

Critères de jugement - Calcul d'effectif

Mercier Gwennaëlle
Soins intensifs Hôpital Erasme
Université Libre de Bruxelles-Belgique

Plan

Critères de jugement

- Pourquoi un critère de jugement ?
- Définition
- Qualité des critères de jugement
- Exemples

Taille d'échantillon

- Éléments à prendre en considération
- Type d'erreur
- Exemples

Critères de jugement : pourquoi ?

Différentes étapes de l'essai
thérapeutique :

Contexte scientifique (littérature, biblio)



Question de recherche



Hypothèse



Objectif principal



Définition du Critère de jugement

Exemple:

✓ Hypothèse :

La douleur diminue significativement pour les patients faisant partie du groupe d'intervention (avec l'utilisation du nouveau protocole de gestion de la douleur)

✓ Objectif principal:

Evaluer l'effet de l'utilisation d'un nouveau protocole de gestion de la douleur chez les patients USI.

Définition

- Indispensable dans les essais cliniques randomisés (utilisé même dans une étude observationnelle)
- **Critère de jugement principal** = variable choisie pour évaluer l'effet de l'intervention à l'étude. Va permettre de conclure sur l'objectif principal de l'étude
- Variable (également appelé variable réponse ou outcome) pour laquelle les groupes sont comparés.
- Nécessité de juger l'effet d'une intervention de façon mesurable afin de pouvoir réaliser des calculs statistiques (significativité d'une différence, intensité d'amélioration ...)

Exemple

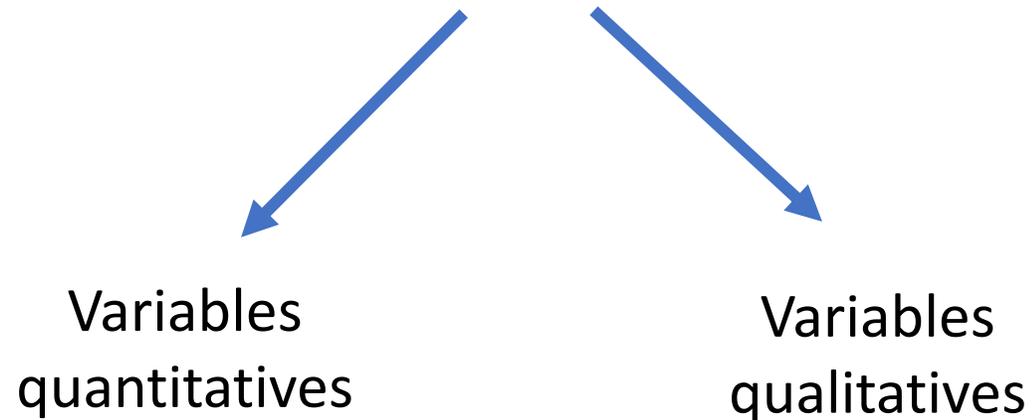
Définition peu précise	Définition plus précise
Fièvre Hyperéosinophilie diarrhée	Fièvre > 39° pendant 3 jours Éosinophilie > 1000c/mm ³ > 3 selles liquides par jour
Le patient a plus ou moins mal	Variation de l'EVA
Amélioration de l'escarre	Diminution ou augmentation de la taille de l'escarre (en cm ²)
Sommeil	Nombre d'heures (sur base d'un EEG)

- Utiliser une définition consensuelle plutôt qu'originale mais non reconnue par la communauté

Définition : variables

Une variable = une caractéristique susceptible de prendre une valeur différente selon les individus (ou les unités statistiques)

Ex : taille d'une personne, couleur des cheveux, durée d'une maladie, dosage sanguin...



Définition : variables quantitatives

✓ Définition :

Variable caractérisée par une valeur numérique

✓ 2 types de variables quantitatives :

➤ Variable quantitative continue : qui peut prendre n'importe quelle valeur numérique (il existe donc une infinité de valeurs possibles)

Rem : la précision va dépendre de l'instrument de mesure.

Exemple : poids, taille, taux de cholestérol, PAS ...

➤ Variable quantitative discrète : c'est un nombre entier, résultat d'un dénombrement

Exemple : nombre d'enfants (0 à ...), rappel de vaccin (0-1-2-3) ...

Définition : variables qualitatives

✓ Définition :

Variable qui n'a pas de valeur numérique, peut prendre différentes valeurs prédéfinies

Exemple : couleur des yeux, niveau d'étude, genre ...

✓ 3 types de variables qualitatives ou catégorielles :

➤ Variable qualitative nominale : catégorie non ordonnée

Exemple : groupe sanguin, état civil, nationalité ...

➤ Variable qualitative ordinale : classes ou catégories qui peuvent être ordonnées

Exemple : groupe d'âge, niveau d'étude ...

➤ Variable binaire ou dichotomique : type particulier d'une variable qualitative nominale qui ne peut prendre que 2 valeurs

Exemple : état de santé (malade/non-malade), décédé (oui/non), genre (M/F)...

Définition : différents types

➤ ***Critère de jugement principal :***

variable considérée comme la plus importante, doit être parfaitement décrit Et est utilisé dans le calcul de la taille d'échantillon

➤ ***Critères de jugement secondaires :***

évaluent les effets additionnels de l'intervention (effets non attendus de l'intervention, bénéfices secondaires, mécanismes d'action, effets indésirables...)

➤ ***Critères de jugement composite:***

critère de jugement qui prend en considération simultanément plusieurs événements cliniques

intérêt : mesure le rapport bénéfice/risque ou combiner des événements également pertinent (ex: infarctus – AVC – ischémie aiguë de membre)

Original Investigation | CARING FOR THE CRITICALLY ILL PATIENT

Effect of High-Dose Vitamin D₃ on Hospital Length of Stay in Critically Ill Patients With Vitamin D Deficiency The VITdAL-ICU Randomized Clinical Trial

Karin Amrein, MD, MSc; Christian Schnedl, MD; Alexander Holl, MD; Regina Riedl, MSc; Kenneth B. Christopher, MD; Christoph Pachler, MD;
Tadeja Urbanic Purkart, MD; Andreas Waltensdorfer, MD; Andreas Münch, MD; Helga Warnkross, MD; Tatjana Stojakovic, MD; Egbert Bisping, MD;
Wolfgang Toller, MD; Karl-Heinz Smolle, MD; Andrea Berghold, PhD; Thomas R. Pieber, MD; Harald Dobnig, MD

OBJECTIVE To investigate whether a vitamin D₃ treatment regimen intended to restore and maintain normal vitamin D status over 6 months is of health benefit for patients in ICUs.

