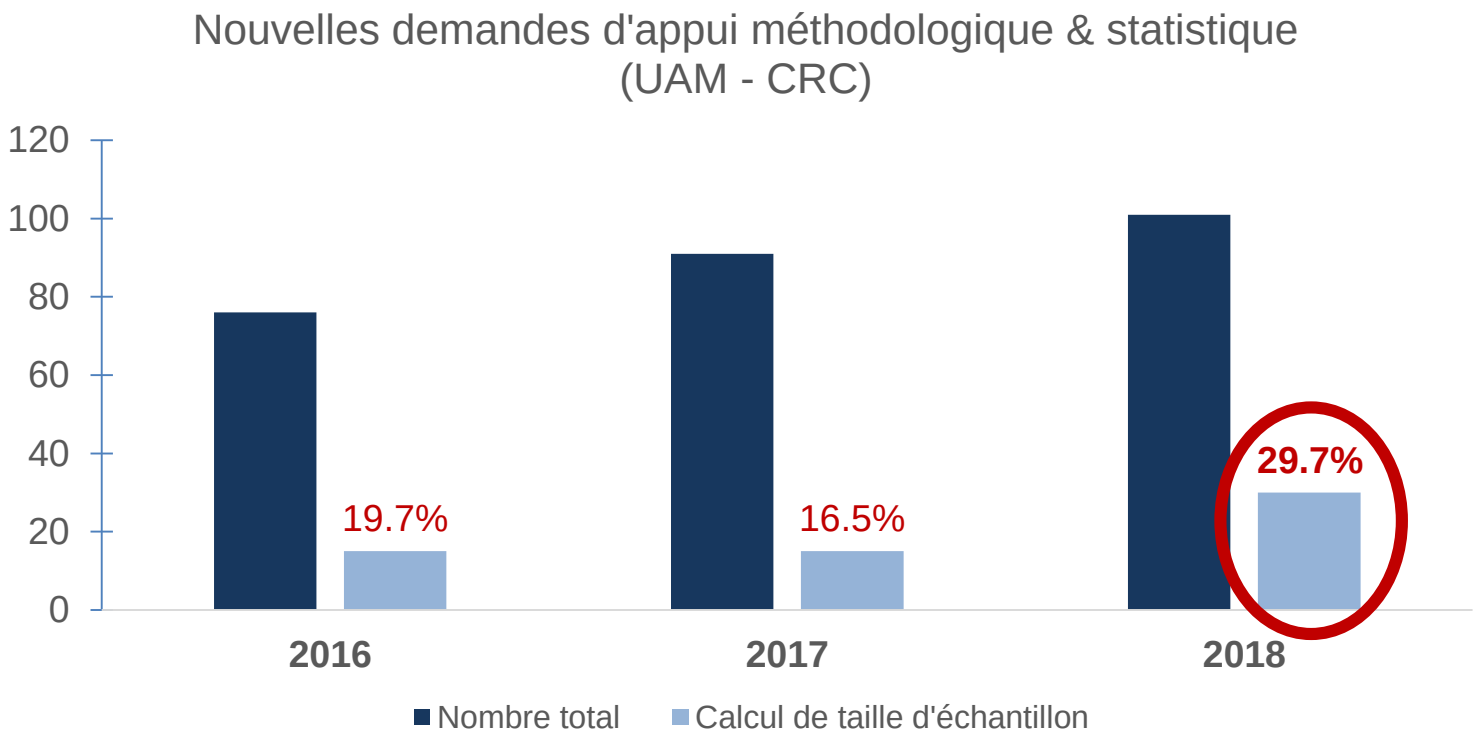
Puissance et taille d'échantillon

Quels sont les points essentiels à considérer?

