

Critère d'évaluation/ Taille de l'étude

Geneviève Chêne

17 mars 2009

Abidjan

Atelier de formation à la recherche clinique

Plan du cours

- Conception
- Composantes
- Principes du calcul de la taille d'étude
- Cas particuliers
 - Critère composite
 - Critères intermédiaires/substitution
- Synthèse/conclusion
- Bibliographie

Définition d'un critère de jugement (1)

Traduction d'un évènement d'intérêt
en une **variable** permettant de
répondre à la question de l'étude

- directement reliée à la question de l'étude
- fondée sur un évènement d'intérêt

Définition d'un critère de jugement (2)

Question de l'étude



Évènement d'intérêt



Variable = Critère de jugement

Élaboration d'un critère de jugement : exemple 1

question de l'étude



évènement d'intérêt



critère de jugement

efficacité anticancéreux ?



décès



% patients vivant à la
fin de l'étude

Élaboration d'un critère de jugement : exemple 2

question de l'étude



évènement d'intérêt



critère de jugement

efficacité substitut
nicotinique ?



statut tabagique



% patients abstinents
à la fin de l'étude

Formulation d'un critère de jugement

- Évènement d'intérêt mesuré à l'échelon de l'individu
- Critère de jugement exprimé au niveau de l'échantillon (le groupe décrit ou les groupes comparés)
- Une question = un objectif = un critère de jugement

Formulation d'un critère de jugement 3 éléments

- Temps de mesure pertinent
- Définition du (des) diagnostics qui le composent
- Modalités pratiques des tests ou procédures permettant de le mesurer
- Exemple : essai comparant traitement chirurgical / médicament pour la prévention des AVC
 - « proportion de patients ayant présenté un accident vasculaire cérébral constitué au cours de la 1ère année suivant l'inclusion dans l'essai et systématiquement confirmé dans les 48h suivant les 1ers symptômes par un scanner cérébral »

Conditions essentielles d'un critère de jugement (1)

- Pertinence clinique
 - ce qu'il mesure fait une réelle différence dans la vie du patient ou sa prise en charge
 - Exemples
 - anticancéreux : survie, rechute
 - régime alimentaire : poids, glycémie, cholestérolémie, survie, complications (cancer, maladies cardiovasculaires) ...

Conditions essentielles d'un critère de jugement (2)

- **Fiabilité**
 - la mesure de l'évènement d'intérêt doit être répétable et reproductible
 - exemple : survie, poids, cholestérolémie...
- **Validité**
 - L'évènement d'intérêt doit être mesuré sans biais

Calcul de la taille de l'essai

- Objectif
 - définir *a priori* un nombre de sujets nécessaire pour pouvoir mettre en évidence une certaine différence entre les traitements
 - il existe une **différence minimale** dont on ne veut pas laisser échapper la mise en évidence grâce à l'essai
 - notion de **puissance statistique satisfaisante**
- Importance de cette différence
 - pertinence clinique
 - **différence clinique** et non différence statistique
 - connaissance approfondie de la littérature scientifique
- Calcul en fonction
 - **type de critère jugement**
 - **du test statistique**

- Principe du test statistique

- H_0 : il n'y pas de différence entre les traitements comparés
 H_a : il existe une différence
- Raisonnement par **l'absurde** : on suppose que H_0 est vérifiée et on cherche si l'on peut infirmer cette hypothèse
- Ne permet pas de démontrer que H_0 est vraie ou fausse
- Indique si H_0 est vraisemblable
- Conclusions : **modérées par un risque de se tromper** (risque d'erreur)

Les risques α et β

Réalité

Décision prise

	Absence de différence / liaison	Existence d'une différence / liaison
H₀	Bonne décision	Erreur de 2 ^{ème} espèce (risque β)
H_a	Erreur de 1 ^{ère} espèce (risque α)	Bonne décision (puissance $1-\beta$)

