

Hétérogénéité en méta-analyse particularités des variables temps-dépendantes



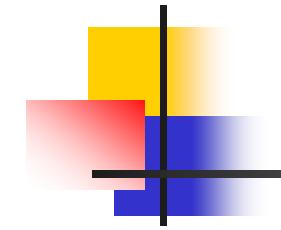
Dr. Christophe Combescure
Dr Guy Haller
Hôpitaux Universitaires de Genève



OBJECTIFS

- Rappel méthodologique
- Points critiques classiques
- Point critique particulier: l'hétérogénéité
- Les variables temps-dépendantes

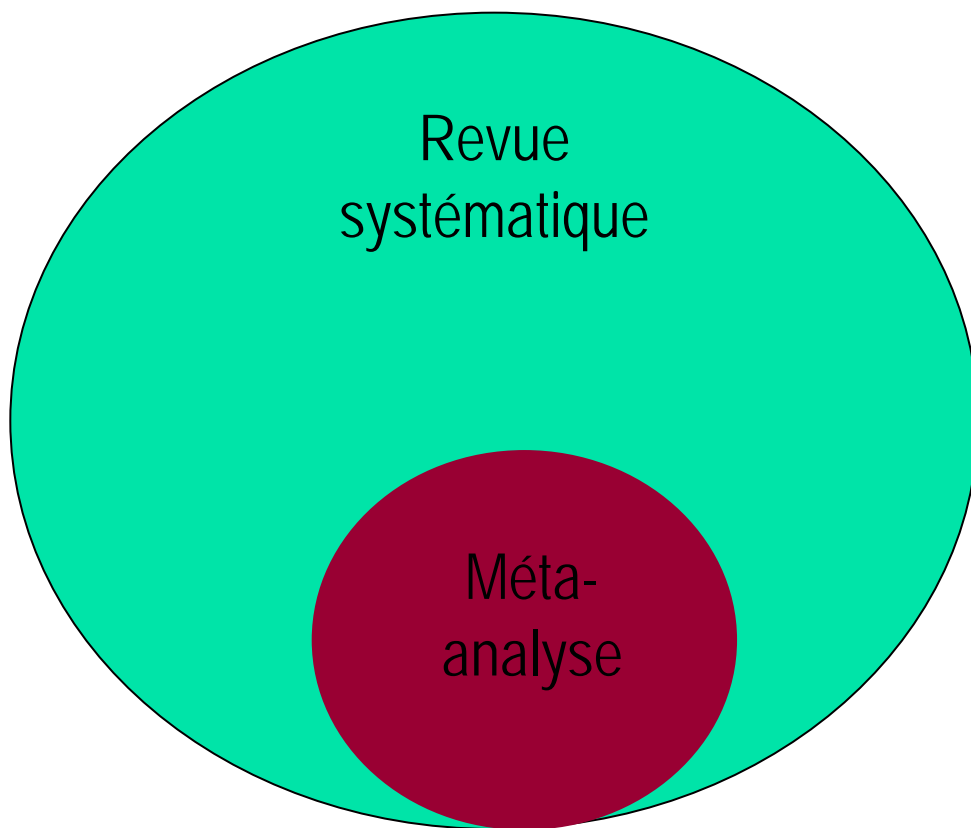
DEFINITION – Revues systématiques méta-analyses



Méthode systématisée et clairement définie d'identification de sélection et de validation de l'information intégrée dans une revue de littérature.

Lorsque l'information contenue dans les différents articles sélectionnés est analysée statistiquement et résumée de manière numérique, on parle de méta-analyse

Rappel



Procédure

1. *Définition question de recherche*
2. *Critères d'inclusion - exclusion et recherche systématique d'articles correspondant*
3. *Résumé de chaque article*
4. *Evaluation de la qualité des articles sélectionnés*
5. *Contact avec auteurs articles si nécessaire*
6. *Analyse des résultats des études sélectionnées*
7. *Combinaison des études (+/- méta-analyse)*
8. *Choix approprié des méthodes statistiques*

Efficacy and safety of ephedra and ephedrine for weight loss and athletic performance: a meta-analysis

Shekelle PG, Hardy ML, Morton SC, Maglione M, Mojica WA, Suttorp MJ et al
JAMA. 2003 Mar 26;289(12):1537-45

Introduction

Ephedra and ephedrine sometimes are used for weight loss or enhanced athletic performance, but the efficacy and safety of these compounds are uncertain. The study purpose was to assess the efficacy and safety of ephedra and ephedrine used for weight loss and enhanced athletic performance.

Method

We searched 9 databases using the terms ephedra, ephedrine, adverse effect, side effect, efficacy, effective, and toxic. We included unpublished trials and non-English-language documents. Adverse events reported to the US Food and Drug Administration MedWatch program were assessed. Eligible studies were controlled trials of ephedra or ephedrine used for weight loss or athletic performance. Two reviewers independently identified trials of efficacy and safety of ephedra and ephedrine on weight loss or athletic performance; disagreements were resolved by consensus.

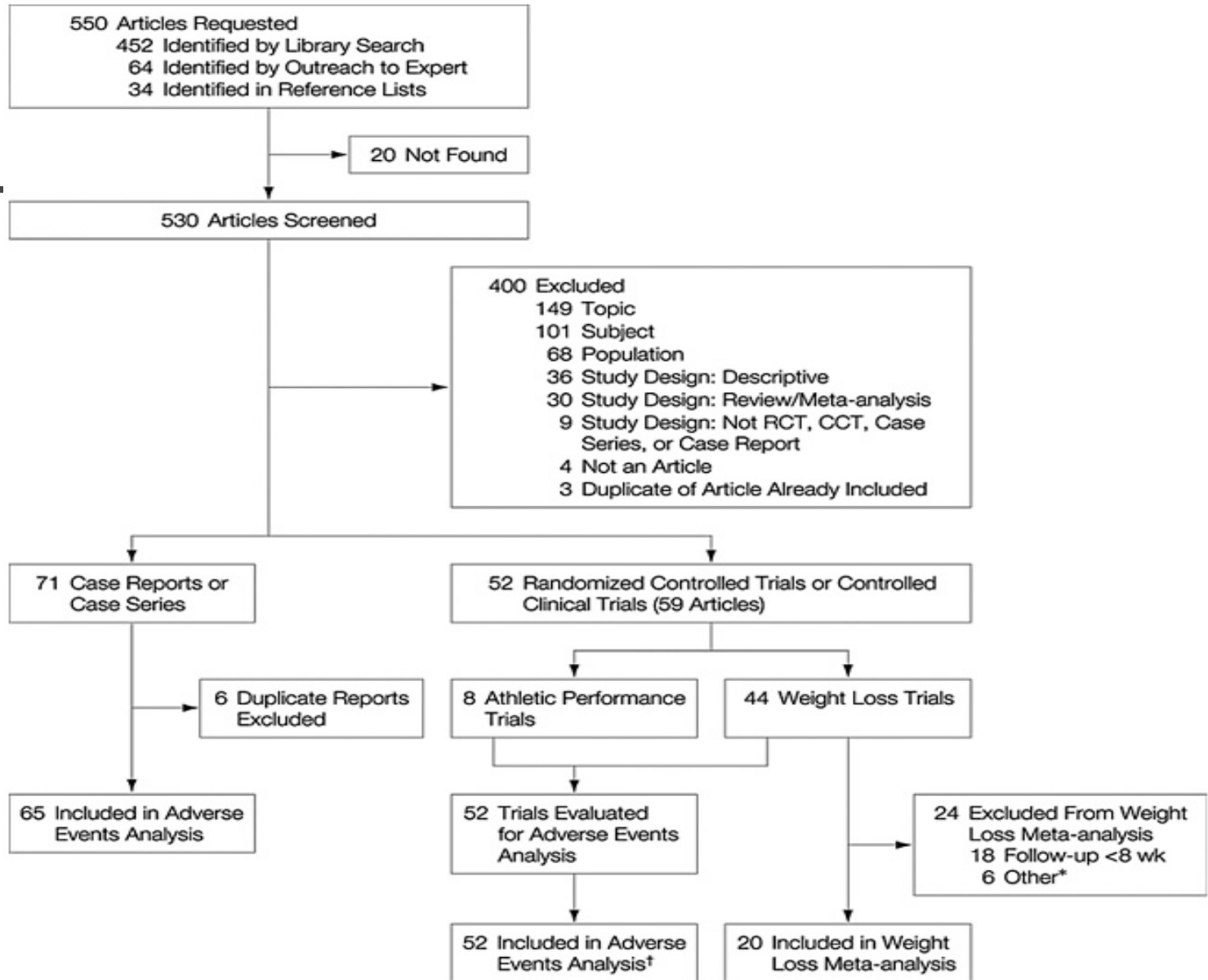
Results

Pooled results for trials comparing placebo with ephedrine (n = 5), ephedrine and caffeine (n = 12), ephedra (n = 1), and ephedra and herbs containing caffeine (n = 4) yielded estimates of weight loss (more than placebo) of 0.6 (95% confidence interval, 0.2-1.0), 1.0 (0.7-1.3), 0.8 (0.4-1.2), and 1.0 (0.6-1.3) kg/mo, respectively. Safety data from 50 trials yielded estimates of 2.2- to 3.6-fold increases in odds of psychiatric, autonomic, or gastrointestinal symptoms, and heart palpitations..

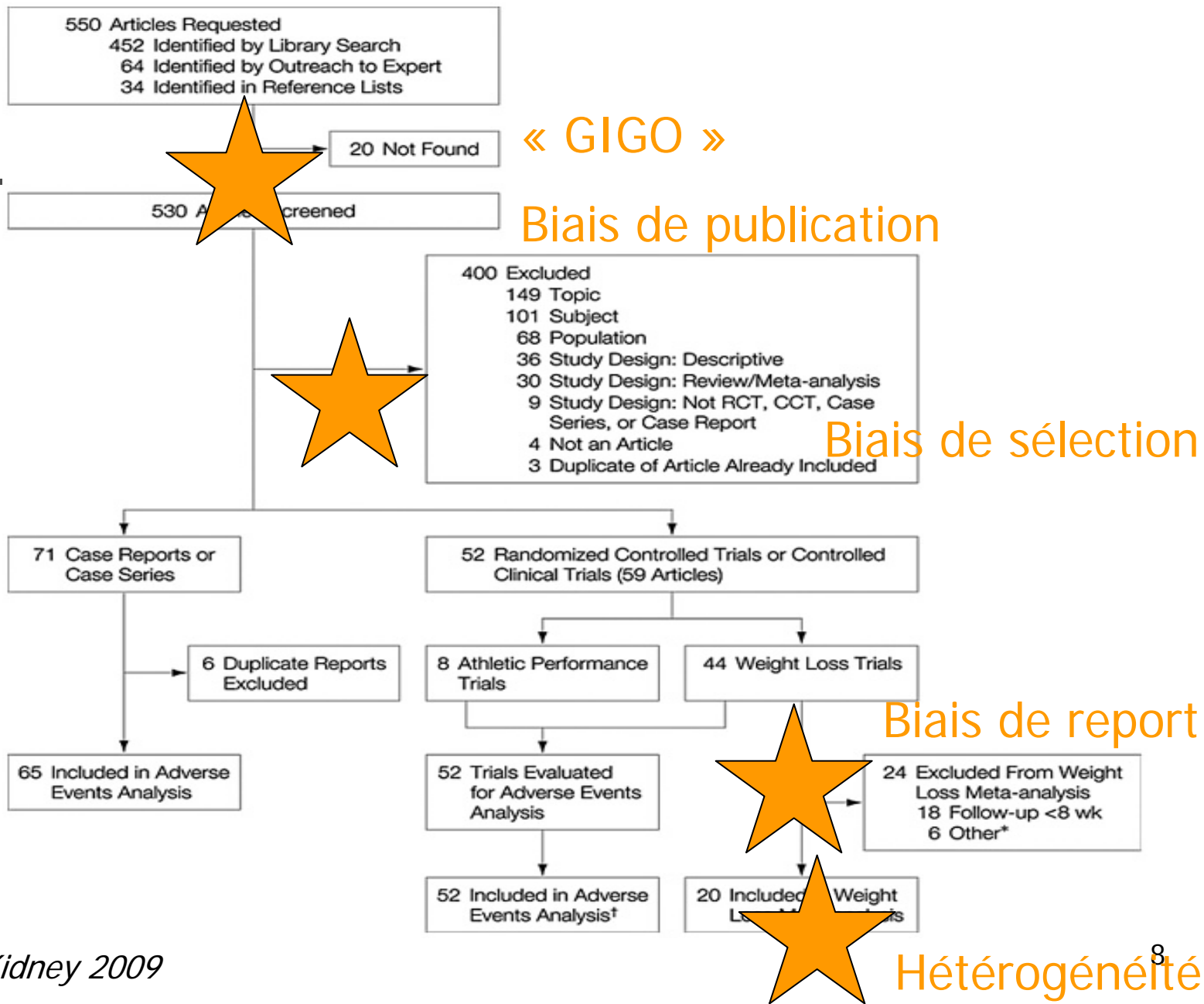
Discussion

Ephedrine and ephedra promote modest short-term weight loss (approximately 0.9 kg/mo more than placebo) in clinical trials. Use of ephedra or ephedrine and caffeine is associated with increased risk of psychiatric, autonomic, or gastrointestinal symptoms.