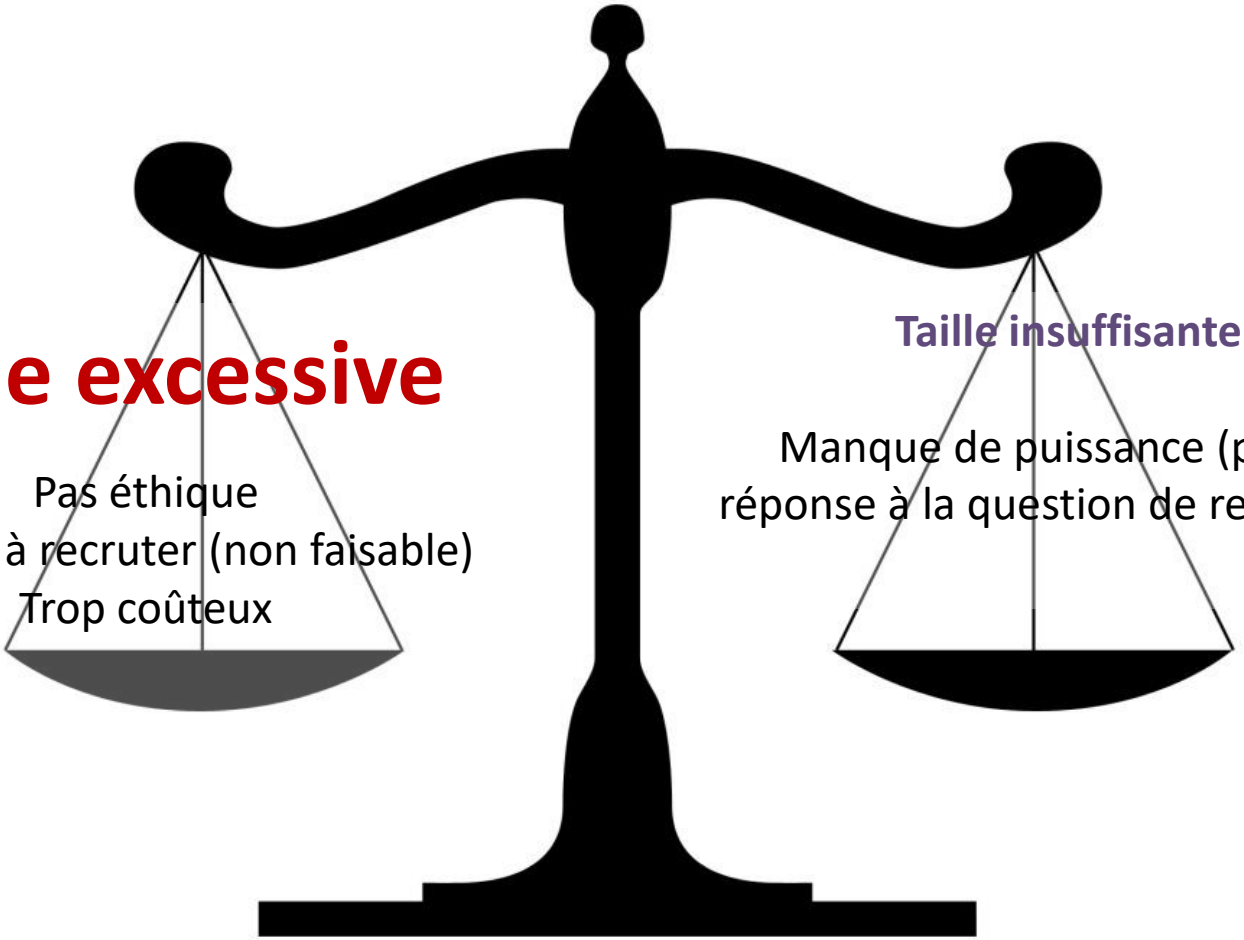
Angèle GAYET-AGERON, MD, PhD
Elise DUPUIS-LOZERON, PhD
Unité d'Appui Méthodologique - CRC
Service d'épidémiologie clinique