DESIGN, SETTING, AND PARTICIPANTS A randomized double-blind, placebo-controlled, single-center trial, conducted from May 2010 through September 2012 at 5 ICUs that included a medical and surgical population of 492 critically ill adult white patients with vitamin D deficiency (≤ 20 ng/mL) assigned to receive either vitamin D₃ (n = 249) or a placebo (n = 243).

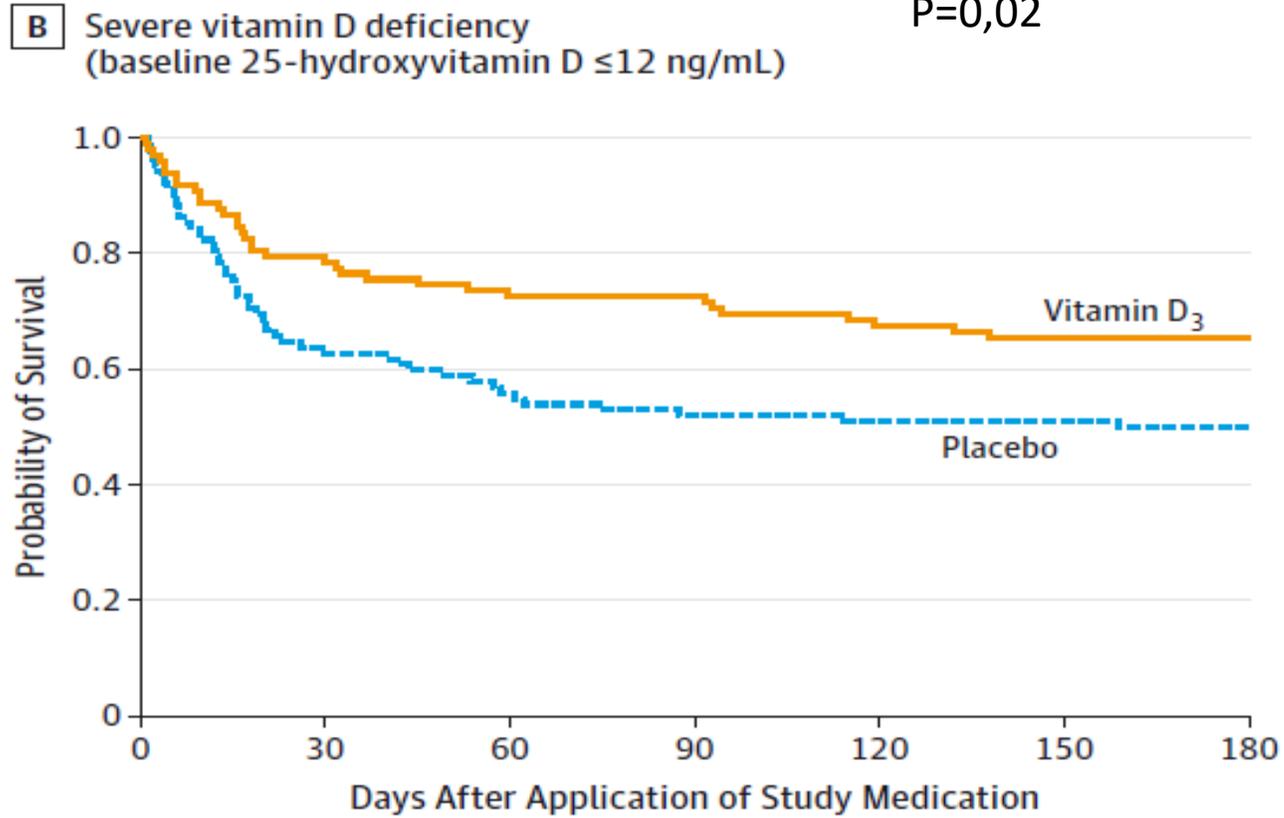
INTERVENTIONS Vitamin D₃ or placebo was given orally or via nasogastric tube once at a dose of 540 000 IU followed by monthly maintenance doses of 90 000 IU for 5 months.

MAIN OUTCOMES AND MEASURES The primary outcome was hospital length of stay.

Secondary outcomes included, among others, length of ICU stay, the percentage of patients with 25-hydroxyvitamin D levels higher than 30 ng/mL at day 7, hospital mortality, and 6-month mortality.