Diagramme d'une méta-analyse



Sources de biais





Hétérogénéité

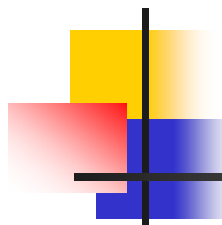
C'est la variation des résultats observés
parmi toutes les études sélectionnées

Se diagnostique par....

Hétérogénéité: diagnostique

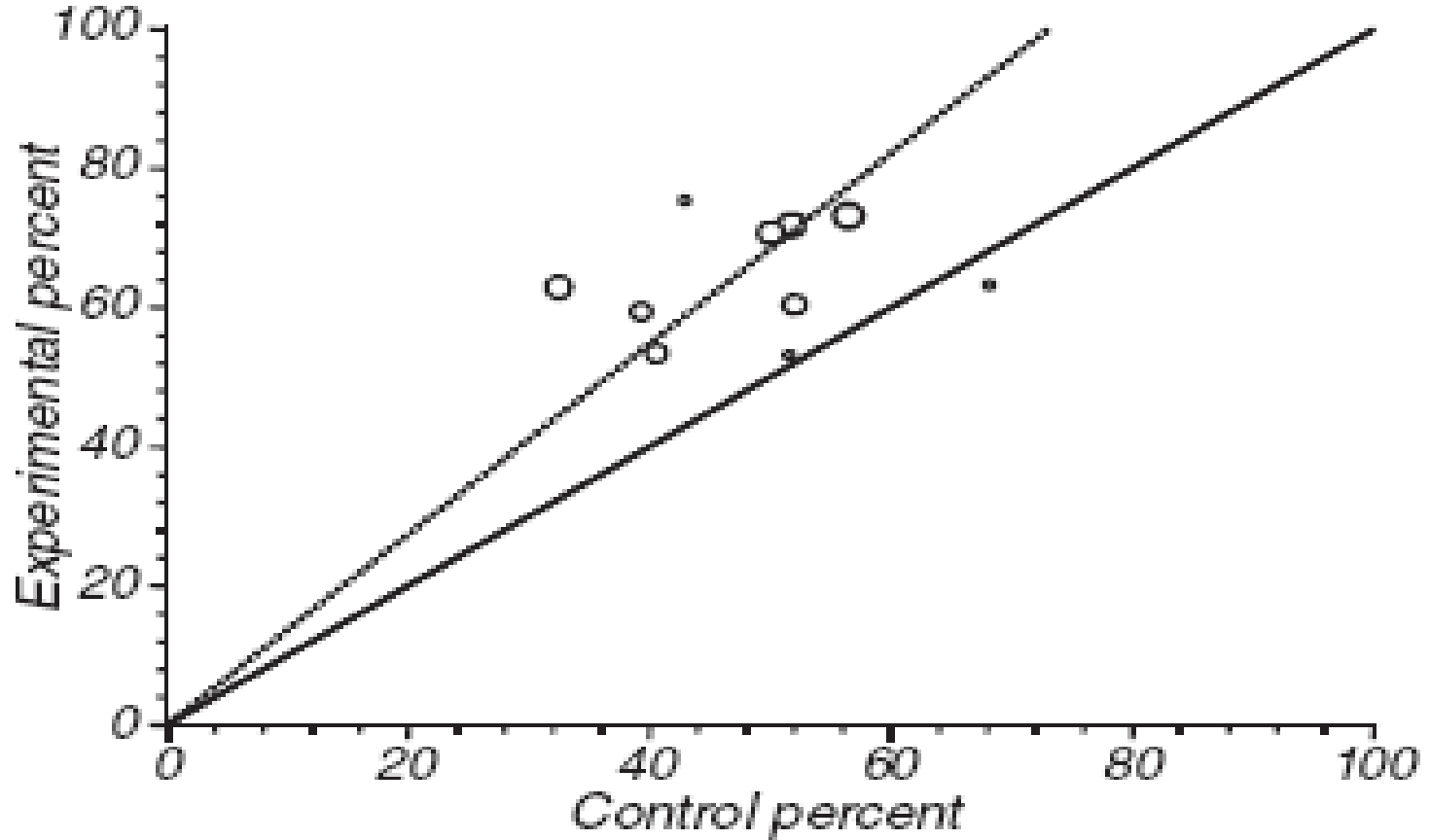


- Connaissance clinique
- L'Abbe-plots
- Test Cochran's Q et I²
- Tests de sensibilité (outliers)
- Meta-regression/analyse de sous groupe



L'Abbé plot for treatment

L'Abbe plot (symbol size represents sample size)



Test d'hétérogénéité

■ Test de Cochran

$$Q = \sum_i w_i (\theta_i - \theta_{Commun})^2$$

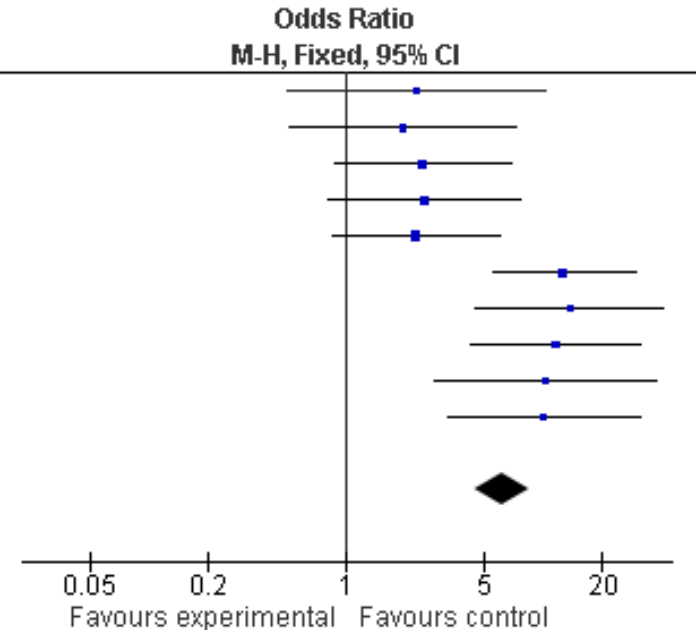
- Où $w = 1/\text{Var}$ et θ = effet mesuré
Sous H_0 : "Homogénéité" ; Q suit une loi du Chi-deux à $k-1$ ddl effet mesuré
- Un résultat statistiquement significatif signifie que l'écart entre les études est plus important que leur effet

■ Indicateur I^2

- L'indicateur I^2 est une alternative au test: il mesure la proportion d'hétérogénéité dans les études qui ne peut pas être expliquée par le hasard seul mais en tenant compte du nombre d'études analysées
- Les valeurs de I^2 égales à 25%, 50%, et 75% représentent respectivement une hétérogénéité basse, modérée et forte.

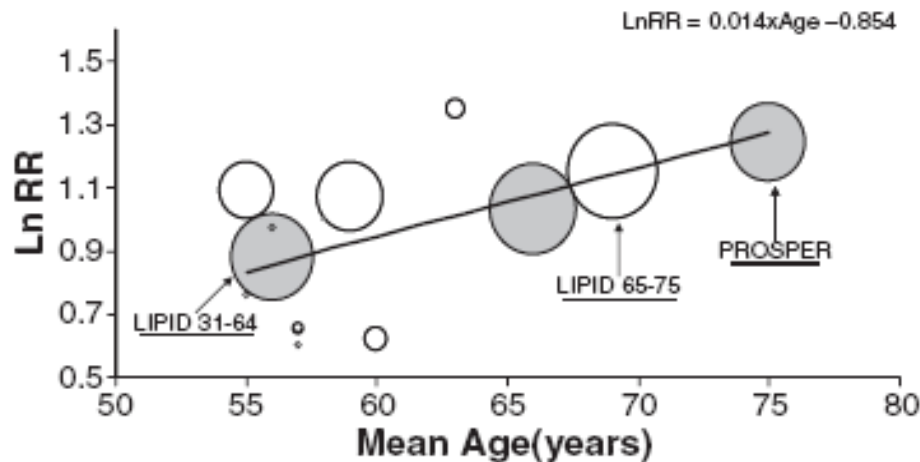
Exemple fictif d'une méta-analyse avec hétérogénéité

Study or Subgroup	Experimental		Control		Weight	Odds Ratio M-H, Fixed, 95% CI
	Events	Total	Events	Total		
Etude 1	6	28	3	28	7.0%	2.27 [0.51, 10.18]
Etude 10	7	35	4	35	9.4%	1.94 [0.51, 7.33]
Etude 2	13	65	6	65	14.2%	2.46 [0.87, 6.93]
Etude 3	11	58	5	58	11.9%	2.48 [0.80, 7.66]
Etude 4	14	71	7	71	16.6%	2.25 [0.85, 5.95]
Etude 5	56	140	7	140	12.4%	12.67 [5.51, 29.10]
Etude 6	34	84	4	84	7.0%	13.60 [4.55, 40.64]
Etude 7	37	93	5	93	8.9%	11.63 [4.31, 31.36]
Etude 8	20	52	3	52	5.4%	10.21 [2.80, 37.18]
Etude 9	26	67	4	67	7.2%	9.99 [3.25, 30.73]
Total (95% CI)		693		693	100.0%	6.19 [4.45, 8.60]
Total events	224		48			
Heterogeneity: Chi ² = 22.0, df = 9 (P = 0.009); I ² = 59%						
Test for overall effect: Z = 10.86 (P < 0.00001)						



Méta-régression

Modélise le RR/OR de l'outcome selon un prédicteur commun aux différentes études





Causes d'Hétérogénéité: différences cliniques

Liées

- population
- type d'intervention
- type d'outcome



Causes d'hétérogénéité: différences méthodologiques

Liées

- au design d'étude
- contenu (placebo/meilleur ttt),
- qualité de l'étude
- durée de suivi

Hétérogénéité: comment s'en accommoder



- Sélection études appropriées-Scores qualités
- Grouper les études par design/outcome
- Echelles de transformation validées d'outcome
- Tests statistiques généraux appropriés
(effets fixes-aléatoires)
- Tests particuliers

Méta-analyse d'outcomes binaires lorsque les longueurs de suivi varient

■ Contexte

- Pour chaque étude: tableau 2*2 (Bras * Événement)
- Classiquement, méta-analyses d'ORs ou de RRs

■ Problème

- Dans les méta-analyses, les études peuvent avoir des durées de suivi hétérogènes
- Lettre de D'Amico « Length of follow up is poorly reported »
 - *BMJ* 1999;319:1199 (30 October)
 - Sa conclusion « Our findings suggest that, as so many trial reports omit mentioning length of follow up, in practice it may not be possible to adjust for length of follow up in a systematic review »

■ Objectifs:

- Etudier l'impact sur les méta-analyses, de l'hétérogénéité des durées de suivi
- Proposer une solution simple et robuste par rapport à des longueurs de suivi différentes



Plan

- ORs et RRs dans un contexte de données de survie
- Un nouveau candidat: le relative log survival (RLS)
- Méthode de méta-analyse: inverse de la variance
- Comparaison des méta-analyses d'ORs, RRs et RLSs par une étude de simulations
- Exemple: chimiothérapie palliative vs « supportive care » dans le cancer colorectal
- Conclusion