Calcul de la taille de l'essai

- Cas d'une variable quantitative continue
 - Test statistique : . comparaison de 2 moyennes
. test de Student

$$n = 2 \frac{\sigma^2}{\Delta^2} (U_{\alpha} + U_{2\beta})^2$$

n= taille de chaque groupe (identique)
 σ = écart-type de la variable dans la pop
 Δ =différence attendue

- Le nombre de sujet est d'autant plus grand que
 - variabilité du critère de jugement est grande (σ)
 - Δ est petit
 - risque α est faible
 - risque β est faible (et donc puissance $1-\beta$ est grande)

Calcul de la taille de l'essai

- Cas d'une variable quantitative continue

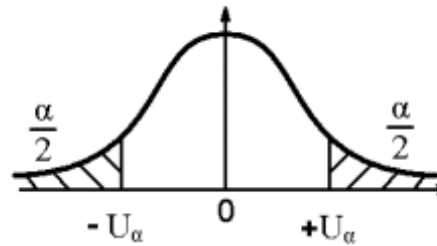
- Exple : vitamine D vs placebo chez femmes enceintes pour prévenir hypocalcémie néonatale

- Calcémie moyenne femmes non traitées : 9,0 mg/100 ml
- **Δ intéressant** : augmentation moyenne de **0,5 mg/100 ml** au moins
- **$\sigma = 1$ mg/100 ml**
- **$\alpha = 5\%$, $\beta = 10\%$**

$$n = 2 \frac{\sigma^2}{\Delta^2} (U_{\alpha} + U_{2\beta})^2 = n = 2 \frac{1^2}{0,5^2} (1,96 + 1,282)^2 = 84 \text{ par groupe}$$

Table de la loi normale centrée réduite

α ou β
↓



α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	infini	2,5758	2,3263	2,1701	2,0537	1,9600	1,8808	1,8119	1,7507	1,6954
0,1	1,6449	1,5982	1,5548	1,5141	1,4758	1,4395	1,4051	1,3722	1,3408	1,3106
0,2	1,2816	1,2536	1,2265	1,2004	1,1750	1,1503	1,1264	1,1031	1,0803	1,0581
0,3	1,0364	1,0152	0,9945	0,9741	0,9542	0,9346	0,9154	0,8965	0,8779	0,8596
0,4	0,8416	0,8239	0,8064	0,7892	0,7722	0,7554	0,7388	0,7225	0,7063	0,6903
0,5	0,6745	0,6588	0,6433	0,6280	0,6128	0,5978	0,5828	0,5681	0,5534	0,5388
0,6	0,5244	0,5101	0,4958	0,4817	0,4677	0,4538	0,4399	0,4261	0,4125	0,3989
0,7	0,3853	0,3719	0,3585	0,3451	0,3319	0,3186	0,3055	0,2924	0,2793	0,2663
0,8	0,2533	0,2404	0,2275	0,2147	0,2019	0,1891	0,1764	0,1637	0,1510	0,1383
0,9	0,1257	0,1130	0,1004	0,0878	0,0753	0,0627	0,0502	0,0376	0,0251	0,0125

α	0,001	0,0001	0,00001	0,000001	0,0000001	0,00000001	0,000000001
U_α	3,2905	3,8906	4,4172	4,8916	5,3267	5,7307	6,1094

Calcul de la taille de l'essai

- **Cas d'une variable catégorielle dichotomique**

- Test statistique :
 - . comparaison de 2 proportions
 - . test du Chi2

$$n = \frac{\left(U_{\alpha} \sqrt{2\pi(1-\pi)} + U_{2\beta} \sqrt{\pi_N(1-\pi_N) + \pi_R(1-\pi_R)} \right)^2}{\Delta^2}$$

π_N = % attendu avec nouveau traitement

π_R = % attendu avec traitement de réf.