Exemple

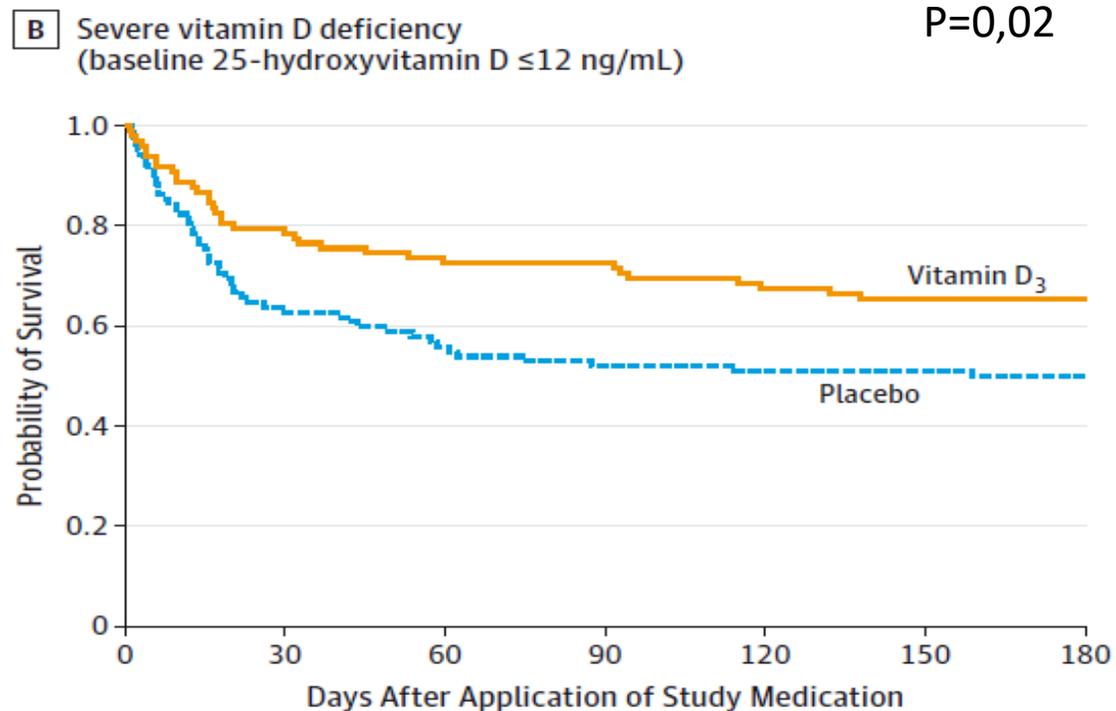
Figure 2. Survival Within 6 Months After Randomization



Bénéfice sur la mortalité ?

Exemple

Figure 2. Survival Within 6 Months After Randomization



Critère de jugement secondaire

Critère de jugement principal

	Placebo (n=238)	Vit D3 (n=237)	P-valeur
Durée de séjour hospitalière médiane (range)	19,3 (0,1-154,1)	20,1 (0,2-181)	0,98

Exemple

THE LANCET

Methylprednisolone in patients undergoing cardiopulmonary bypass (SIRS): a randomised, double-blind, placebo-controlled trial

*Richard PWhitlock, P J Devereaux, Kevin H Teoh, Andre Lamy, Jessica Vincent, Janice Pogue, Domenico Paparella, Daniel I Sessler, Ganesan Karthikeyan, Juan Carlos Villar, Yunxia Zuo, Álvaro Avezum, Mackenzie Quantz, Georgios I Tagarakis, Pallav J Shah, Seyed Hesameddin Abbasi, Hong Zheng, Shirley Pettit, Susan Chrolavicius, Salim Yusuf, for the SIRS Investigators**

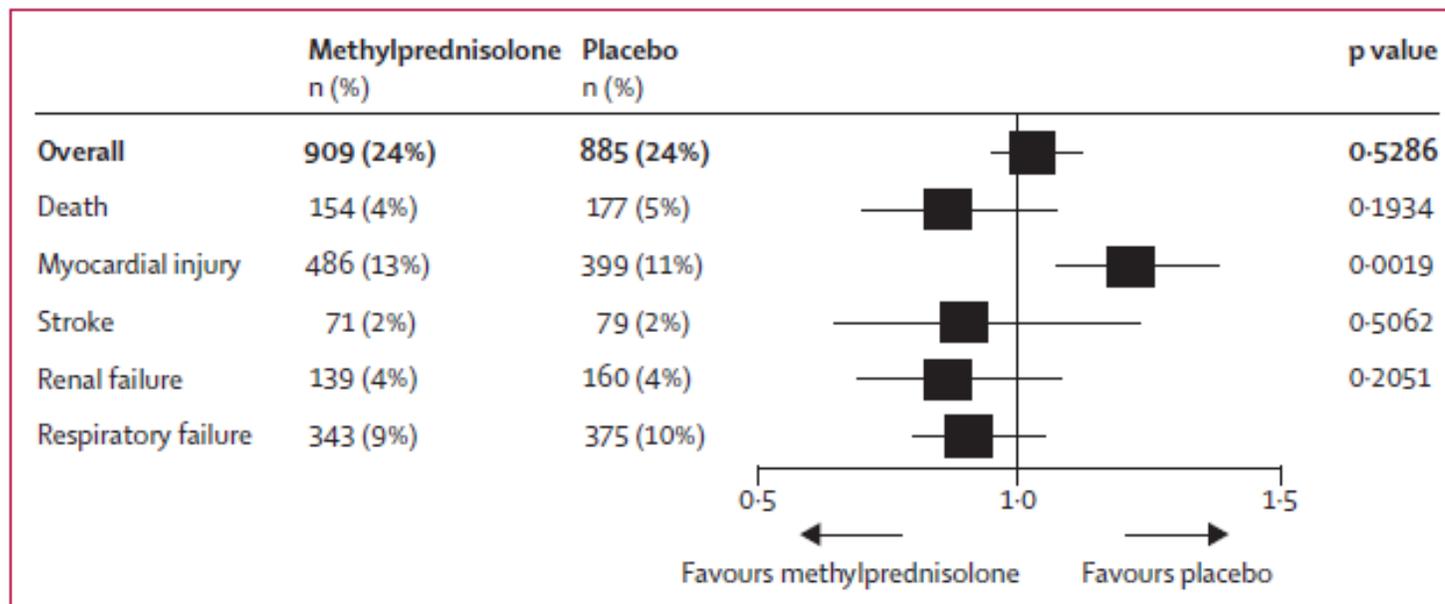
Primary outcomes were mortality at 30 days after randomization

a composite of death, myocardial injury, stroke, renal failure (stage 3 acute kidney injury, 2012 Kidney Disease Improving Global Outcomes [KDIGO] guidelines), or respiratory failure (uninterrupted postoperative mechanical ventilation for more than 48 h) at 30 days after randomisation.