Odds ratios et risk ratios dans un contexte de données de survie

- Notations :

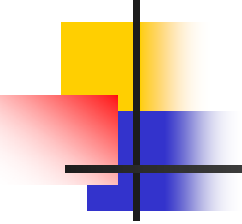
- effectifs observés à un temps t donné dans une étude :

	Absence d'événement	Événement	
Bras 1	a	b	N_1
Bras 2	c	d	N_2

- Alors:

$$OR = \frac{ad}{bc}$$

$$RR = \frac{d/N_2}{b/N_1}$$

- 
- Mais les effectifs observés a , b , c et d dépendent du temps: plus le temps avance, plus on observe d'événements
 - Probabilités de survie sont notées $S_1(t)$ et $S_2(t)$ dans les bras 1 et 2

	Absence d'événement	Evénement	
Bras 1	$N_1 S_1(t)$	$N_1 [1-S_1(t)]$	N_1
Bras 2	$N_2 S_2(t)$	$N_2 [1-S_2(t)]$	N_2

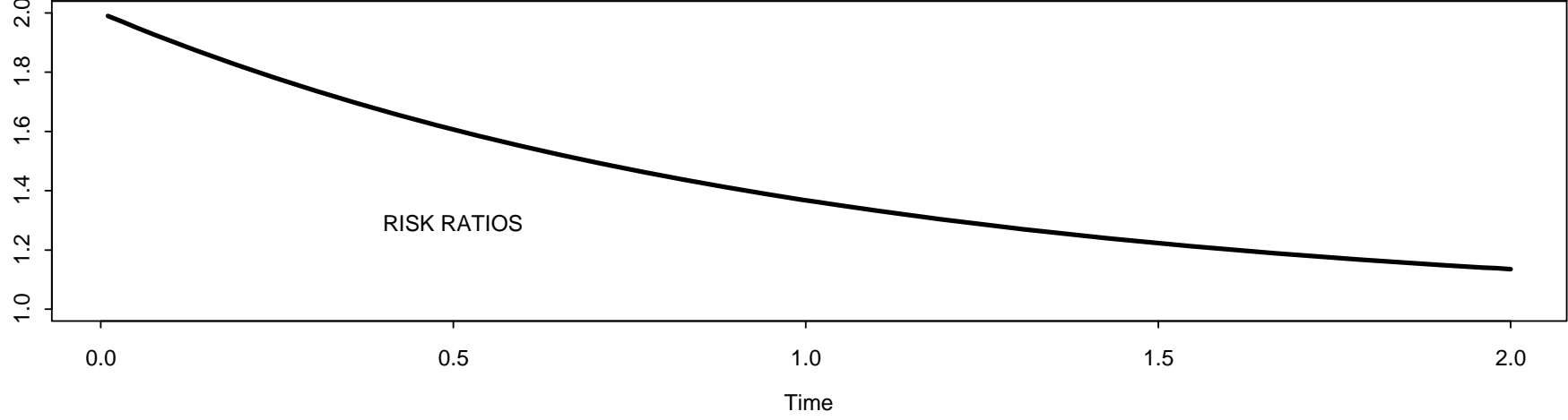
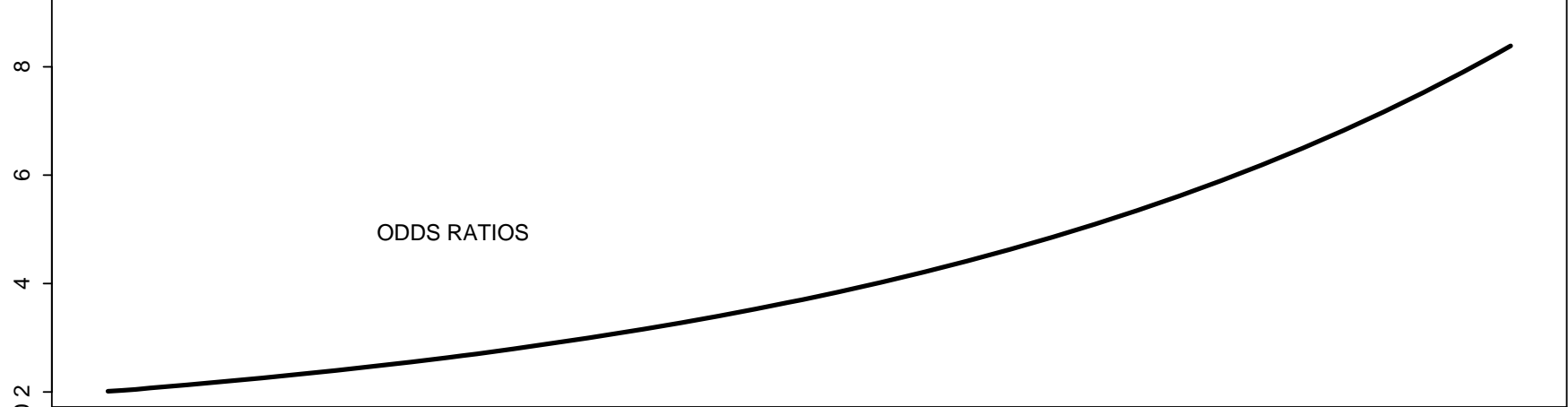
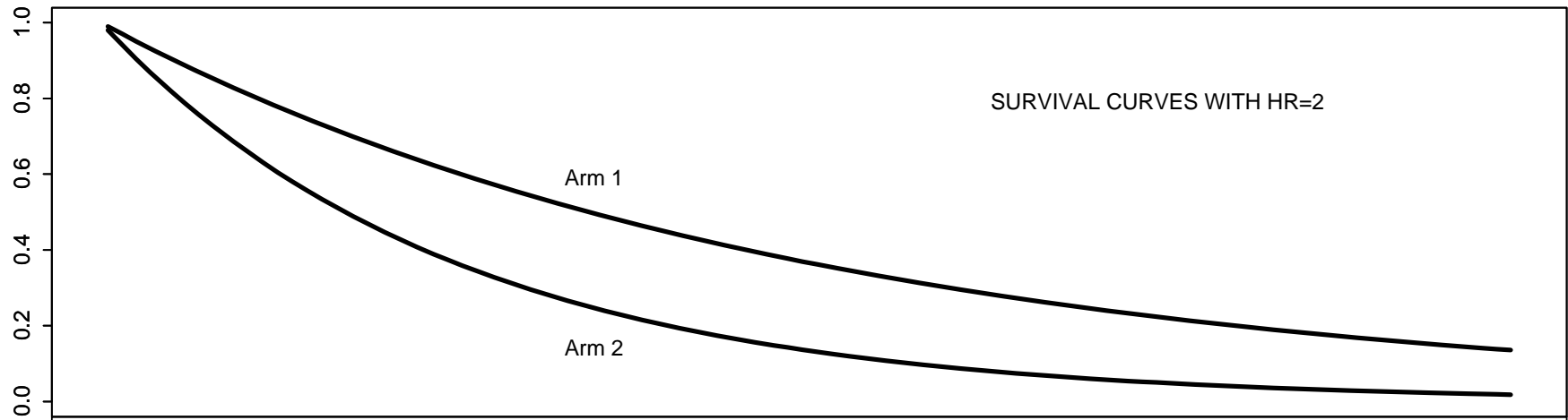


- Par conséquent:

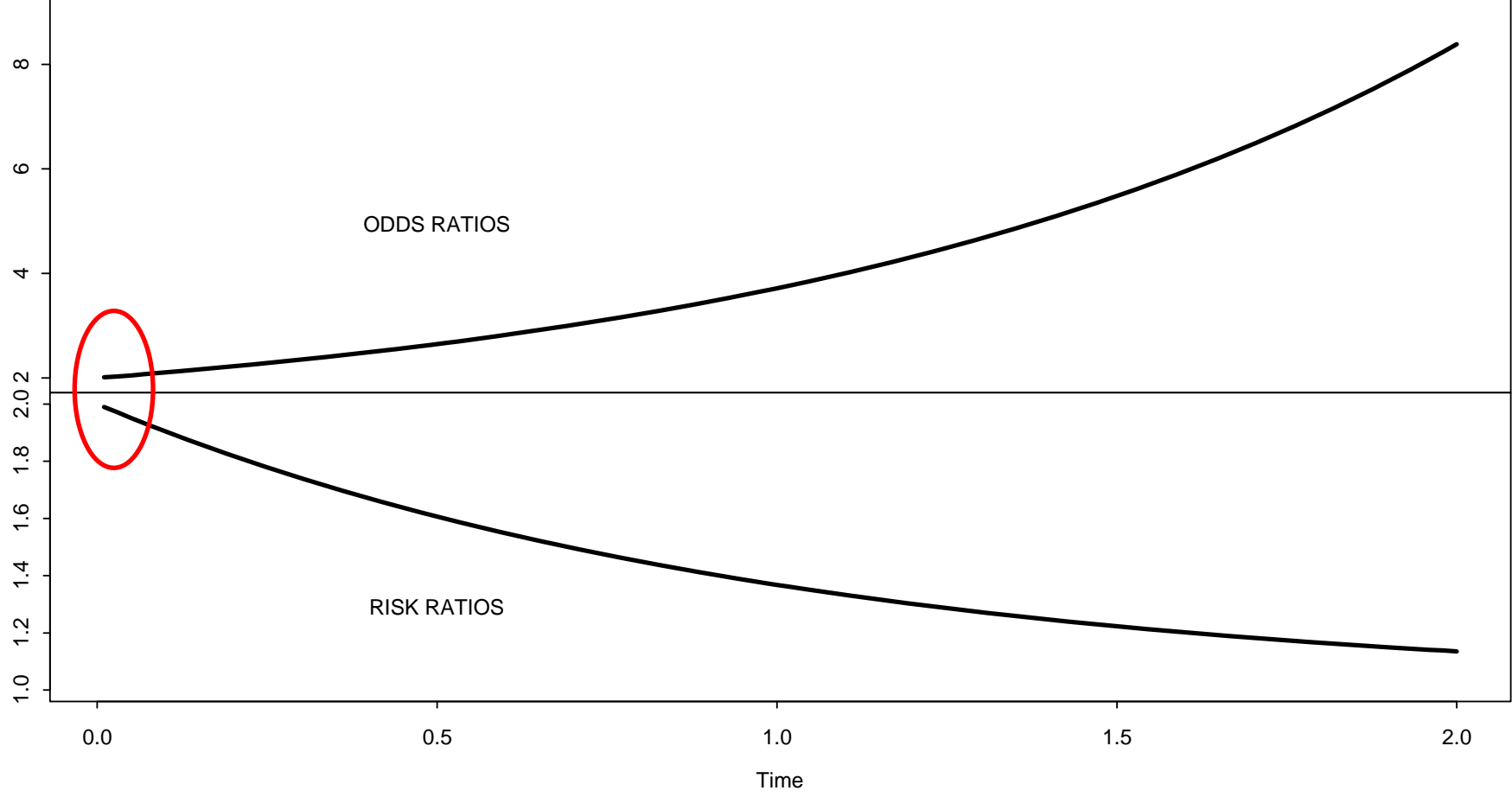
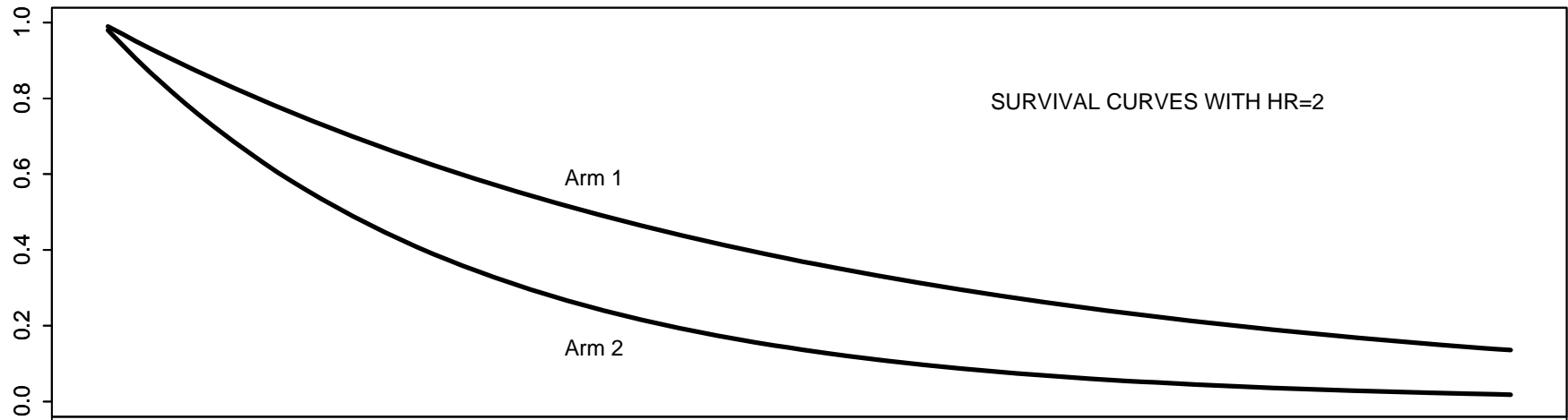
- $OR(t) = S_1(t) [1 - S_2(t)] / S_2(t) [1 - S_1(t)]$
- $RR(t) = [1 - S_2(t)] / [1 - S_1(t)]$