$\Delta = \pi_N - \pi_R$

π_N = moyenne de π_N et π_R

Calcul de la taille de l'essai

- Cas d'une variable catégorielle dichotomique

- Exple : Sclérose latérale amyotrophique

- étude nouveau traitement vs placebo
- survie : 10% à 18 mois
- $\Delta = 10\%$
- $\alpha=5\%$, $\beta=10\%$

$$n = \frac{\left(U_{\alpha} \sqrt{2\pi(1-\pi)} + U_{2\beta} \sqrt{\pi_N(1-\pi_N) + \pi_R(1-\pi_R)} \right)^2}{\Delta^2}$$
$$= \frac{\left(1,96 \sqrt{2 \times 0,15(1-0,15)} + 1,282 \sqrt{0,1(1-0,1) + 0,2(1-0,2)} \right)^2}{0,1^2}$$
$$= 266 \text{ par groupe, } 532 \text{ au total}$$

Définition d'un critère de jugement composite

Traduction de **plusieurs évènements** d'intérêt en **une variable** permettant de répondre à la question de l'étude

- variable regroupant plusieurs évènements
- directement reliée à la question de l'étude
- basée sur un évènement d'intérêt

Définition d'un critère de jugement composite

Question de l'étude



Plusieurs évènements d'intérêt



Variable = Critère de jugement composite

Critère de jugement composite

Exemple 1

question de l'étude

efficacité anticancéreux ?



Plusieurs évènements
d'intérêt

décès, rechute



critère de jugement
composite

% patients décédés ou
ayant rechuté à la fin
de l'étude

Critère de jugement composite

Exemple 2

question de l'étude

efficacité antirétrovirale ?



Plusieurs évènements
d'intérêt



Décès, stade SIDA



critère de jugement
composite



% patients décédés ou
étant passé au stade
SIDA au cours du suivi

Critères essentiels à l'élaboration d'un CDJ composite (1)

- Pertinence clinique
 - chaque évènement mesuré doit représenter une réelle différence dans la vie du patient ou dans sa prise en charge
- Fiabilité
 - la mesure de chaque évènement d'intérêt doit être répétable et reproductible
- Validité
 - chaque évènement d'intérêt doit être mesuré sans biais

Critères essentiels à l'élaboration d'un CDJ composite (2)

- Le critère de jugement composite doit, dans son ensemble, être :
 - pertinent cliniquement
 - fiable
 - Valide
- Cohérence des évènements du critère de jugement composite
 - Tous les composants du CDJ composite doivent aller dans le même sens