Exemple

THE LANCET

	Methylprednisolone (n=3755)	Placebo (n=3752)	Relative risk (95% CI)	p value
Primary outcomes				
Death	154 (4%)	177 (5%)	0.87 (0.70-1.07)	0.19
Death, myocardial injury, stroke, new renal failure, or respiratory failure	909 (24%)	885 (24%)	1.03 (0.95-1.11)	0.53



Critère de jugement principal

Critères de jugement composite

Figure 2: Relative risk of the composite primary outcome and its components
The horizontal lines represent the 95% CI.

Exemple

✓ Critère de jugement principal :

La douleur du patient évaluée pendant une période déterminée 2 fois par shift (au début et à la fin de la prise en charge) à l'aide de l'EVA.

✓ Critères de jugement secondaires :

- Durée de séjour à l'USI
- Repos des patients
- Durée de la ventilation mécanique

Qualité des critères de jugement

Tenir compte des qualités spécifiques afin d'opérer un choix rationnel du critère de jugement

- **Pertinence** : réponse au problème posé
- **Signifiant** en termes cliniques (morbidity, AVC, épuration extra-rénale, mortalité, qualité de vie,...)
- **Objectif** : ex dosage sanguin et non bon ou mauvais
- **Unique** : ex : cicatrisation d'un ulcère duodénal

Qualité des critères de jugement

➤ Mesures des variables chez le patient →

Exactitude : donne la vraie valeur – justesse (ex : balance)

Précision : faible variabilité (si répétition de la mesure → donne le même résultat)

Reproductibilité : aptitude à produire la même valeur en cas de conditions identiques (variation biologique intra-individuelle)

Sensibilité : capacité d'un test à détecter les vrais malades

Spécificité : capacité d'un test à détecter les vrais non-malades

Qualité des critères de jugement

- Réalisation d'examens chez les sujets → *acceptables, aisés et peu coûteux*
- Ethique, pratique (non douloureux pour le patient) et important pour le patient.

Ex : concentration d'un ATB dans le LCR = refaire PL ...

Il ne faut pas choisir plusieurs critères de jugement principal car cela amène à des problèmes d'interprétation liés à de multiples analyses

Kathleen Puntillo
Shoshana R. Arai
Bruce A. Cooper
Nancy A. Stotts
Judith E. Nelson

A randomized clinical trial of an intervention to relieve thirst and dry mouth in intensive care unit patients

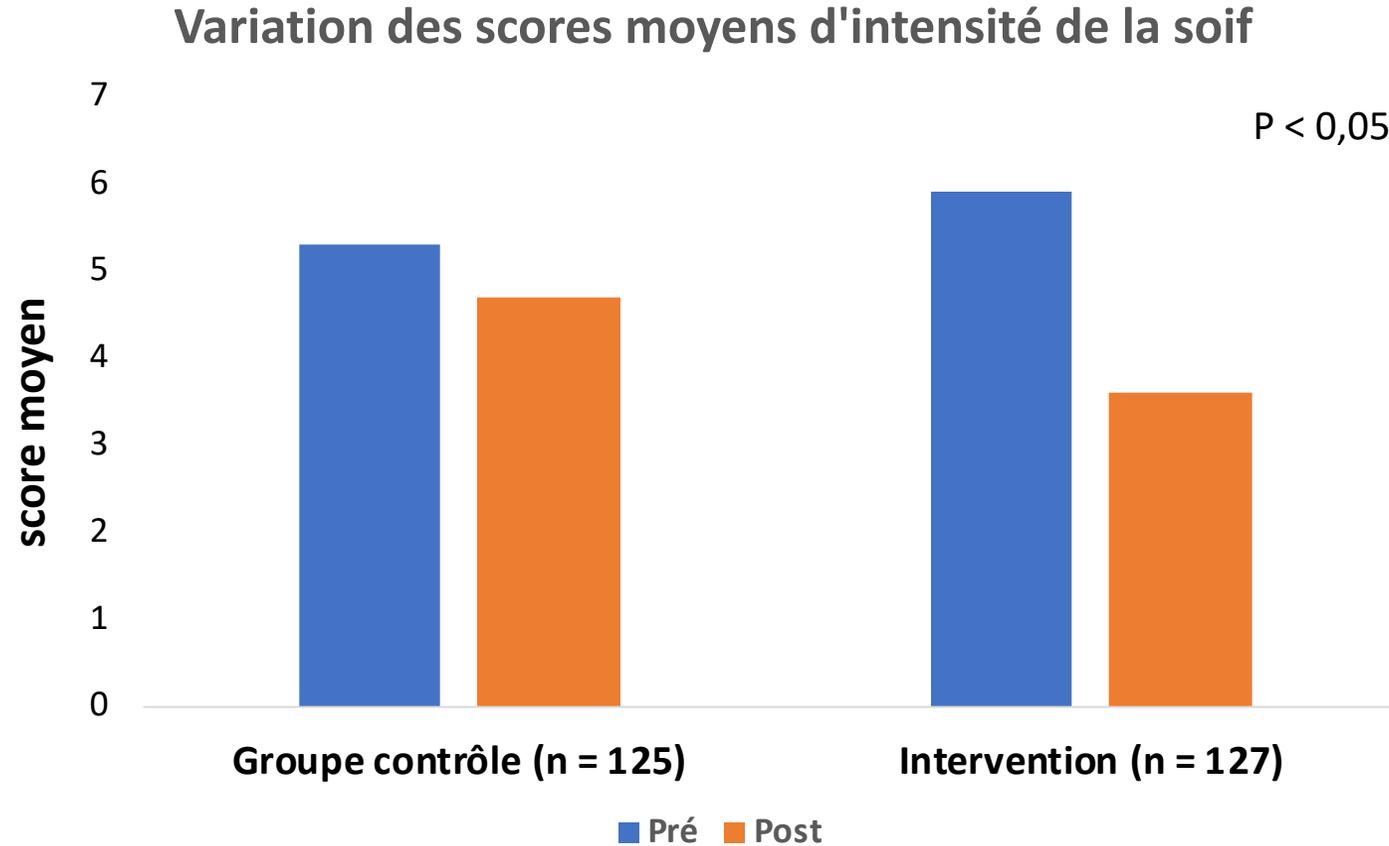
Objectif: To test an intervention bundle for thirst intensity, thirst distress, and dry mouth