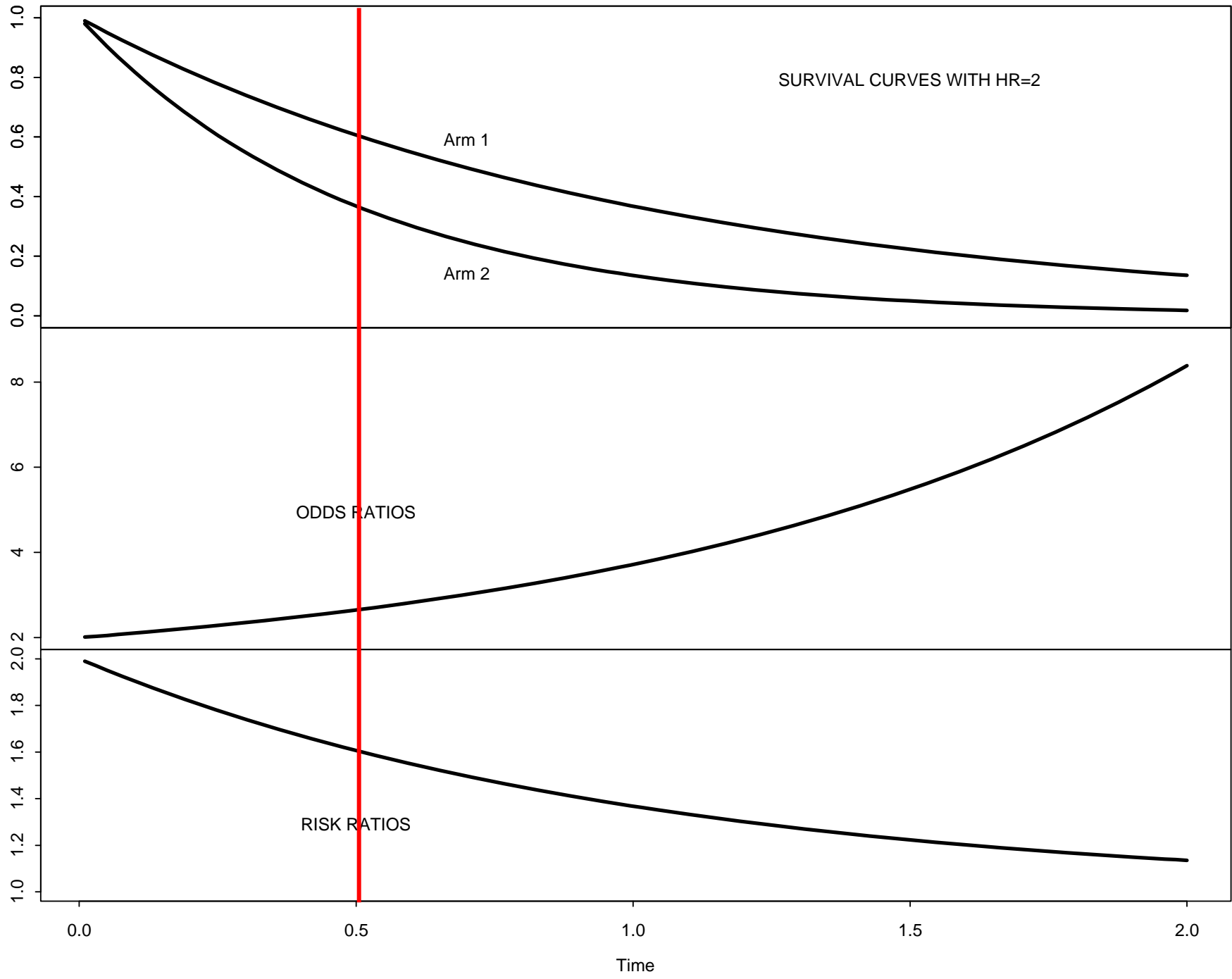
- OR et RR sont fonction du temps

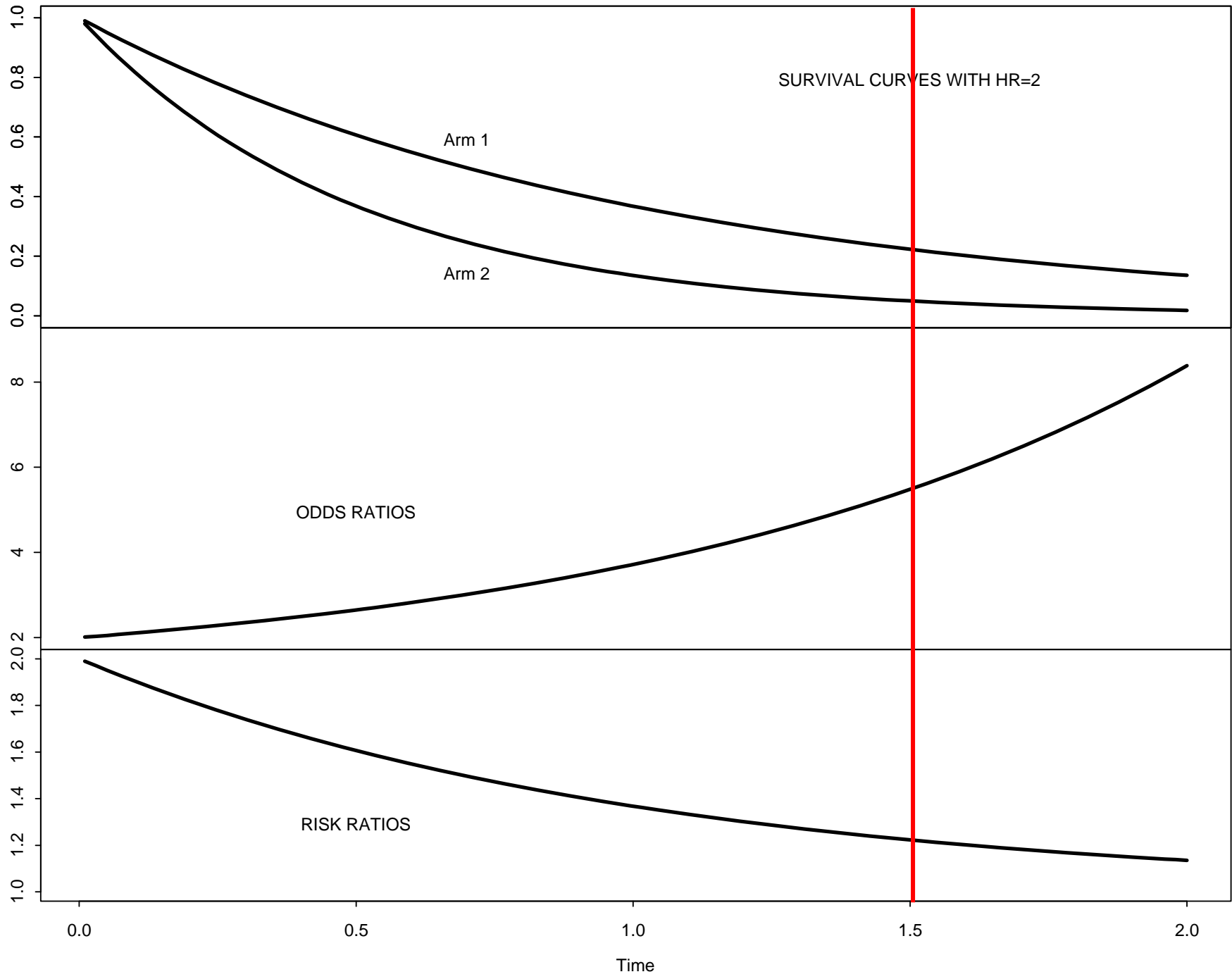
- Question: comment évoluent l'OR et le RR en fonction du temps ?