- Calcul de taille d'échantillon : demande fréquente à l'UAM



- Avant de démarrer toute étude clinique (expérimentale comme observationnelle)
- Partie importante du protocole de recherche (« Méthodes statistiques »)
  - Permet de **formaliser les hypothèses de recherche**
  - Inclut une **justification du choix de l'*outcome* principal**
  - Doit inclure une **justification du choix des valeurs de *baseline*** (avant de monter l'étude = citer la littérature ou des résultats d'étude pilote)
  - Doit inclure **une justification du choix de la différence** minimale à détecter (ou de la marge de non-infériorité)
  - Permet de juger de la **faisabilité** de l'étude, de la nécessité ou pas de mener une étude multicentrique

- La **BONNE** taille d'échantillon est un équilibre entre…



Sample size and statistical power of randomised, controlled trials in orthopaedics

K. B. Freedman, S. Back, J. Bernstein

From the University of Pennsylvania School of Medicine, Philadelphia, USA

We reviewed all 717 manuscripts published in the 1997 issues of the British and American volumes of the *Journal of Bone and Joint Surgery* and in *Clinical Orthopaedics and Related Research*, from which 33 randomised, controlled trials were identified. The results and sample sizes were used to calculate

Résultats « négatifs » (inconclusifs)

= manque de puissance

= taille d'échantillon trop petite

size and 12 (48%) lacked the power necessary to detect a large effect size. Of the 25 studies which did not have an adequate size of sample to detect small differences, the average used was only 10% of the required number

Our findings suggest that randomised, controlled trials in clinical orthopaedic research utilise sample sizes which are too small to ensure statistical significance for what may be clinically important results.

J Bone Joint Surg [Br] 2001;83-B:397-402.
Received 16 September 1999; Accepted after revision 15 June 2000

INTERNATIONAL CONFERENCE ON HARMONISATION OF TECHNICAL REQUIREMENTS FOR REGISTRATION OF PHARMACEUTICALS FOR HUMAN USE

ICH HARMONISED TRIPARTITE GUIDELINE

STATISTICAL PRINCIPLES FOR CLINICAL TRIALS
E9

Current *Step 4* version

dated 5 February 1998

3.5 Sample Size

The number of subjects in a clinical trial should always be large enough to provide a reliable answer to the questions addressed. This number is usually determined by the primary objective of the trial. If the sample size is determined on some other basis, then this should be made clear and justified. For example, a trial sized on the basis of safety questions or requirements or important secondary objectives may need larger numbers of subjects than a trial sized on the basis of the primary efficacy question (see, for example, ICH E1a).

Using the usual method for determining the appropriate sample size, the following items should be specified: a primary variable, the test statistic, the null hypothesis, the alternative ('working') hypothesis at the chosen dose(s) (embodying consideration of the treatment difference to be detected or rejected at the dose and in the subject population selected), the probability of erroneously rejecting the null hypothesis (the type I error), and the probability of erroneously failing to reject the null hypothesis (the type II error), as well as the approach to dealing with treatment withdrawals and protocol violations. In some instances, the event rate is of primary interest for evaluating power, and assumptions should be made to extrapolate from the required number of events to the eventual sample size for the trial.

- Tout part de la **question de recherche**
  - Répond à l'objectif principal de l'étude
  - Idéalement **UNE étude** répond à **UNE question** de recherche
  - Guider toute la structure du protocole d'étude
  - Format type (le fameux **PICO**)
    - A quelle **population** s'adresse l'étude ?
    - Qu'est-ce qui est évalué ? (nouvelle **intervention**, nouveau traitement, exposition à un facteur de risque…)
    - A quoi est-ce comparé ? (groupe de **comparaison**)
    - Sur quoi est évalué l'effet de l'intervention ? (***outcome*** principal)

Use of aspirin to reduce risk of initial vascular events in patients at moderate risk of cardiovascular disease (ARRIVE): a randomised, double-blind, placebo-controlled trial

J Michael Gaziano, Carlos Brotons, Rosa Coppolecchia, Claudio Cricelli, Harald Darius, Philip B Gorelick, George Howard, Thomas A Pearson, Peter M Rothwell, Luis Miguel Ruilope, Michal Tendera, Gianni Tognoni; the ARRIVE Executive Committee

Lancet 2018; 392: 1036–46

Summary

Background The use of aspirin in the primary prevention of cardiovascular events remains controversial. We aimed to assess the efficacy and safety of aspirin versus placebo in patients with a moderate estimated risk of a first cardiovascular event.