Apports des critères de jugements composites

Évènements d'intérêt multiples

Augmentation de la quantité d'informations relatives à la pathologie étudiée

Réduction du nombre de sujets nécessaire

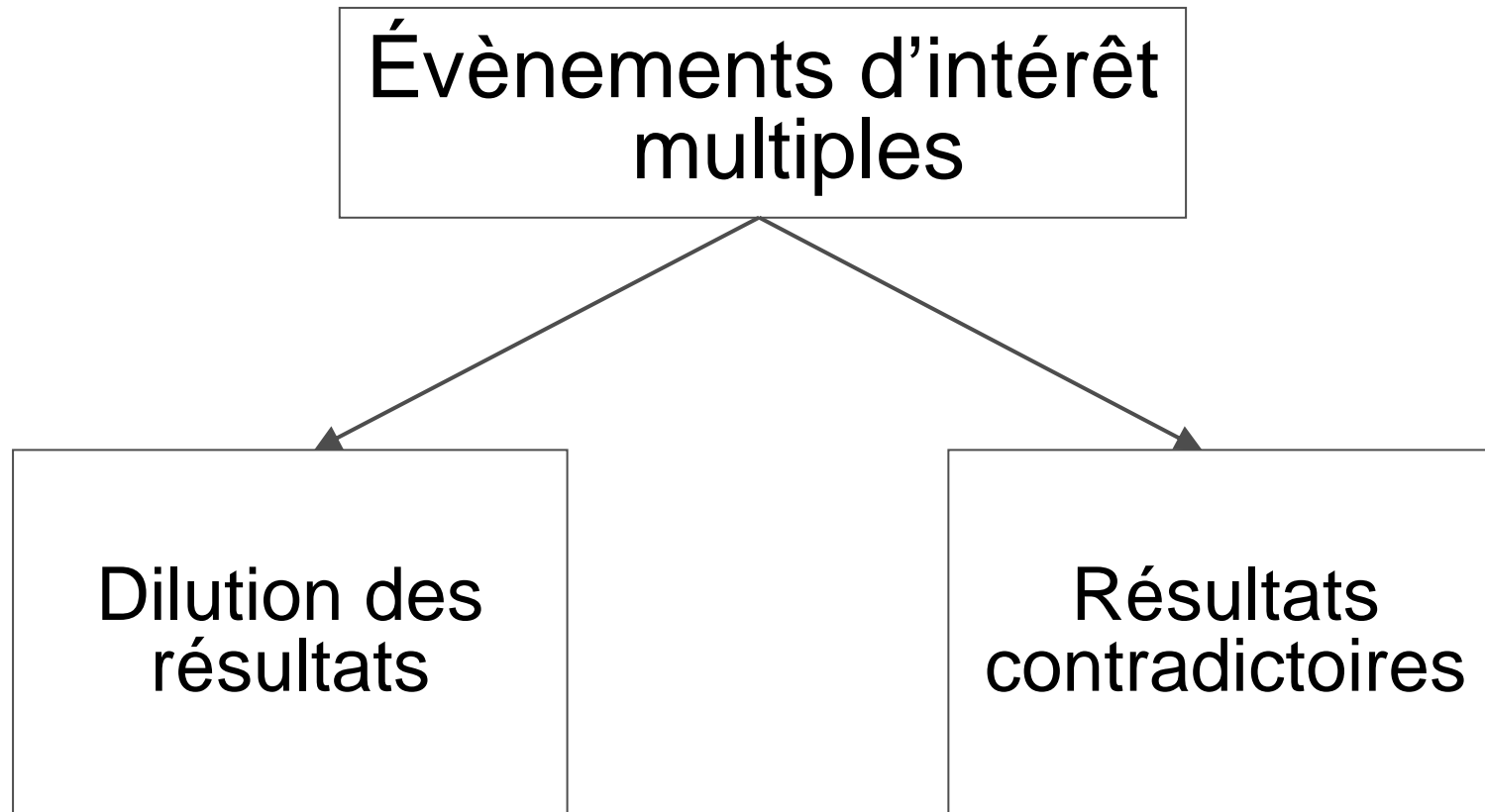
Augmentation de la quantité d'informations relatives à la pathologie étudiée

- Une maladie se résume rarement à un seul évènement clinique
- Un critère de jugement composite permet d'intégrer plusieurs aspects d'une même pathologie
- évaluation de l'effet global d'un traitement
- sans ajustement du risque alpha

Réduction du nombre de sujets nécessaire

- Exemple : analyse de survie. Délai de survenue d'évènements cliniques
- Puissance = 0,9
- Risque alpha = 0,05
- Risque relatif = 0,65
- Proportion d'évts = 20%
- N = 954
- Puissance = 0,9
- Risque alpha = 0,05
- Risque relatif = 0,65
- Proportion d'évts = 40%
- N = 496

Limites des critères de jugement composites



Problème de la dilution des résultats


Lorsqu'un évènement fréquent du CDJ composite n'est pas substantiellement modifié par l'effet du traitement


Problème de la dilution des résultats

Exemple

- Évaluation des effets du carvedilol (β bloquant) chez 1959 patients présentant une insuffisance ventriculaire gauche après infarctus du myocarde
- Critère de jugement composite
 - mortalité de toute cause
 - hospitalisation pour cardiopathie

Résultats	hospitalisation	mortalité	total
carvedilol	23%	12%	35%
contrôle	22%	15%	37%