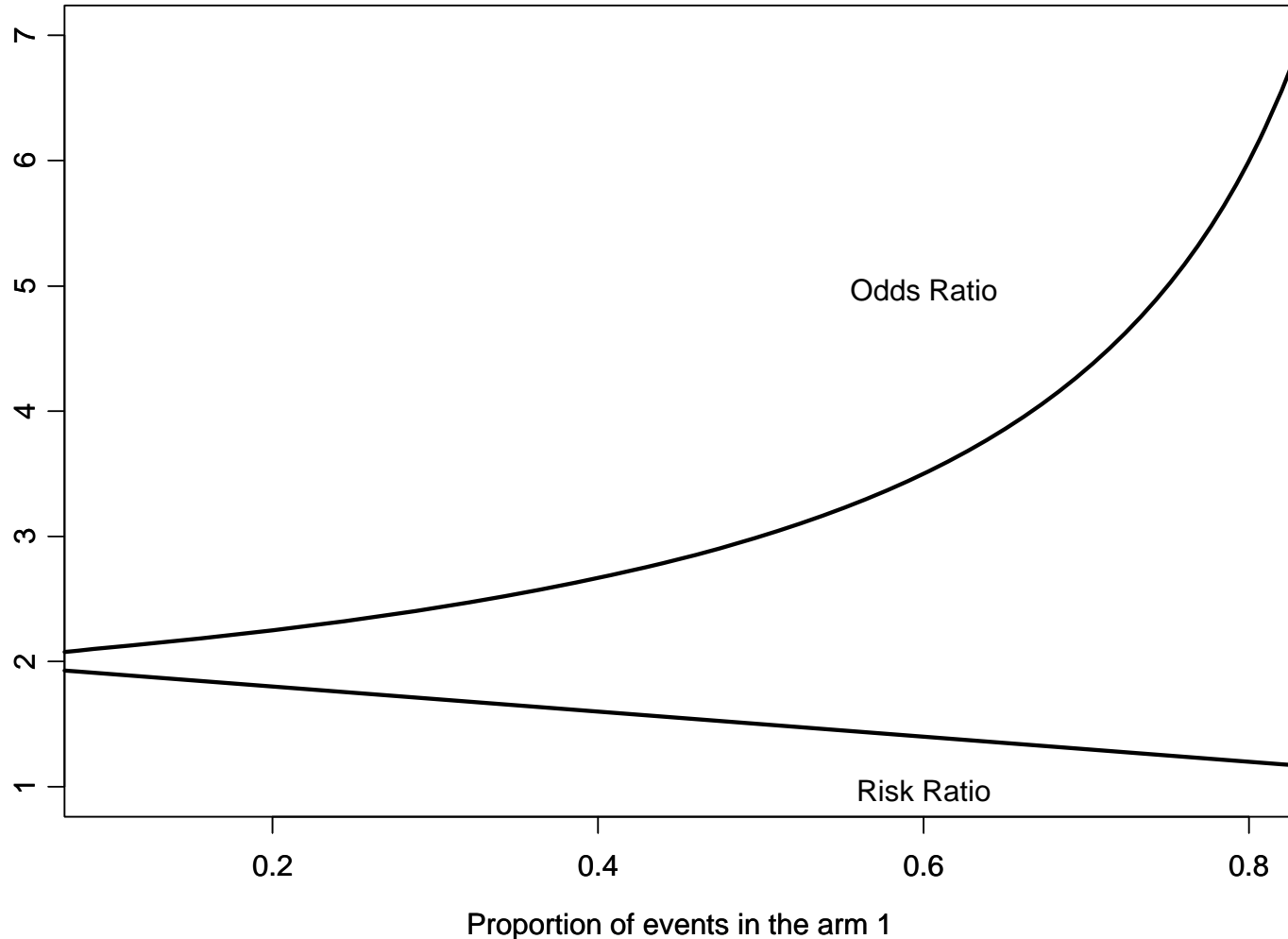
}



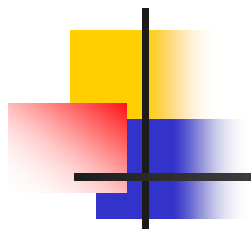




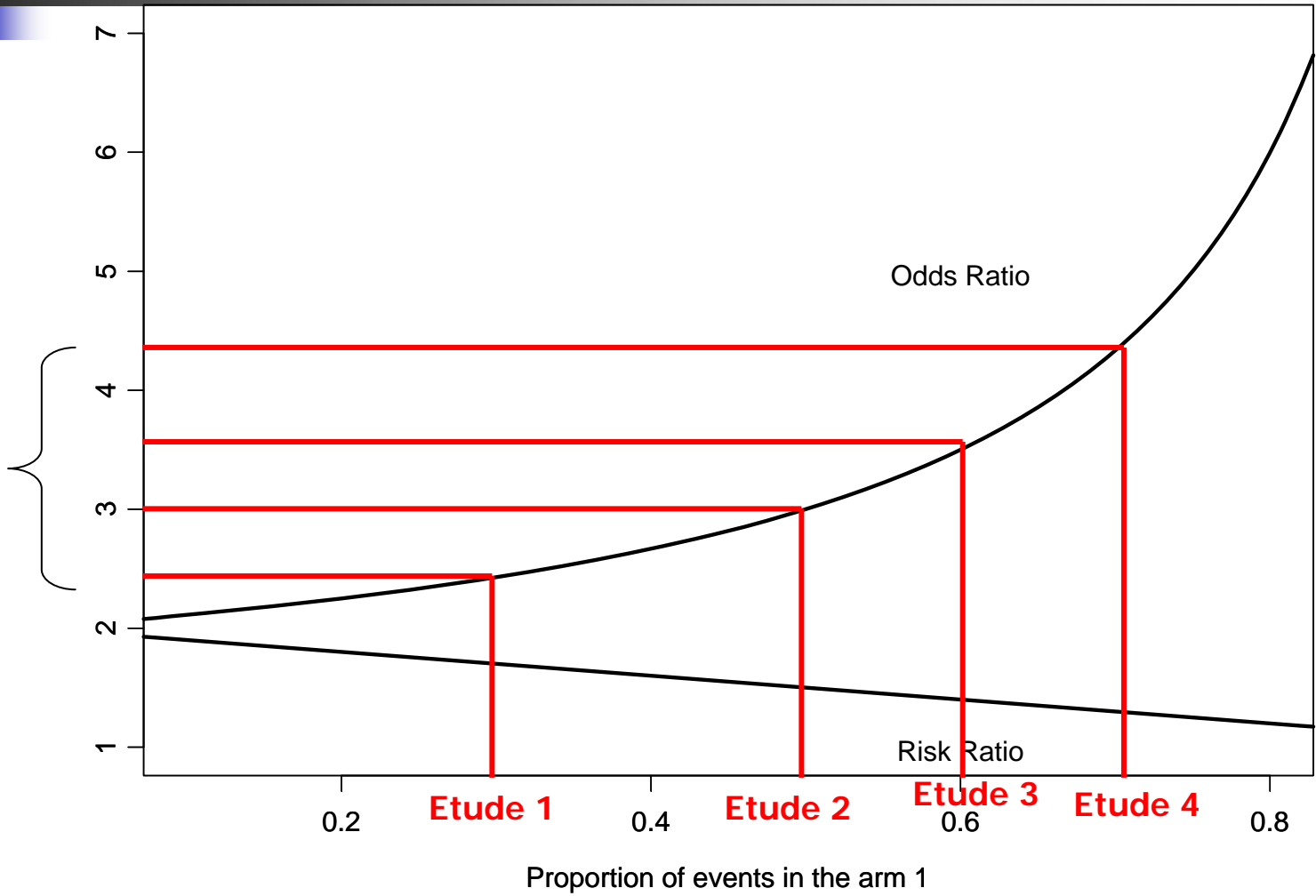
- Sous l'hypothèse d'un modèle à hasard proportionnel (avec $HR=2$):

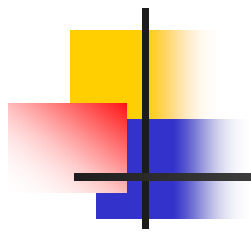


- 
-
- Impact dans les méta-analyses ?

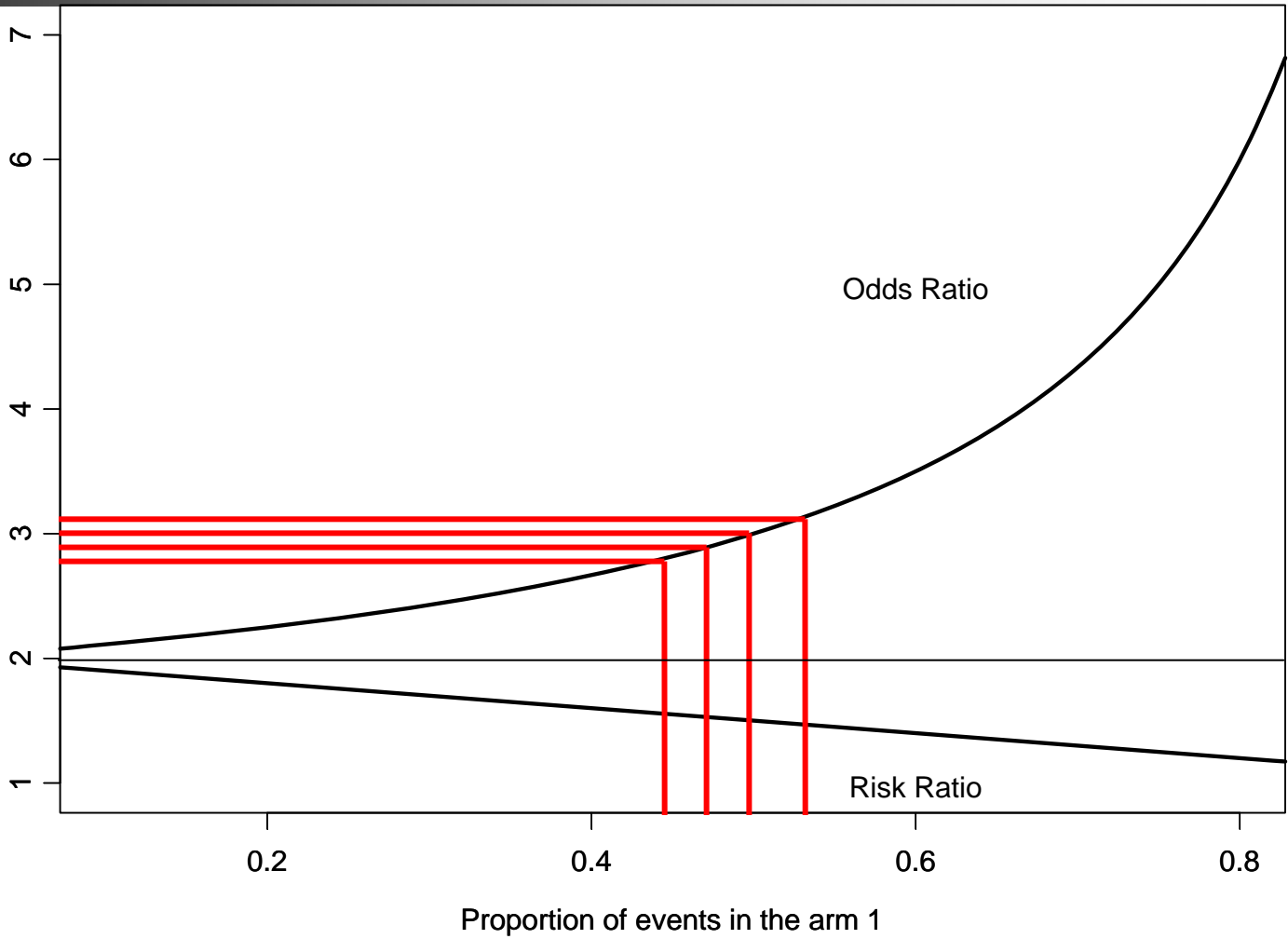


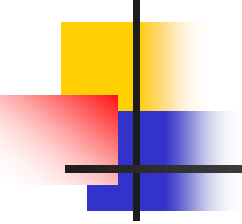
Hétéro-généité





Sur-
estimation
de l'effet



- 
-
- Dans une méta-analyse d'ORs ou de RRs, lorsque les longueurs de suivi diffèrent entre les études, on peut s'attendre à:
 - Détecter de l'hétérogénéité
 - Un biais dans l'estimation de l'effet
 - Sur-estimation si méta-analyse d'ORs
 - Sous-estimation si méta-analyse de RRs



Un nouveau candidat

- Sous l'hypothèse de hasard proportionnel:

$$S_2(t) = [S_1(t)]^{\exp(\beta)}$$

$$\text{Log}[S_2(t)] = \exp(\beta)\text{Log}[S_1(t)]$$

$$\frac{\text{Log}[S_2(t)]}{\text{Log}[S_1(t)]} = \exp(\beta)$$

- Le ratio du logarithme des survies (RLS) est égal au HR

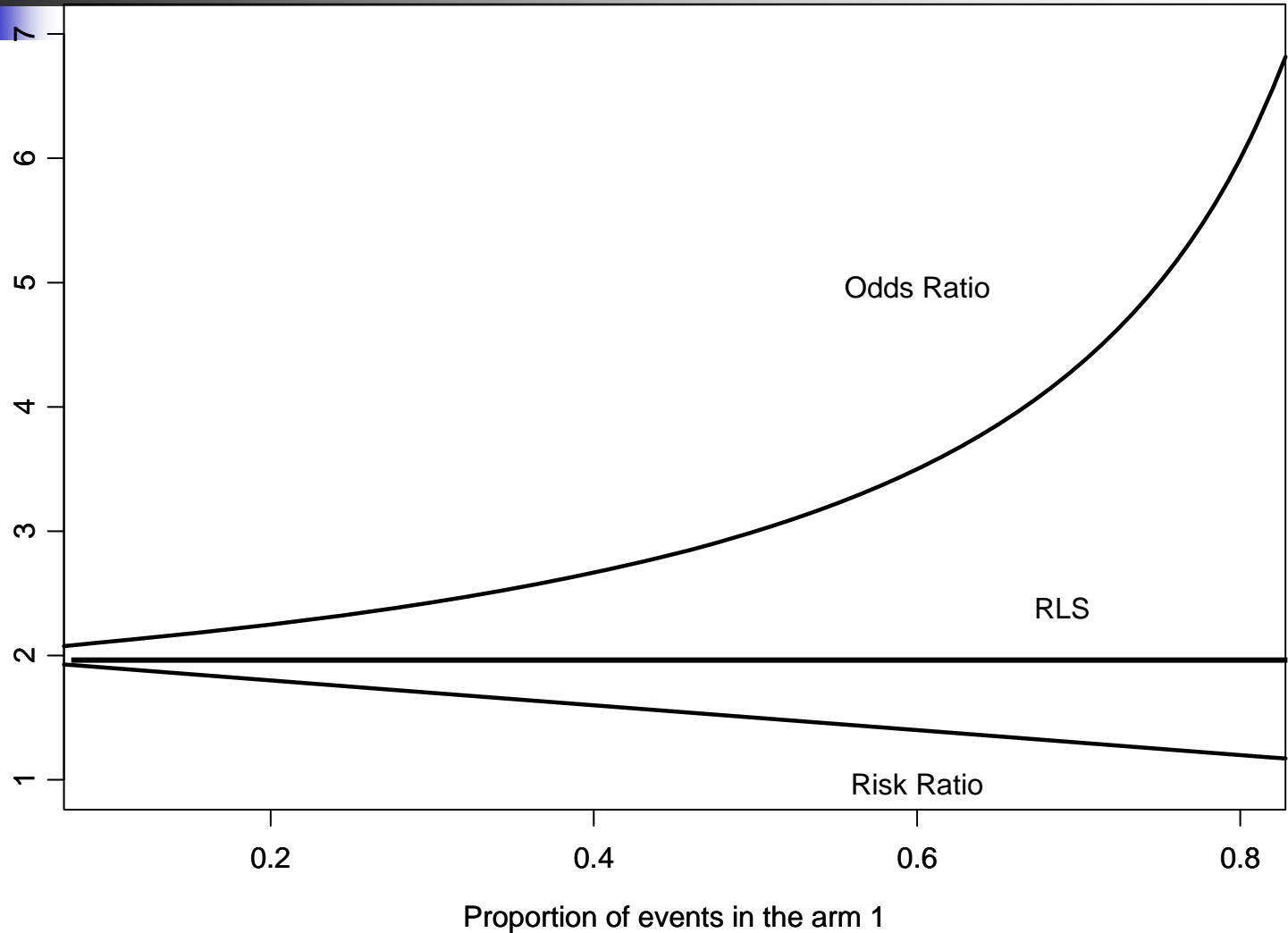
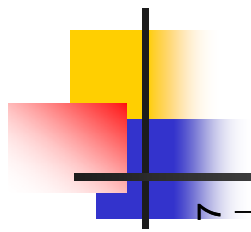
- 
-
- En pratique:

$$R\hat{L}S = \frac{\log\left(\frac{c}{c+d}\right)}{\log\left(\frac{a}{a+b}\right)}$$

est un estimateur du HR, qui est calculable à partir d'un tableau 2*2

- Variance (formule simplifiée):

$$Var\left[\log(R\hat{L}S)\right] = \frac{1}{b} + \frac{1}{d}$$

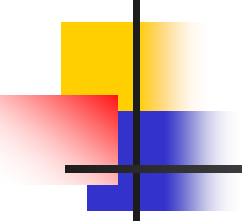


Méthode de méta-analyse: inverse de la variance

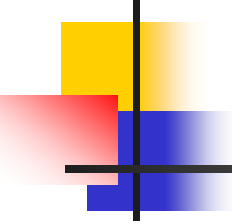
- K études
- 1 tableau 2*2 pour chaque étude:
- Dans chaque étude, on calcule l'indicateur OR, RR ou RLS, noté θ_k
- Méthode de méta-analyse: inverse de la variance

$$\text{Log}(\theta_{Pooled}) = \frac{\sum_{k=1}^K W_k \text{Log}(\theta_k)}{\sum_{i=1}^K W_i}$$

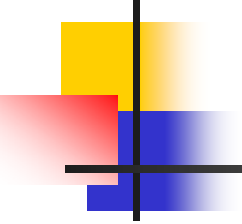
avec W_k l'inverse de la variance de θ_k

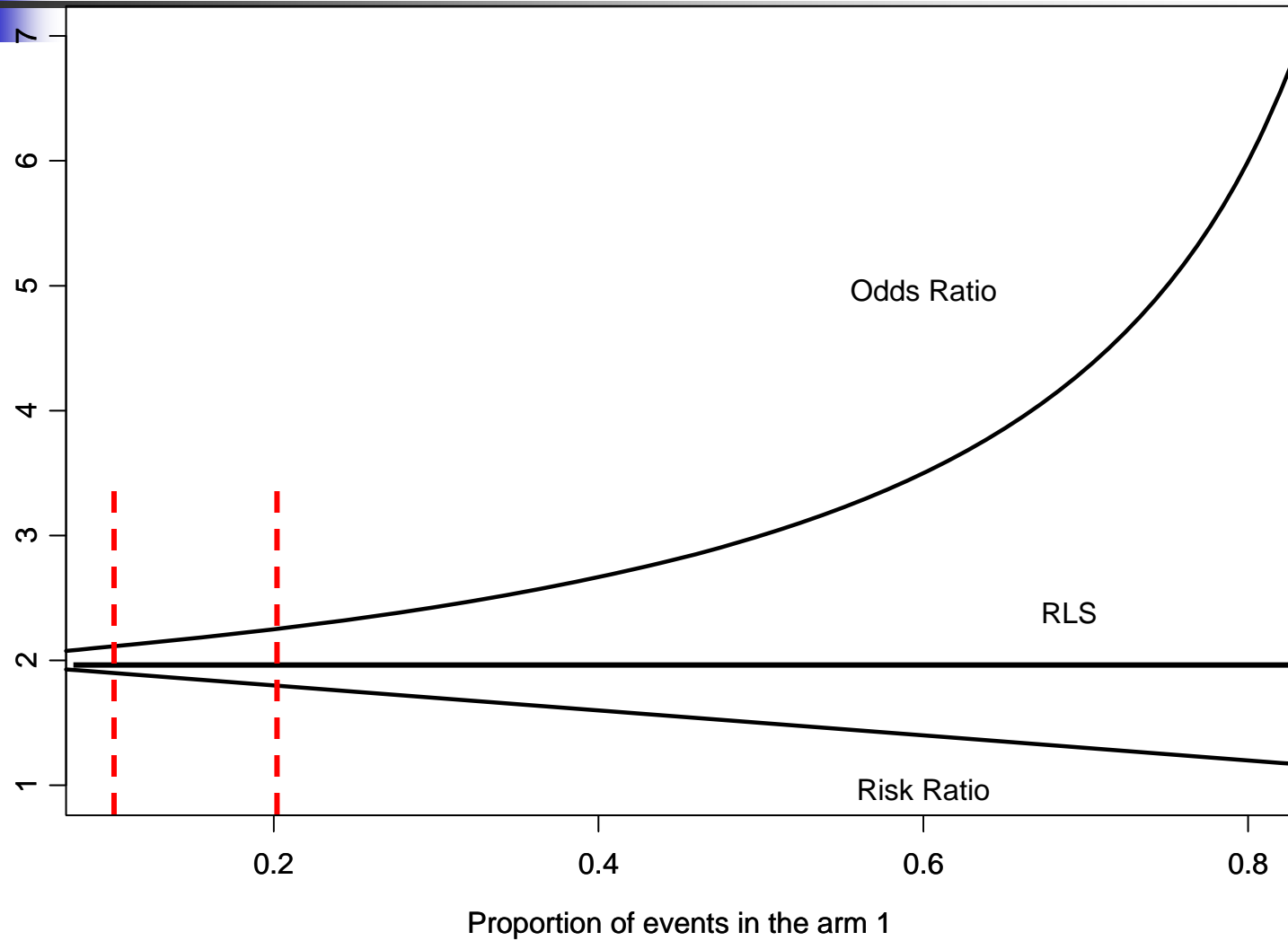
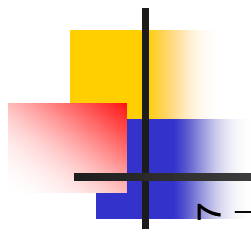
- 
-
- On peut aussi:
 - Calculer l'intervalle de confiance autour de θ_{Pooled}
 - Tester l'effet
 - Mesurer et tester l'hétérogénéité (I-square, test de Cochran)
 - Utiliser un modèle à effets aléatoires (Der Simonian)
 - La démarche est identique pour OR, RR et RLS. Seules les expressions de θ_k et W_k changent

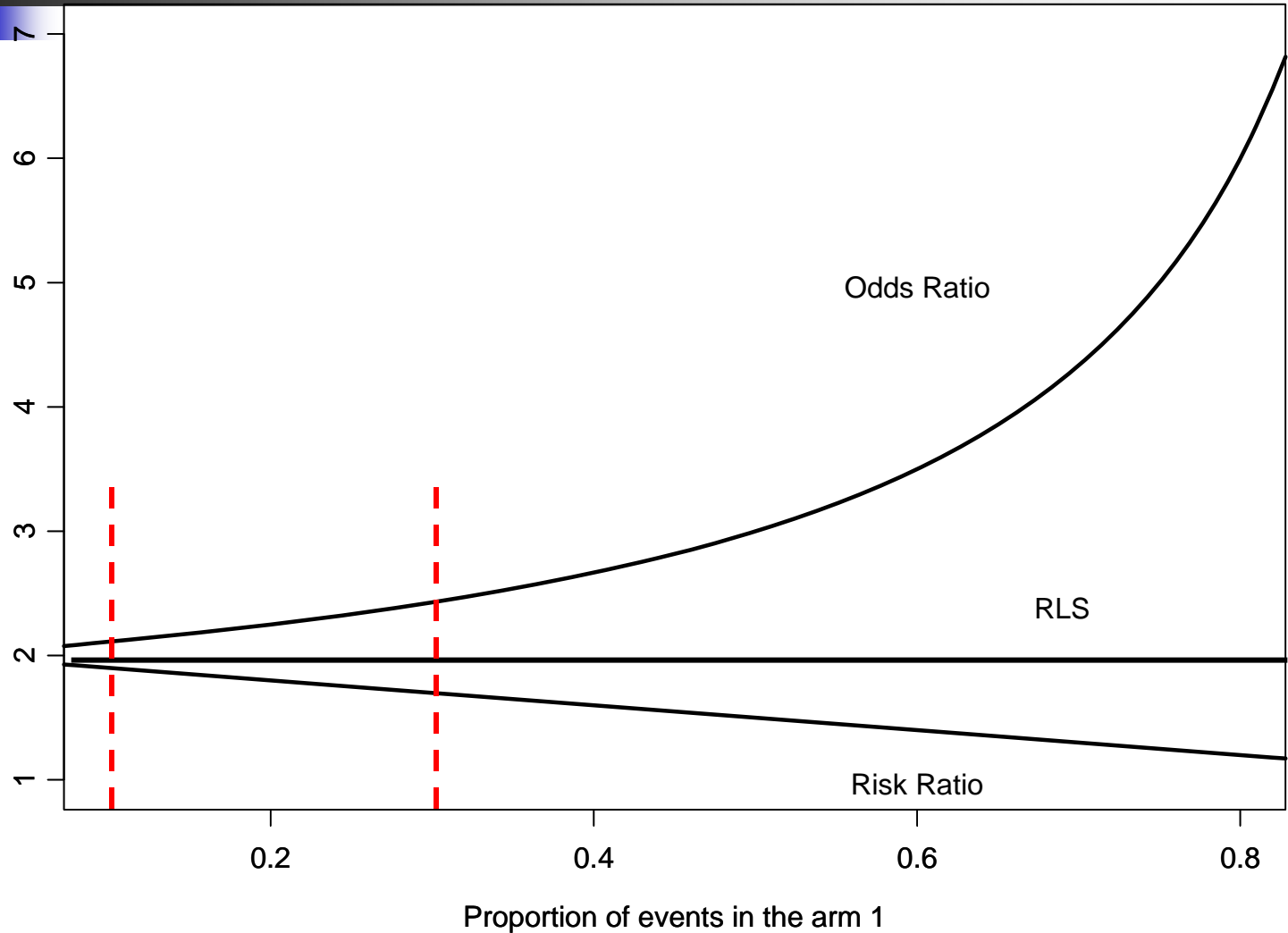
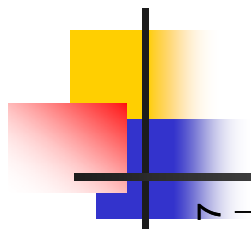
Comparaison des 3 types de méta-analyses par une étude de simulations