Description du critère de jugement principal :

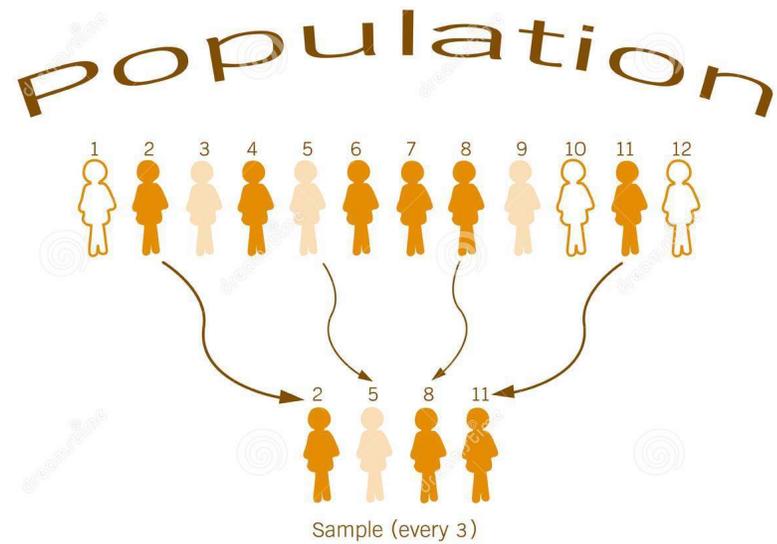
A research team nurse (RTN) used **NRSs (0–10)** to obtain patient thirst ratings at baseline and during subsequent assessments. The NRS has been widely used to assess pain intensity and other symptoms; has face, construct, and concurrent validity [32]; and is **feasible for use in ICUs**. To measure TI, the RTN asked the patient, “how intense is your thirst, on this scale, where 0 = no thirst and 10 = worst possible thirst?”

For patients unable to communicate orally owing to endotracheal intubation, the RTN held up a laminated sheet showing the NRS and/or asked the patient to nod affirmatively when she pointed to the number that corresponded to their TI or TD. Independent scoring by two RTNs for 50 assessments was initially compared to assure inter-rater **reliability**

Exemple



Taille d'échantillon



Taille d'échantillon

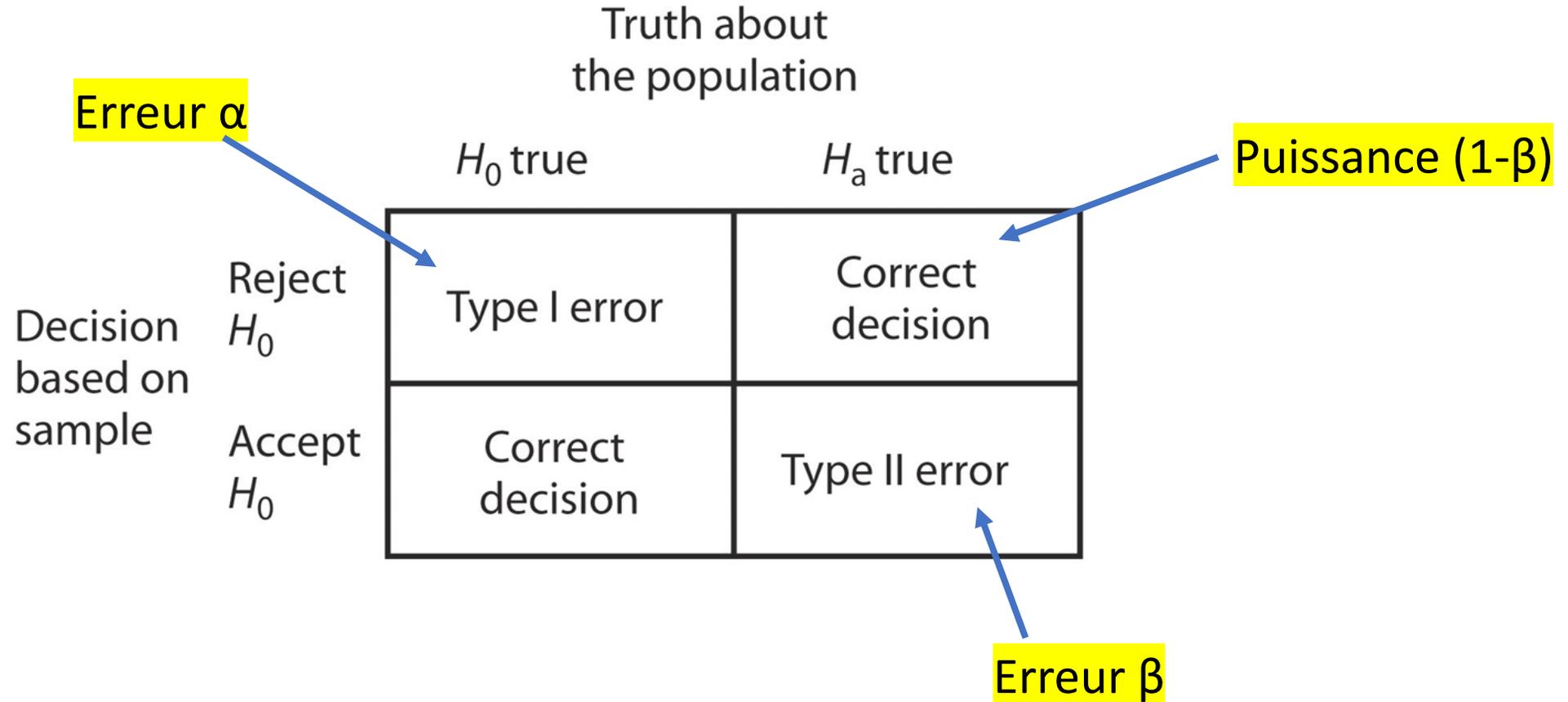
- Dans les essais cliniques randomisés le nombre de sujets à inclure (n) se définit à l'avance
- Le n doit être assez large afin d'obtenir une haute probabilité de détecter une différence statistiquement significative = puissance de l'essai, par convention > 80% (1-risque de conclure à l'absence d'effet alors qu'il y en a un)
- A effectif constant la puissance est d'autant plus grande que la différence entre les groupes est importante
- Si effet attendu est faible (petites différences) → grande taille d'échantillon
- Se calcule sur base du critère de jugement principal