P=0,03


NS

Problème des résultats contradictoires

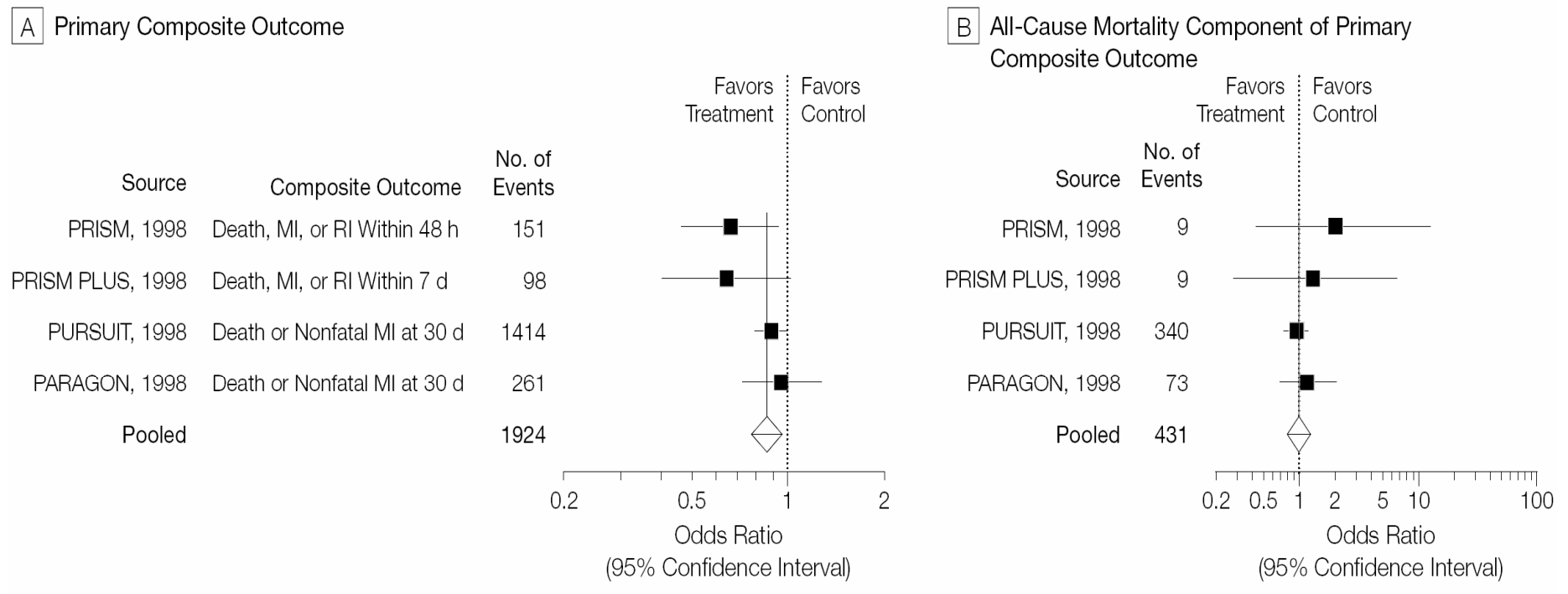
Lorsqu'un composant du CDJ composite met en évidence un effet bénéfique du traitement et qu'un autre met en évidence un effet délétère

→ problème d'interprétation

Problème des résultats contradictoires

Exemple

Figure. Glycoprotein IIb/IIIa Inhibitors in the Management of Acute Coronary Syndromes



Analyse des composants d'un CDJ composite

- Problème de la multiplicité des analyses statistiques
 - ajustement du risque alpha ou analyses secondaires
- Risques compétitifs
 - combinaison d'évènements fatals et non fatals
 - un traitement responsable d'une surmortalité va apparaître comme protecteur vis-à-vis des évènements non fatals
- Attention aux données incomplètes
 - arrêt du suivi après survenue d'un évènement non fatal