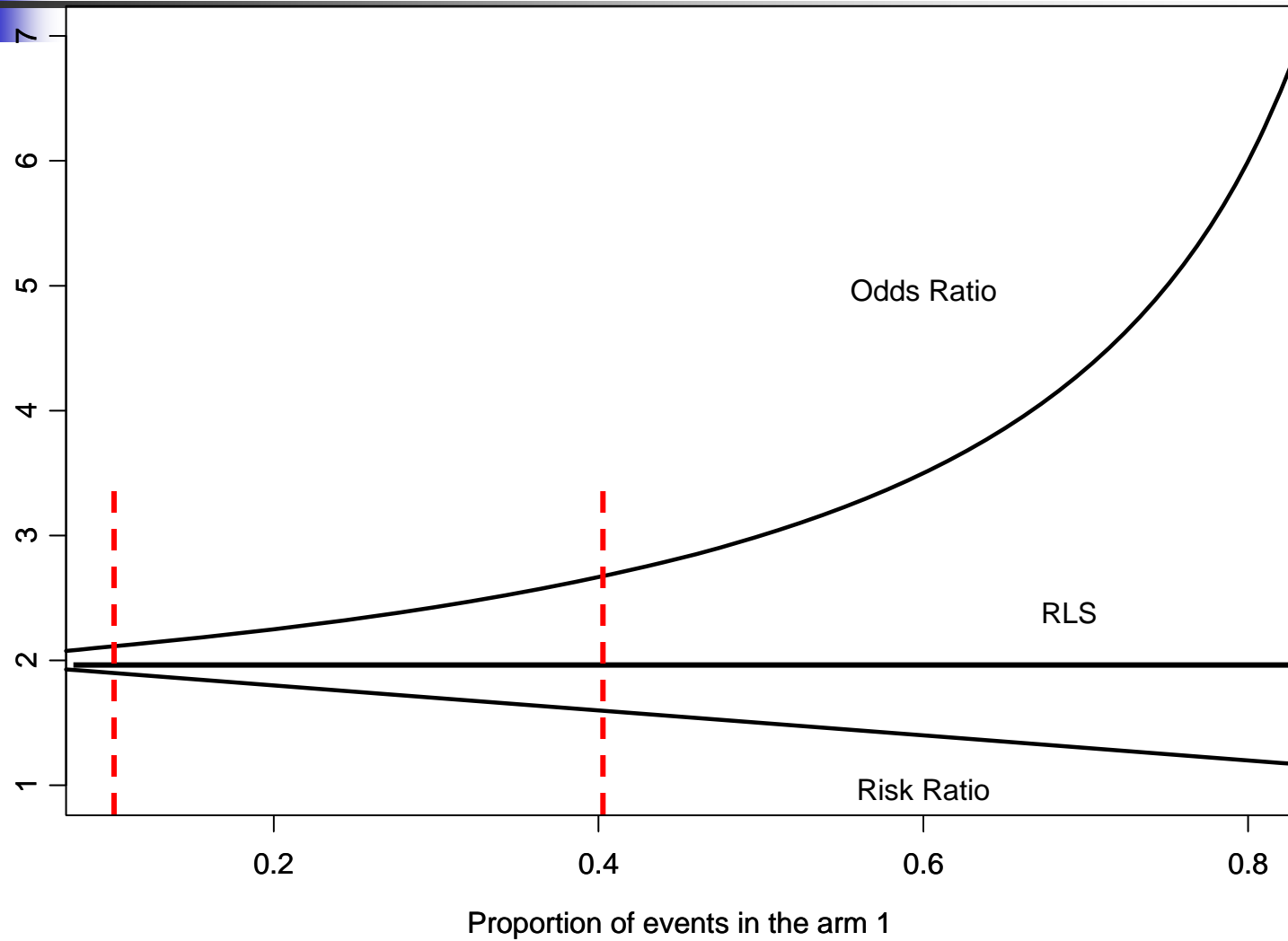
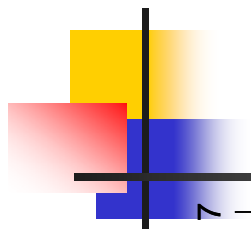


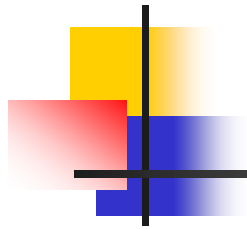
- Critères de comparaison:
 - % de simulations avec une forte hétérogénéité (p Cochran < 0.10 ou $I^2 > 50\%$)
 - Biais relatif: écart en % entre le logarithme de l'estimation et le logarithme du vrai HR
 - % de « couverture »: % de simulations pour lesquelles l'IC95% contient le vrai HR

- 
-
- Scénarios de méta-analyses explorés (45 scénarios):
 - Nombre d'études: 10
 - HR=1.5, 2 et 4
 - N=50, 100 et 300 par bras (bras équilibrés) dans chaque étude
 - Range de % d'événements dans le bras 1 est tiré au hasard dans un intervalle: [10%-20%], [10%-30%], ..., [10%-70%]









■ Génération des données pour un scénario donné:

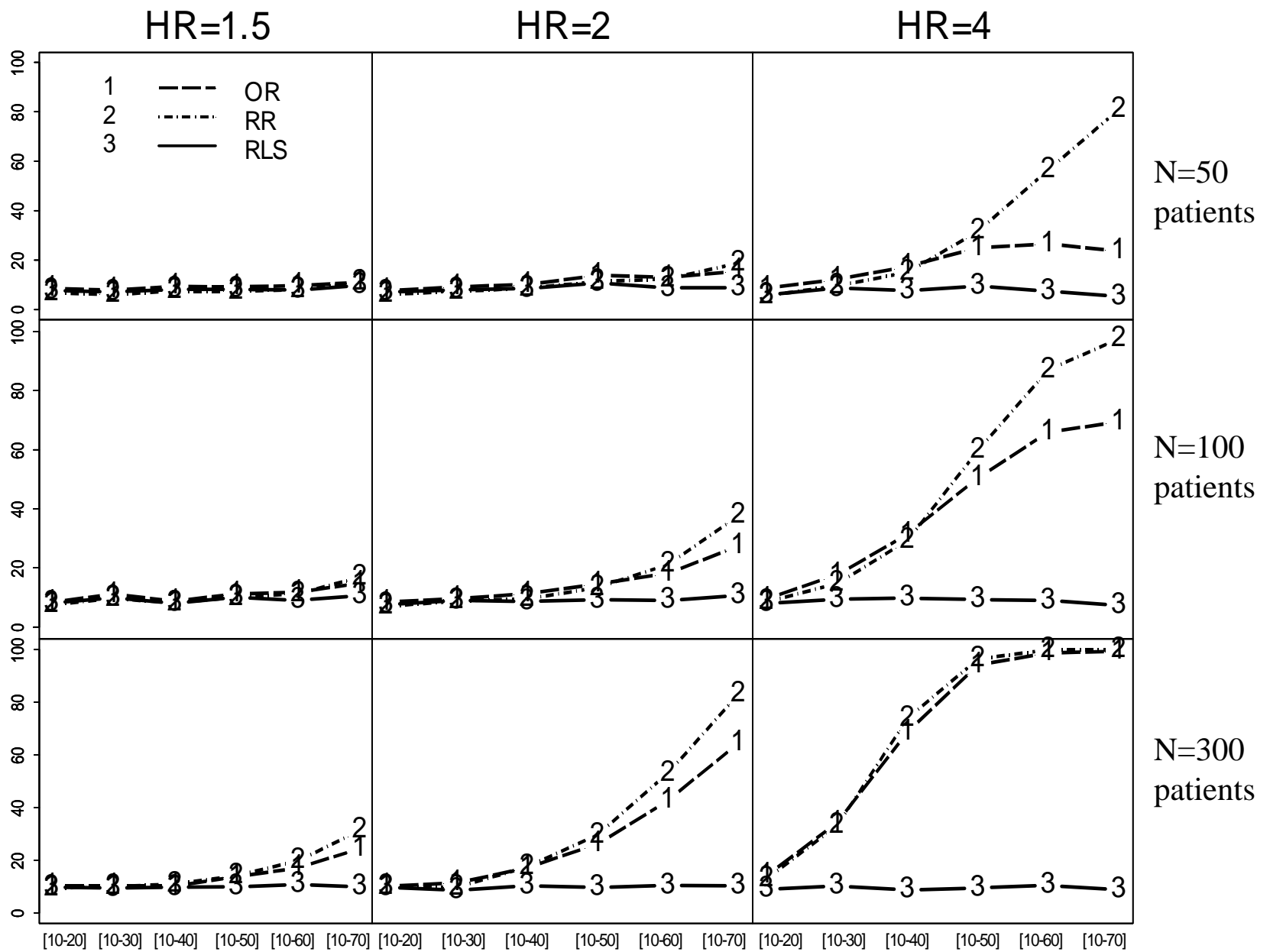
- On génère 10 études.
- Pour chaque étude:
 - On tire au hasard la proportion d'événements dans le bras 1 dans l'intervalle fixé ([10%-20%], [10%-30%],...)
 - On tire au hasard la présence de l'événement chez chaque sujet du bras 1 selon une loi binomiale de paramètre la proportion précédente
 - On déduit la proportion d'événements dans le bras 2 à partir du HR et de la proportion d'événements dans le bras 1
 - On tire au hasard la présence de chaque sujet du bras 2 selon une loi binomiale de paramètre la proportion précédente
 - On reconstruit pour chaque étude le tableau 2×2
- On répète l'opération 2000 fois pour avoir un échantillon de 2000 méta-analyses pour chaque scénario



- Résultats des simulations

- % de simulations avec une forte hétérogénéité (I-square > 50% ou $p < 0.10$ au test de Cochran)

Percentage of simulations with heterogeneity



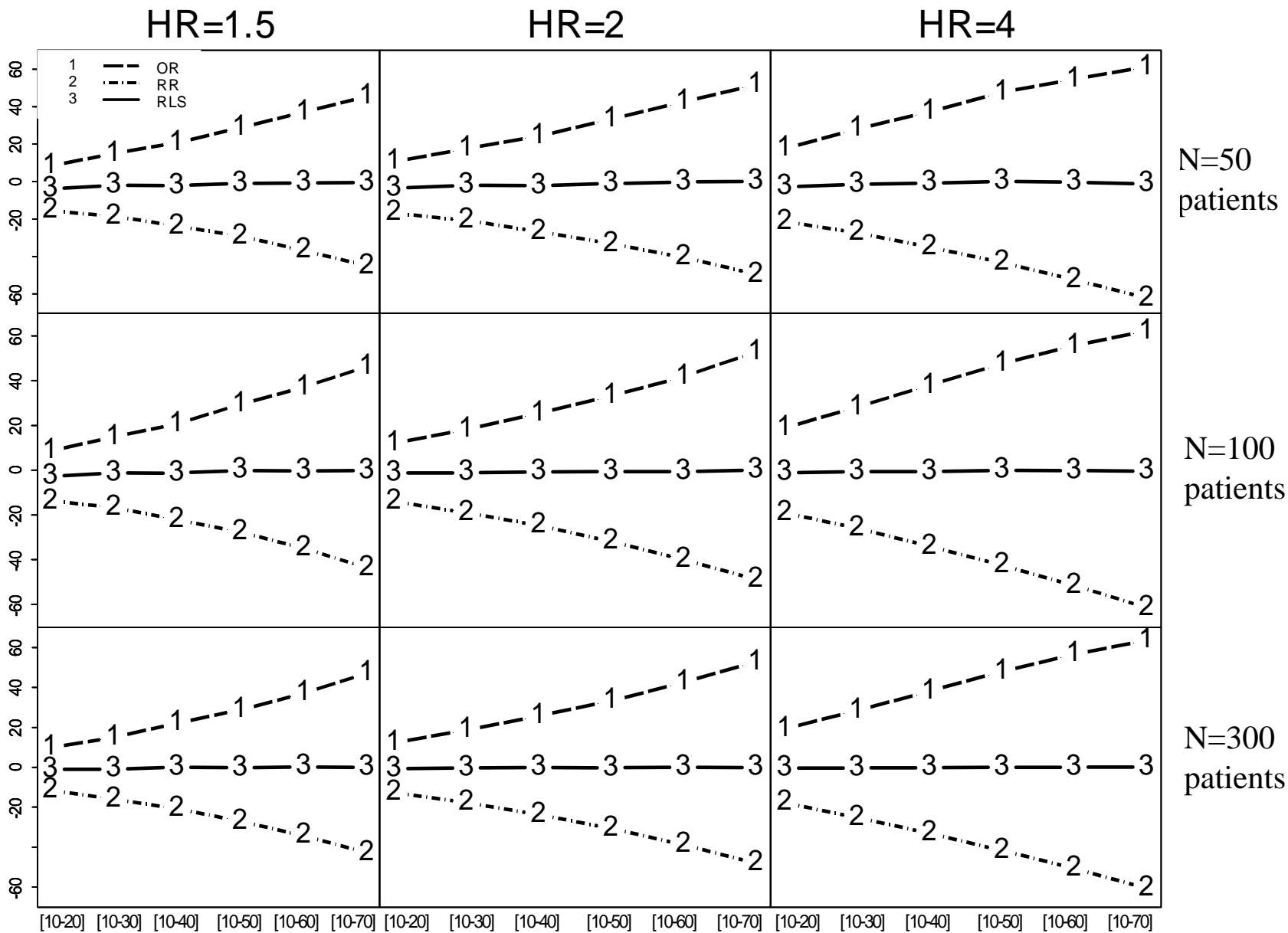
Percentage of events in arm 1



- Résultats des simulations

- % de simulations avec une forte hétérogénéité (I-square > 50% ou $p < 0.10$ au test de Cochran)
- Biais relatif: écart en % entre le logarithme de l'estimation et le logarithme du vrai HR