Eléments à prendre en compte

Le calcul de la taille d'échantillon doit tenir compte de :

- L'effet attendu dans chaque groupe (différences attendues)
- L'erreur α (type I) = 5%
- La puissance statistique : 1- erreur β (type II), > 80%
- La fréquence de l'évènement dans le groupe contrôle et pour les variables quantitatives continues = déviation standard (dispersion)
- Souvent on augmente la taille d'échantillon pour faire face aux perdus de vue, examens non réalisés etc.

Type d'erreur



Exemple

Fréquence évènement dans le groupe contrôle ou moy (ds)

Statistical Analysis

Sample size calculation was performed using the Mann-Whitney test. It was based on 2008 routine data collected by medical and surgical ICUs participating in our trial from patients who stayed 48 hours or more. A mean length of hospital stay of 14 days with a standard deviation of 7 days was inferred. The effect size of a 2-day shorter hospital stay was set arbitrarily and was delineated as what we decided was the smallest unit of time that would be of benefit to the patient. The number of patients needed for the study was calculated to be 468 at a 2-sided significance level of $\alpha = .05$ and 80% power. Including a drop-out rate of 5%, 490 patients needed to be randomized.

Effet attendu

Risque α et puissance

Nombre de patients

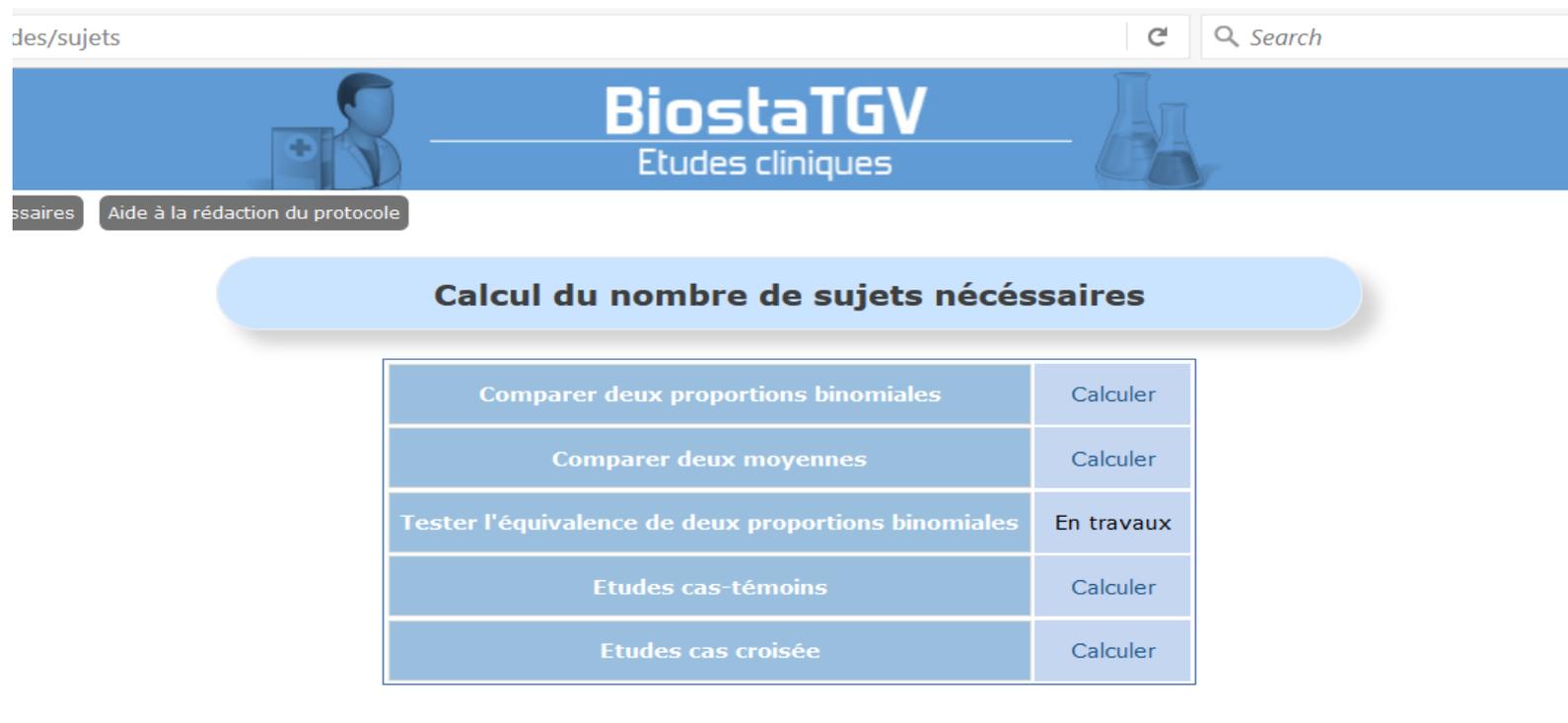
Exemple

- ✓ **L'effet attendu** dans chaque groupe : diminution de l'EVA de 2
- ✓ **L'erreur α (type I) = 5%**
- ✓ **La puissance statistique = 80%**
- ✓ **Fréquence dans le groupe contrôle:**
 - Variables quantitatives continues : étude descriptive : EVA moyen = 6 (± 2)
 - Variable dichotomique: groupe avec EVA > 4 versus EVA \leq 4 avec fréquence chez le groupe contrôle = 60%
- ✓ **Ajoute 5% de sujets**