Population : Patients présentant un risque cardiovasculaire modéré

Intervention : Aspirine

Comparateur : Placebo

Outcome : 1er évènement cardiovasculaire

(principal)

Outcome principal

- Homonymes
  - **Mesure d'issue** (essai randomisé et contrôlé)
  - **Critère de jugement** (essai randomisé et contrôlé)
  - **Cas** (étude cas-témoins)
  - **Évènement** (étude de cohorte prospective)
  - …
- Variable essentielle dans une étude clinique
  - **Ce que l'on cherche à étudier/expliquer** par l'étude
  - Oriente sur le choix du ***design*** d'étude
  - Justifie l'**importance** de l'étude (*outcome* clinique versus paraclinique, logistique, économique…)
  - Conditionne la **faisabilité** de l'étude (taille d'étude, coût…)

Use of aspirin to reduce risk of initial vascular events in patients at moderate risk of cardiovascular disease (ARRIVE): a randomised, double-blind, placebo-controlled trial

J Michael Gaziano, Carlos Brotons, Rosa Coppolecchia, Claudio Cricelli, Harald Darius, Philip B Gorelick, George Howard, Thomas A Pearson, Peter M Rothwell, Luis Miguel Ruilope, Michal Tendera, Gianni Tognoni; the ARRIVE Executive Committee