Recommandations

- Veiller à la pertinence, la fiabilité et la validité des composants d'un CDJ composite
- Ne pas arrêter le suivi d'un patient après la survenue du premier évènement d'intérêt
- Préciser ce qui concernera l'analyse principale et l'analyse secondaire
- l'analyse principale, sa conclusion et son interprétation doivent se faire par rapport au CDJ composite
- chacun des composants du CDJ composite doit être décrit

Critères de jugement de substitution

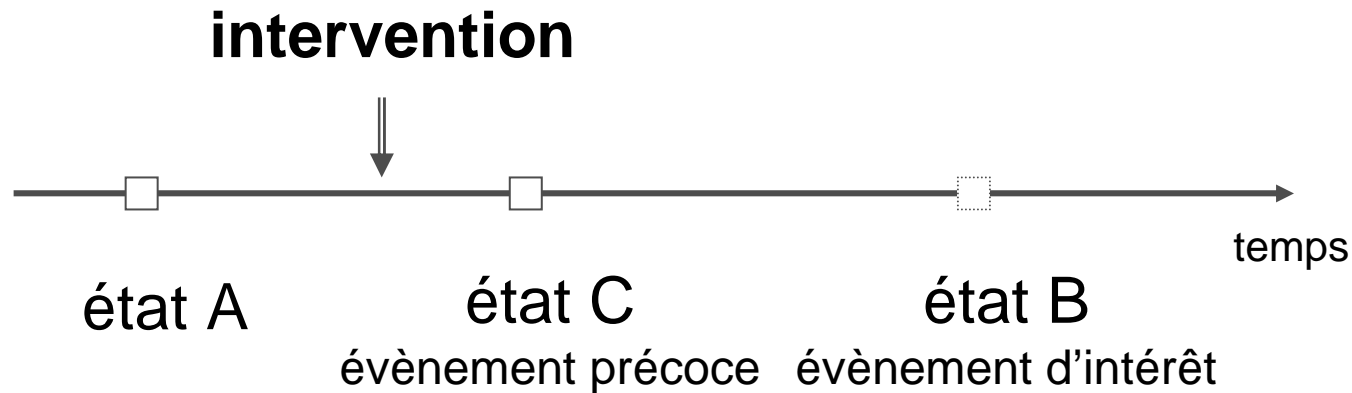
La problématique (1)



Exemple : essai thérapeutique d'anti-cancéreux / évènement d'intérêt : décès par cancer

Observation de l'évènement d'intérêt souvent tardive, rare

Critères de jugement de substitution La problématique (2)



Exemple : essai thérapeutique d'anti-cancéreux / événement précoce : masse tumorale

Observation de l'évènement précoce, fréquente

Définition d'un critère de jugement de substitution

Critère (ou marqueur) de substitution
Surrogate endpoint, surrogate marker

Traduction d'une **évènement précoce, prédisant un évènement d'intérêt**, en une **variable** permettant de répondre à la question de l'étude

- fondée sur un évènement de substitution, plus facile à mesurer que le réel évènement d'intérêt
- directement reliée à la question de l'étude

Élaboration d'un critère de jugement de substitution

Question de l'étude



Evènements précoces, prédisant l'évènement d'intérêt



Variable = Critère de jugement de substitution

Critère de jugement de substitution

Exemple 1

question de l'étude



évènement précoce,
prédisant l'évènement d'intérêt



critère de jugement
de substitution

efficacité
antihypertenseur ?



pression artérielle
diastolique (\Leftrightarrow survie)



moyenne de pression
artérielle diastolique
au temps t

Critère de jugement de substitution

Exemple 2

question de l'étude



évènement précoce,
prédisant l'évènement d'intérêt



critère de jugement
de substitution

efficacité antirétrovirale ?