Calcul de la taille d'échantillon

Calculator free online

<http://marne.u707.jussieu.fr/biostatgv/?module=etudes/sujets>

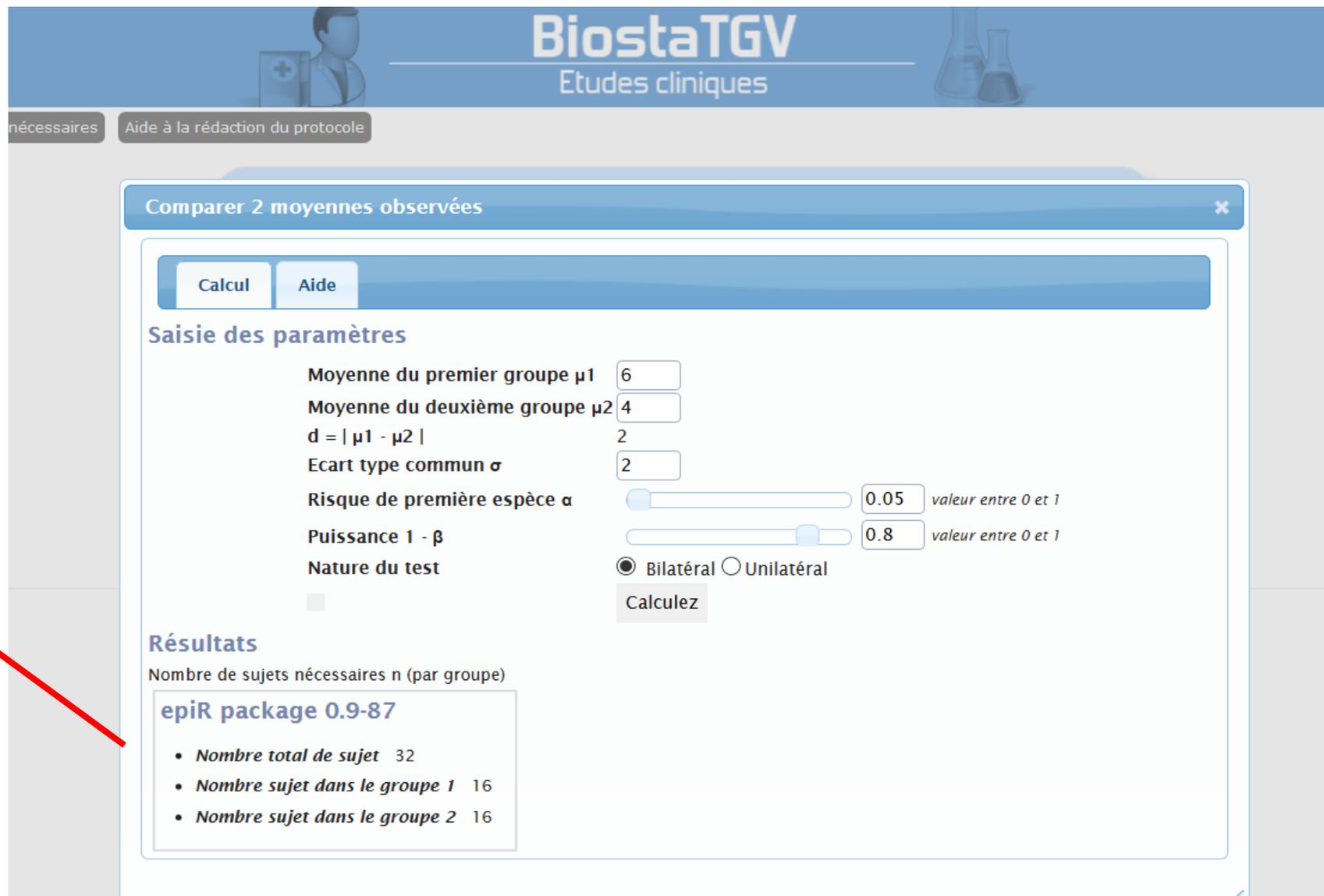


The screenshot shows the BiostaTGV website interface. At the top, there is a navigation bar with the text 'des/sujets' on the left and a search bar on the right. Below this is a blue header with the 'BiostaTGV' logo and the text 'Etudes cliniques'. Underneath the header, there are two buttons: 'essaires' and 'Aide à la rédaction du protocole'. The main content area features a light blue rounded rectangle with the title 'Calcul du nombre de sujets nécessaires'. Below this title is a table with five rows, each representing a different type of study and a corresponding action button.

Calcul du nombre de sujets nécessaires	
Comparer deux proportions binomiales	Calculer
Comparer deux moyennes	Calculer
Tester l'équivalence de deux proportions binomiales	En travaux
Etudes cas-témoins	Calculer
Etudes cas croisée	Calculer

Calcul de la taille d'échantillon

Variables quantitatives : comparaison de 2 moyennes



The screenshot shows the 'BiostaTGV Etudes cliniques' interface. The main window is titled 'Comparer 2 moyennes observées'. It has two tabs: 'Calcul' (selected) and 'Aide'. Under 'Saisie des paramètres', the following values are entered: Moyenne du premier groupe μ_1 = 6, Moyenne du deuxième groupe μ_2 = 4, $d = |\mu_1 - \mu_2|$ = 2, Ecart type commun σ = 2. The 'Risque de première espèce α ' is set to 0.05, and 'Puissance $1 - \beta$ ' is set to 0.8. The test is set to 'Bilatéral'. A 'Calculez' button is visible. The 'Résultats' section shows 'Nombre de sujets nécessaires n (par groupe)' and 'epiR package 0.9-87' with the following results:

- Nombre total de sujet 32
- Nombre sujet dans le groupe 1 16
- Nombre sujet dans le groupe 2 16