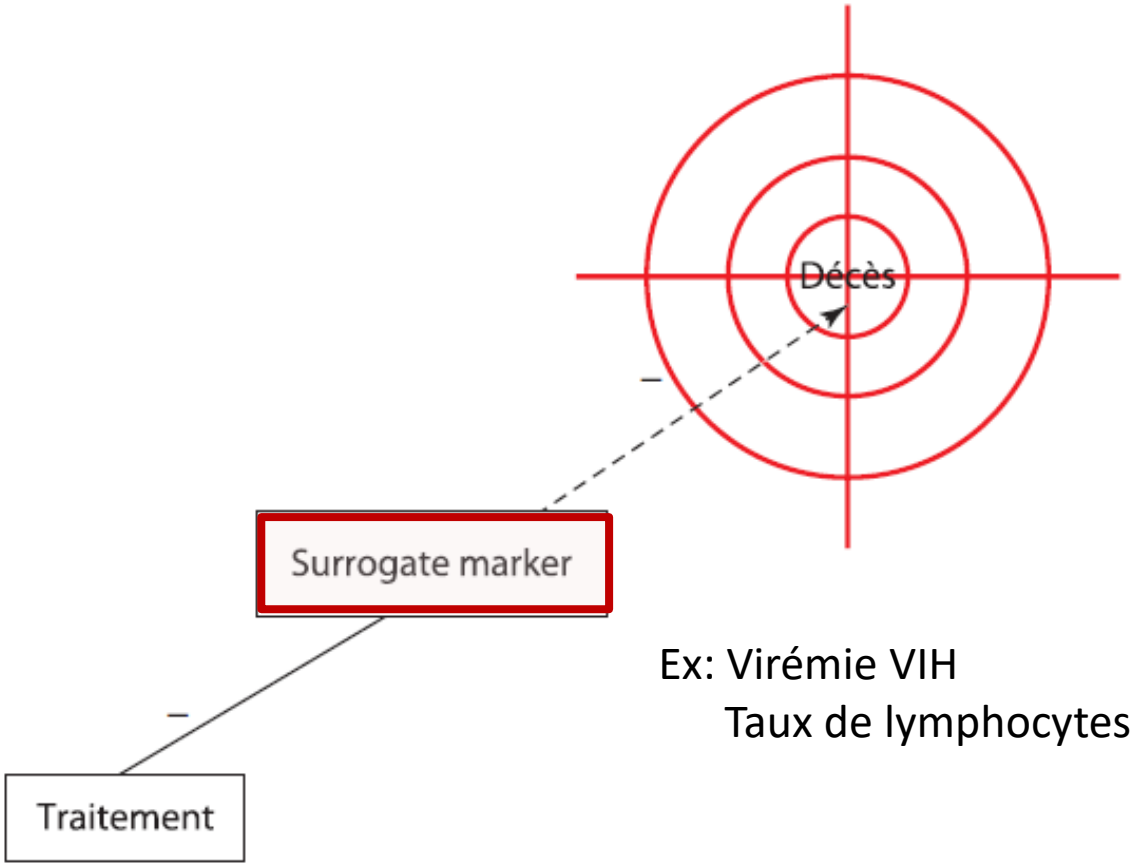
Lancet 2018; 392: 1036–46

Methods ARRIVE is a randomised, double-blind, placebo-controlled, multicentre study done in seven countries. Eligible patients were aged 55 years (men) or 60 years (women) and older and had an average cardiovascular risk, deemed to be moderate on the basis of the number of specific risk factors. We excluded patients at high risk of gastrointestinal bleeding or other bleeding, or diabetes. Patients were randomly assigned (1:1) with a computer-generated randomisation code to receive enteric-coated aspirin tablets (100 mg) or placebo tablets, once daily. Patients, investigators, and others involved in treatment or data analysis were masked to treatment allocation. The primary efficacy endpoint was a composite outcome of time to first occurrence of cardiovascular death, myocardial infarction, unstable angina, stroke, or transient ischaemic attack. Safety endpoints were haemorrhagic events and incidence of other adverse events, and were analysed in the intention-to-treat population. This study is registered with ClinicalTrials.gov, number NCT00501059.

Quel *outcome* principal ?

- Selon la perspective du patient
  - ***Patient-related outcomes***
  - Importance clinique dans la situation médicale étudiée
  - Ce qui compte du point de vue du patient
    - *Outcome* **subjectif** (douleur, qualité de vie …)
    - *Outcome* **objectif** (décès, guérison clinique, fonctionnalité d'une articulation sur une échelle, autonomie…)

- Situations particulières
  - ***Outcome* composite** (combinaison d'*outcomes*, évènements rares)
  - ***Surrogate markers*** (marqueurs intermédiaires associés à un *outcome* clinique rare difficile à étudier)



Ex: Virémie VIH
Taux de lymphocytes CD4

Figure 1. Lien entre traitement, *surrogate marker* et événement d'intérêt.

Extrait du livre «L'épidémiologie expliquée aux cliniciens», Edition Médecine & Hygiène 2018

Quel *outcome* principal ?

- Selon la perspective du patient
  - ***Patient-related outcomes***
  - Importance clinique dans la situation médicale étudiée
  - Ce qui compte du point de vue du patient
    - *Outcome* subjectif (douleur, fonctionnalité, qualité de vie, autonomie…)
    - *Outcome* objectif (décès, guérison clinique, fonctionnalité, autonomie…)

- Situations particulières
  - ***Outcome* composite** (combinaison d'*outcomes*, évènements rares)
  - ***Surrogate markers*** (marqueurs intermédiaires associés à un *outcome* clinique rare difficile à étudier)

- Selon une **perspective de santé publique** ou **économique**
  - Exemples: prévalence de souches bactériennes résistantes, durée d'hospitalisation, coûts médicaux directs…

3.5 Sample Size

The number of subjects in a clinical trial should always be large enough to provide a reliable answer to the questions addressed. This number is usually determined by the primary objective of the trial. If the sample size is determined on some other basis, then this should be made clear and justified. For example, a trial sized on the basis of safety questions or requirements or important secondary objectives may need larger numbers of subjects than a trial sized on the basis of the primary efficacy question (see, for example, ICH E1a).