nombre de CD4/mm³
(\Leftrightarrow survie)



médiane de CD4/mm³
au temps t

Conditions essentielles d'un critère de jugement de substitution

- Facilité de mesure
- Fiabilité
 - reproductibilité et répétabilité
- Mesurable à tous les stades de la maladie
- Validité
 - mesure sans biais
 - l'évènement d'intérêt doit être parfaitement prédit par l'évènement précoce

Démarche de sélection

1. Relation physiopathologique du marqueur avec CJ vrai ?
 - recherche fondamentale
2. Fiabilité de la mesure ?
 - étude de reproductibilité et répétabilité
3. Mesurable à tous les stades de la maladie ?
 - étude transversale / longitudinale
4. Rôle pronostique vis-à-vis du CJ vrai ?
 - étude de cohorte
5. Evolution rapide de concentration sous traitement ?
 - étude de cohorte
6. Importance du rôle substitutif de l'effet du traitement sur le CJ vrai ?
 - essai thérapeutique

Estimation de la part d'effet du traitement

- sur CJ vrai expliquée par effet sur CJ de substitution

– exemple

étude des CD4+ comme CJ de substitution du sida dans essai ZDV/Placebo (*Choi et al. Ann Intern Med 1993; 118:674-80*)

	Placebo/ZDV		part expliquée
	β	RR	
modèle brut	0,53	1,70	
modèle ajusté sur CD4+ à S16	0,29	1,33	$1-0,29/0,53=45\%$
modèle ajusté sur % CD4+ à S16	0,14	1,15	$1-0,14/0,53=74\%$

Choix du critère de jugement

- dépend de l'intervention étudiée (*Prentice. Stat Med 1989; 8:431-40*)
 - exemples
 - CJ vrai : incidence maladies coronariennes
 - intervention : anti-hypertenseurs
 - ⇒ CJ substitution : mesures pression artérielle
 - intervention : hypocholestérolémiantes
 - ⇒ CJ substitution : cholestérolémie
 - pas de CJ de substitution universel

Avantages des critères de substitution

- **Dans l'évaluation d'interventions**
(essais thérapeutiques, évaluation campagnes de prévention)
 - ⇒ diminution taille d'échantillon, durée d'étude, coûts
 - ⇒ facilitation interprétation en limitant problème risques compétitifs, interventions secondaires
- **Dans l'adaptation de stratégies de prise en charge sur le plan individuel**
 - ⇒ adaptation plus précoce

Autres intérêts : selon le cas, mesure moins invasive, moins coûteuse

Limites des critères de jugement de substitution (en cas de mauvais choix)

- Rejeter à tort une intervention efficace sur le « vrai » CJ
- Retenir à tort une intervention inefficace
- CAST (Cardiac Ahythmia Suppression Trial)
 - encainide et flecainide : suppression des arythmies chez 1727 des 2309 patients traités (75%).
 - Évaluation de l'efficacité de encainide et flecainide, contre placebo, dans le traitement des arythmies après infarctus du myocarde
 - Groupes traités : 56 décès (8%)
 - Groupe placebo : 22 décès (3%)
 - RR = 2.5, CI [1.6-4.5]

Critères de jugement pour évaluer les antirétroviraux

- Mortalité
- Événements sida, morbidité
- Lymphocytes CD4+
- ARN VIH plasmatique
- Qualité de vie
- Etc.

% patients / ARN VIH-1 < seuil en fin d'essai

$$\mathbb{1} (\text{ARN}_{\text{fin}} < \text{seuil}) \rightarrow \%$$

interprétation clinique

valeurs < seuil

étendue valeurs > seuil

données intermédiaires

dynamique

valeurs manquantes

indétectabilité

oui

non

non

non

très sensible

changement moyen d'ARN VIH-1 en fin d'essai

$$\begin{aligned} & \text{si } \text{ARN} < \text{seuil} \rightarrow \text{seuil} \\ & \text{ARN}_{\text{fin}} - \text{ARN}_{\text{S00}} \rightarrow \text{moyenne} \end{aligned}$$

