approx 10$;
- $p_C = 0,10$;
- $p_T = 0,10$;
- $\Delta = 0,10$.

Application numérique :

$$N \approx \frac{2 \times 10 \times [0,10(1 - 0,10) + 0,10(1 - 0,10)]}{(0,10)^2}$$

$$N \approx \frac{20 \times (0,09 + 0,09)}{0,01}$$

$$N \approx \frac{20 \times 0,18}{0,01} = 360$$

On retrouve donc **environ 360 sujets au total**, ce qui est très proche des 372 cités dans l'article.

Le message méthodologique à faire apparaître

En non-infériorité :

- la **marge acceptable** est le paramètre clé ;
- on cherche à montrer que la stratégie test n'est pas **trop mauvaise** par rapport au contrôle ;
- l'ordre de grandeur du NSN dépend fortement de cette marge.

Réponse type

Le papier annonce un raisonnement de non-infériorité sur les rechutes sévères avec une marge de 10%, conduisant à un besoin d'environ 372 sujets. Le corrigé de l'annale propose une reconstruction simplifiée avec la formule de cours pour deux proportions, en prenant un risque de base autour de 10% et une différence admissible de 10 points. On retrouve alors environ 360 sujets au total, ce qui est cohérent avec l'ordre de grandeur donné par l'article. La vraie idée à mettre en avant est que, dans un calcul de non-infériorité, **la marge cliniquement acceptable structure tout le dimensionnement**.

6.14. 2025 RC

Article et contexte

Article support : **Hautzinger et al., JAMA Psychiatry 2024**, essai randomisé comparant deux psychothérapies adjuvantes pour **prévenir la rechute dans le trouble bipolaire euthymique** :

- SEKT : thérapie cognitivo-comportementale structurée ;
- FEST : thérapie de soutien centrée sur le patient.

L'essai est :

- randomisé ;
- en aveugle pour les évaluateurs ;
- centré sur le **temps jusqu'à la rechute** ;
- conduit sur 18 mois de traitement/suivi, avec une vraie section **Sample size calculation** dans l'article.

Exercice contextualisé

Ici, la meilleure stratégie n'est pas de bricoler une formule de mémoire.

Le bon exercice consiste à lire la section **Sample size calculation** et à savoir la traduire proprement dans une copie.

Autrement dit :

1. repérer le type de test utilisé ;
2. lister les hypothèses temporelles et cliniques ;
3. donner directement le nombre de sujets obtenu par les auteurs.

Ce qu'il faut aller chercher dans l'article

La section de calcul d'effectif donne explicitement :

- un test du **log-rank bilatéral** ;
- une durée totale d'étude de 48 mois ;
- un recrutement pendant les 36 premiers mois ;
- 12% de dropouts dans chaque bras ;
- une amélioration de la proportion de patients **sans rechute** de 50% à 70% pendant la première année ;
- une puissance de 80% ;
- un alpha bilatéral de 5%.

Ce que l'examineur veut vérifier

- savez-vous reconnaître une logique de **survie / temps jusqu'à événement** ;
- savez-vous ne pas forcer inutilement une formule "deux proportions" ;
- savez-vous exploiter correctement un paragraphe de dimensionnement déjà rédigé dans le papier.

Conclusion chiffrée

L'article dit explicitement :

$$N = 300 \text{ sujets au total} = 150 \text{ par bras}$$

L'essai a ensuite effectivement randomisé 305 participants, ce qui confirme que l'ordre de grandeur est exactement le bon.

Réponse type

La bonne réponse consiste surtout à exploiter la section "Sample size calculation" de l'article. Les auteurs utilisent un **test du log-rank bilatéral**, avec une durée totale d'étude de 48 mois, un recrutement sur 36 mois, 12% de dropouts dans chaque bras, et l'hypothèse qu'en première année la proportion de patients sans rechute passerait de 50% à 70%. Sous ces hypothèses, ils concluent qu'il faut **300 sujets au total**, soit **150 par bras**. C'est la valeur à rapporter en priorité.

7. Réponses prêtes à l'emploi par grand type de question

7.1. Si l'on vous demande un vrai NSN

Il s'agit d'un calcul de NSN [pour deux proportions / deux moyennes / un essai de survie]. Il me faut donc un alpha, une puissance, une taille d'effet cliniquement pertinente, et une mesure de la variabilité ou du risque de base. En reprenant les hypothèses de l'article, on obtient un effectif d'environ [x] par bras, à majorer si nécessaire pour [dropout / cluster / multiplicité / autres].

7.2. Si l'on vous demande une puissance a posteriori

Une puissance a posteriori n'est pas l'outil le plus satisfaisant car l'intervalle de confiance contient déjà l'information de précision. Si on force néanmoins un calcul grossier, on remonte

d'abord au SE par SE $\approx \frac{\text{largeur}(IC95\%)}{3,92}$ ou, pour un RR / OR / HR, par SE $(\log \theta) \approx \frac{[\log(\text{borne sup}) - \log(\text{borne inf})]}{3,92}$. On en déduit ensuite un z et une puissance observée approximative. L'important est surtout que l'IC95% [n'exclut pas / exclut] un effet cliniquement important.

7.3. Si l'article ne donne pas assez d'éléments

Le calcul exact n'est pas reconstituable car il manque [paramètre]. En revanche, la logique du dimensionnement est claire : il faut [liste des paramètres] et le résultat attendu est de l'ordre de [ordre de grandeur].

7.4. Si l'étude est une grande cohorte

Dans cette étude, la puissance n'est pas le vrai problème. Avec un effectif et un nombre d'événements aussi élevés, l'information est abondante. La discussion doit porter surtout sur la taille d'effet, les IC95%, et la validité causale.

8. Checklist finale avant de rendre la copie

Checklist de 20 secondes

- Ai-je bien identifié le design ?
- Ai-je dit si l'on parle de NSN prospectif, de puissance observée, ou de simple précision ?
- Ai-je cité α , β , l'effet attendu, et la variabilité ?
- Ai-je pensé aux pertes, au cluster, à la multiplicité, à la surdispersion, ou aux co-primary endpoints ?
- Ai-je conclu par un chiffre ou par une interprétation claire ?
- Ai-je expliqué ce qui manque si le calcul exact n'est pas refaisable ?

9. Conclusion

Pour traiter correctement toutes les questions du dossier sur ce thème, il faut retenir trois idées.

1. Le bon raisonnement dépend du design.

On ne parle pas de la même manière d'un essai randomisé, d'un essai event-driven, d'une cohorte, ou d'une étude transversale.

2. Un bon calcul de NSN repose toujours sur les mêmes briques.

α , β , effet attendu, variabilité ou risque de base, puis ajustements de design.

3. Quand le calcul exact est impossible, une réponse méthodologiquement claire vaut mieux qu'un faux calcul précis.

En examen, cela rapporte des points parce que cela montre que vous avez compris la logique statistique réelle de la question.

Ce poly couvre ainsi l'ensemble des variantes rencontrées dans les annales du dossier, avec une progression allant des formules générales vers les cas d'examen réellement posés.