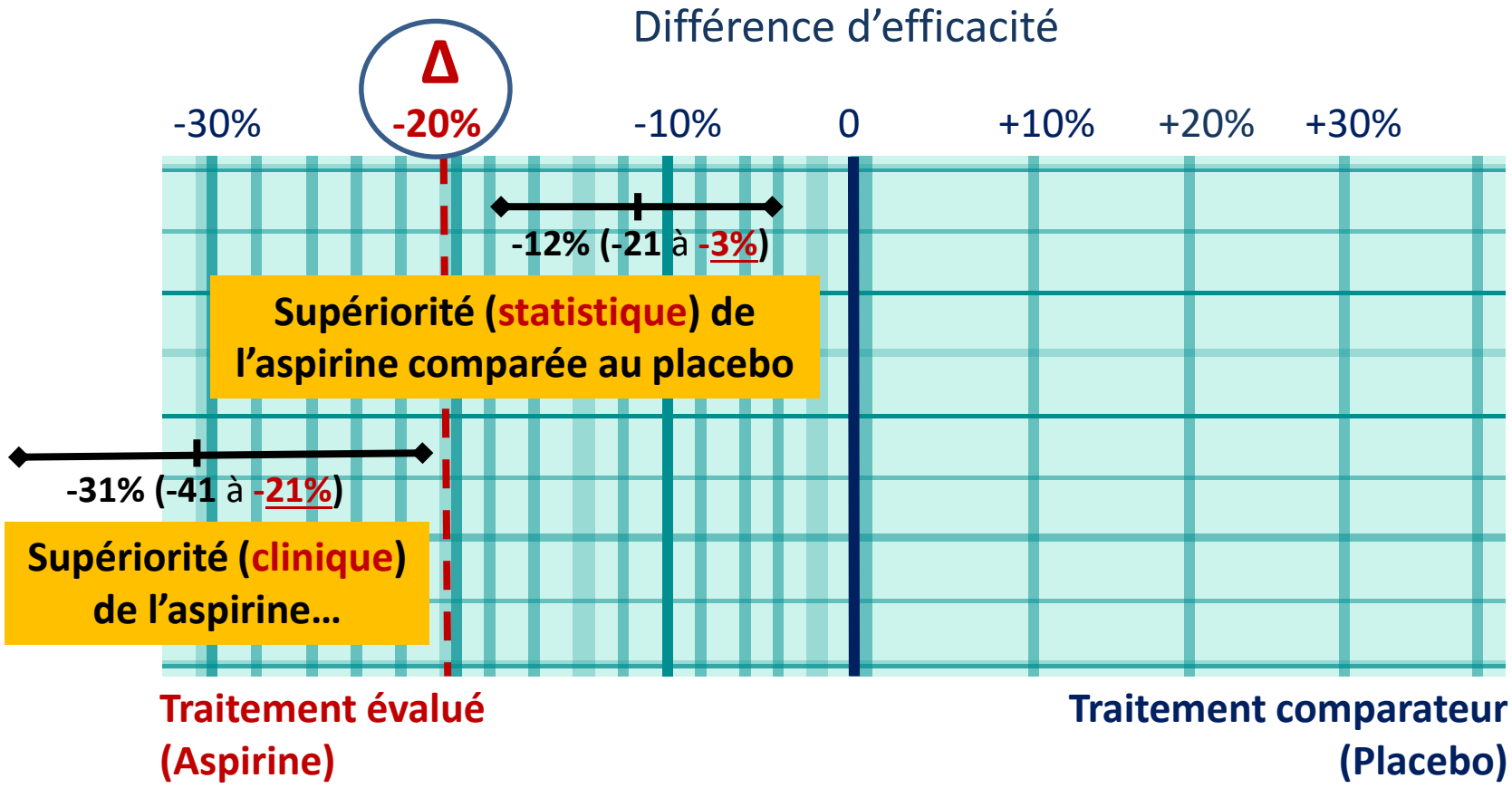
Using the usual method for determining the appropriate sample size, the following items should be specified: a primary variable, the test statistic, the null hypothesis, the alternative ('working') hypothesis at the chosen dose(s) (embodying consideration of the treatment difference to be detected or rejected at the dose and in the subject population selected), the probability of erroneously rejecting the null hypothesis (the type I error), and the probability of erroneously failing to reject the null hypothesis (the type II error), as well as the approach to dealing with treatment withdrawals and protocol violations. In some instances, the event rate is of primary interest for evaluating power, and assumptions should be made to extrapolate from the required number of events to the eventual sample size for the trial.