interprétation clinique

valeurs $<$ seuil

étendue valeurs $>$ seuil

données intermédiaires

dynamique

valeurs manquantes

évolution globale

non

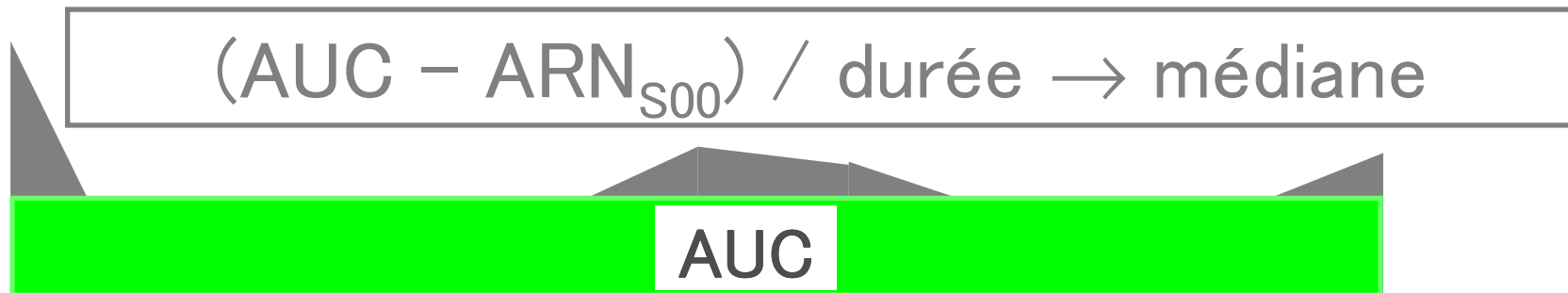
oui

non

oui/non

très sensible

aire médiane sous la courbe d'ARN VIH-1



interprétation clinique

valeurs < seuil

étendue valeurs > seuil

données intermédiaires

dynamique

valeurs manquantes

imprégnation ?

non

oui

oui

non

± sensible

% patients / maintien ARN VIH-1 < seuil

$$\mathbf{1} \text{ (ARN} < \text{seuil}_{S00 \rightarrow \text{fin}}) \rightarrow \%$$

interprétation clinique

valeurs < seuil

étendue valeurs > seuil

données intermédiaires

dynamique

valeurs manquantes

indéfectabilité

oui

non

oui

non

± sensible

TLOVR

(Time to Loss of Virologic Response)

- Pour tous les patients avec une charge virale confirmée en dessous de la limite définie, le temps jusqu'à l'échec est le temps (le plus court) jusqu'à la survenue d'un événement spécifique.
- Ces événements peuvent être :
 - Le décès
 - L'arrêt définitif d'administration du traitement ou les perdus de vue
 - Introduction d'un nouvel ARV (à moins qu'un des ARV du traitement optimisé soit modifié en raison de toxicité ou intolérance clairement imputée à cet ARV)
 - CV confirmée \geq à l'objectif fixé

Bibliographie

- **Critères de jugement**

- Pinatadosi, S. Clinical Trials: A Methodological Perspective. Wiley; New York: 1997
- Pocock SJ. Clinical trials: a practical approach. Chichester: Wiley, 1983

- **Critères de jugement de substitution**

- Pinatadosi, S. Clinical Trials: A Methodological Perspective. Wiley; New York: 1997
- Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med. 1989;8:431-40
- Demol P and Weihrauch TR. Surrogate endpoints, their utility for evaluating therapeutic efficacy in clinical trials. Applied Clin Trials 1998;7:46–56

- **Critères de jugement composites**

- Freemantle N, Calvert M, Wood J et al. Composite outcomes in randomized trials: greater precision but with greater uncertainty? JAMA 2003;289:2554-9
- Chi GY. Some issues with composite endpoints in clinical trials. Fundam Clin Pharmacol 2005;19:609-19