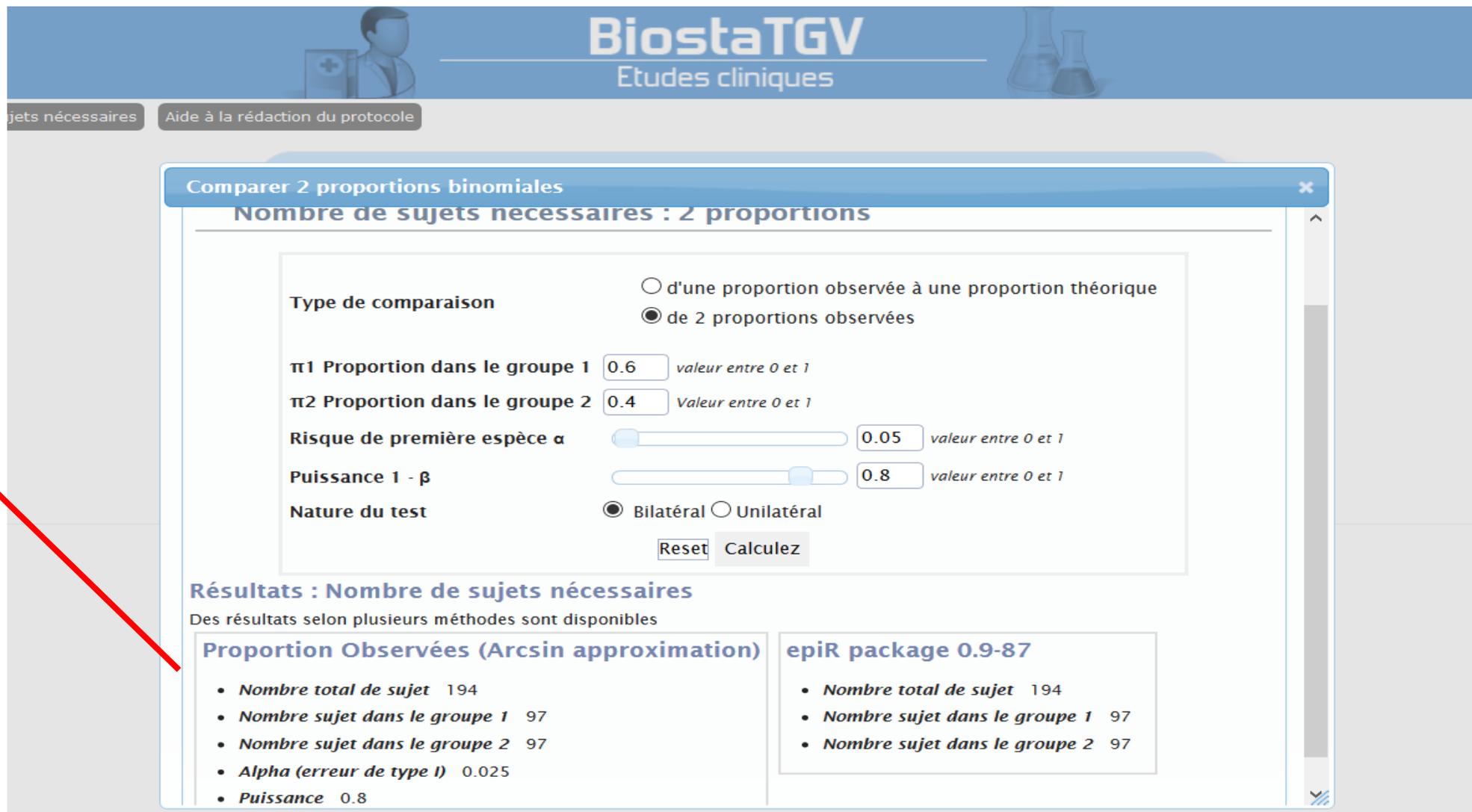
$n = 32 \text{ sujets} + 5\% =$

34 sujets au total



Calcul de la taille d'échantillon

Variables catégorielles : comparaison de 2 proportions



The screenshot shows the BiostaTGV software interface. The main window is titled "Comparer 2 proportions binomiales" and "Nombre de sujets nécessaires : 2 proportions". The interface includes input fields for the proportion in group 1 (0.6), group 2 (0.4), risk of first kind error (0.05), and power (0.8). The test is set to bilateral. The results section shows the number of subjects needed: 194 total, 97 in each group, with an alpha error of 0.025 and power of 0.8.

BiostaTGV
Etudes cliniques

Objets nécessaires Aide à la rédaction du protocole

Comparer 2 proportions binomiales

Nombre de sujets nécessaires : 2 proportions

Type de comparaison

- d'une proportion observée à une proportion théorique
- de 2 proportions observées

π_1 Proportion dans le groupe 1 valeur entre 0 et 1

π_2 Proportion dans le groupe 2 valeur entre 0 et 1

Risque de première espèce α 0.05 valeur entre 0 et 1

Puissance $1 - \beta$ 0.8 valeur entre 0 et 1

Nature du test

- Bilatéral
- Unilatéral

Résultats : Nombre de sujets nécessaires

Des résultats selon plusieurs méthodes sont disponibles

Proportion Observées (Arcsin approximation)

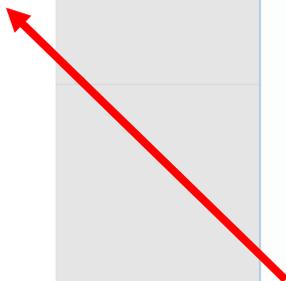
- Nombre total de sujet 194
- Nombre sujet dans le groupe 1 97
- Nombre sujet dans le groupe 2 97
- Alpha (erreur de type I) 0.025
- Puissance 0.8

epiR package 0.9-87

- Nombre total de sujet 194
- Nombre sujet dans le groupe 1 97
- Nombre sujet dans le groupe 2 97

$n = 194 + 5\% =$

204 sujets au total



Conclusion

- ✓ Permet de répondre à l'objectif principal
- ✓ Indispensable à définir de manière très précise
- ✓ Choisir le critère de jugement en tenant compte de ces différentes caractéristiques
- ✓ Va être utilisé pour calculer la taille de l'échantillon

Merci pour votre attention