- Ex. essai clinique randomisé et contrôlé de supériorité

- Prédéfinir quelle est **la plus petite différence à détecter Δ** pour conclure à l'efficacité (supériorité) du nouveau traitement comparé au traitement standard (placebo) ?
  - Choix de la valeur basé sur des **considérations cliniques** et non statistiques (comité d'experts, avis des patients…), doit être **justifiée**
  - Souvent une **différence (de moyennes ou de proportions)** entre les deux groupes comparés
  - Alternativement, une **différence relative** (RR, HR…)
  - Parfois un multiple de l'écart-type des observations obtenues (**différence moyenne standardisée**, ex. 0.5 écart-type = moyen)

Les investigateurs ont défini **une différence de 20% comme cliniquement importante** en terme de réduction des évènements cardiovasculaires avec le traitement par aspirine comparé au placebo (données non réalistes)

Plus petite différence à détecter



Quels ingrédients sont nécessaires au calcul ?

- En plus de la **plus petite différence à détecter Δ**
  1. **Valeur de l'*outcome* dans le groupe de comparaison** (*outcome* binaire)
  2. **Variabilité des observations** (*outcome* continu)
     - Ecart-type de la différence à détecter ou écart-types des mesures d'*outcomes* dans les 2 groupes comparés
     - Souvent grande inconnue (cf. données publiées, étude pilote…)
  3. **Ratio** de patients dans chaque bras d'étude (en général 1:1)
  4. Erreur de type 1 (**α**) : en général 5% parfois 1% (bilatéral)
  5. Erreur de type 2 (β) ou puissance d'étude (**1-β**) : en général maximum 20% (ou minimum 80% de puissance)


DETERMINATION
SAMPLE
SIZE
DEVIATION
INDEPENDENT
ESTIMATOR
COMPONENT
APPROXIMATION
ALLOCATION
SMALLESTERRORPARTLY
EXPERIMENTAL
POPULATION
INFERENCES
INVERSELY
TABLE
TREATMENT
NORMAL
POWER
TOTAL
RESOURCE
PATHOGEN
DISTRIBUTION
STRATIFIED
ANIMALS
BINOMIAL
UNKNOWN
EQUATION
CUTOFF
DEGREES
MEANS
ESTIMATE
STATISTICIANS
TRUE
EQUAL
LARGE
SAMPLING
TRIAL
REJECT
CUMULATIVE
PROPORTIONAL
FORM
STANDARD
LABORATORY
APPROXIMATESIZE
THEOREM
COMPLICATED
STATISTICAL
HYPOTHESIS
OBSERVATIONS
UNITS
STRATIFICATION
VALUES
WIDE
STATISTICS
DIVIDED
DATA
NUMBER
EXPECTED
NK
SUPPORT
CONTROL
REPLICATES
FUNCTION
VARIANCE
PROBABILITY
CONFIDENCE
PROPORTION
MINUS
GROUP
CANDIDATE
TARGET
SHOWN
STRATUM
IDENTICALLY
INDIVIDUALS
COLLECT
PARAMETER
QUANTITATIVELY