Relative bias in % (median)



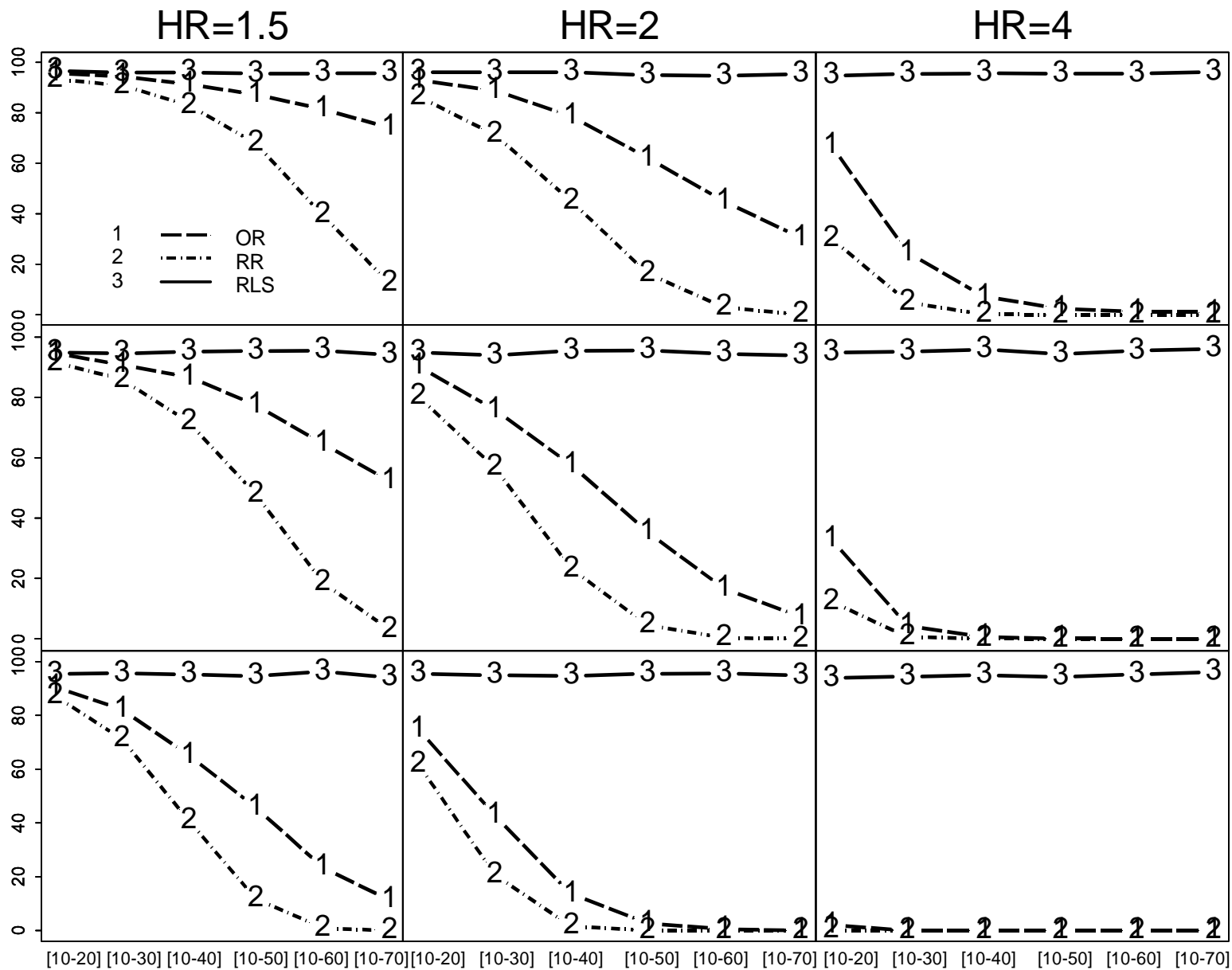
Percentage of events in arm 1



■ Résultats des simulations

- % de simulations avec une forte hétérogénéité ($I^2 > 50\%$ ou $p < 0.10$ au test de Cochran)
- Biais relatif: écart en % entre l'estimation et le vrai HR (logarithme)
- % de « couverture »: % de simulations pour lesquelles l'IC95% contient le vrai HR

Percentage of coverage



N=50 patients

N=100 patients

N=300 patients

Percentage of events in arm 1



- Résumé des résultats

- La méta-analyse de RLSs se comporte bien pour les 3 critères
- Méta-analyse d'ORs et RRs:
 - Pour un effet faible, pas d'hétérogénéité mais biais
 - L'hétérogénéité augmente avec l'effet
 - L'hétérogénéité et le biais augmentent avec l'hétérogénéité des durées de suivi
 - Globalement, le RR se comporte moins bien que l'OR



Exemple

Palliative chemotherapy for advanced colorectal cancer: systematic review and meta-analysis

Colorectal Cancer Collaborative Group

BMJ 2000;321:531-5

Objectives To determine the benefits and harms of palliative chemotherapy in patients with locally advanced or metastatic colorectal cancer and to compare the outcomes for elderly and younger patients.

Studies All randomised controlled trials comparing palliative chemotherapy with supportive care in patients with advanced colorectal cancer that were identified by computerised and hand searches of the literature, scanning references, and contacting investigators.

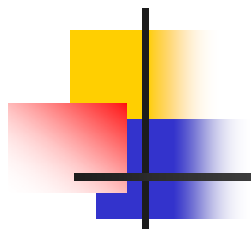
Main outcome measures Survival, disease progression, quality of life, and toxicity.

- 7 études

- Méta-analyse sur la survie

- Dans l'article initial: méta-analyse de HRs

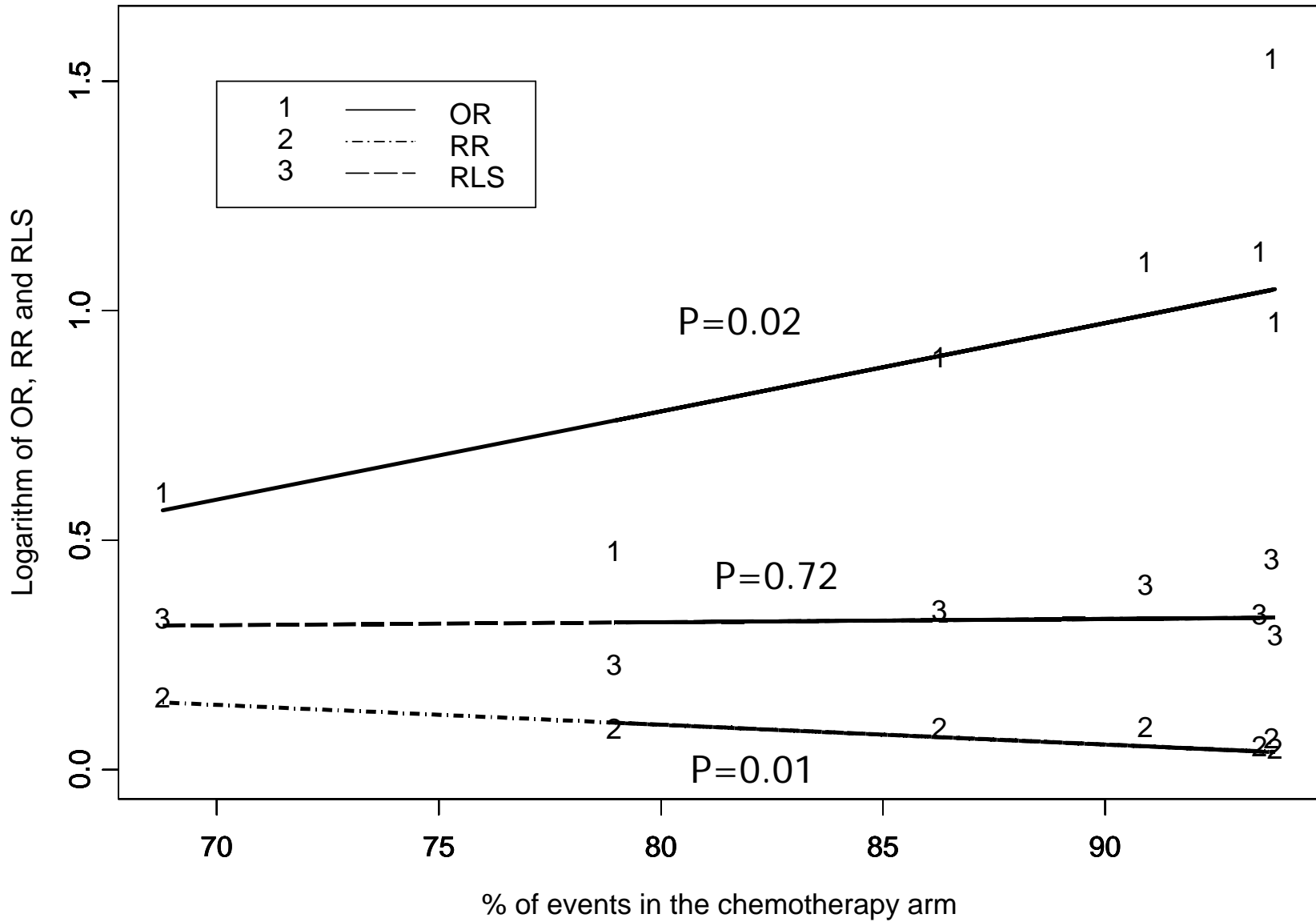
Studies	Palliative chemotherapy		Supportive care	
	N	% Evt	N	% Evt
Cunningham	189	68.8%	90	80.0%
Hunt	19	78.9%	42	90.5%
Allen Marsh	51	86.3%	49	93.9%
Glimelius	11	90.9%	10	100.0%
Nordic group	92	93.5%	90	97.8%
Hafstrom	32	93.8%	28	100.0%
Rougier	81	93.8%	82	97.6%

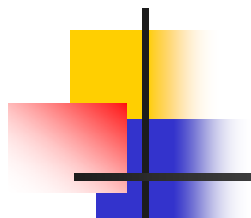


Study	HR*	OR	RR	RLS
Cunningham	1.82 [1.19-2.78]	1.82 [1.00-3.31]	1.16 [0.74-1.83]	1.38 [1.01-1.89]
Hunt	1.54 [0.74-3.23]	1.60 [0.39-6.50]	1.09 [0.55-2.13]	1.25 [0.64-2.46]
Allen Marsh	1.54 [0.88-2.70]	2.44 [0.59-10.03]	1.09 [0.36-3.27]	1.41 [0.83-2.37]
Glimelius	2.22 [0.63-7.69]	3.00 [0.11-82.40]	1.09 [0.07-16.5]	1.49 [0.48-4.62]
Nordic group	1.25 [0.84-1.85]	3.07 [0.60-15.63]	1.05 [0.27-4.12]	1.39 [0.88-2.20]
Hafstrom	2.17 [1.03-4.76]	4.67 [0.21-101.56]	1.06 [0.07-16.7]	1.57 [0.69-3.57]
Rougier	1.39 [0.92-2.13]	2.63 [0.50-14.97]	1.04 [0.26-4.09]	1.33 [0.83-2.15]

*: résultats donnés dans l'article

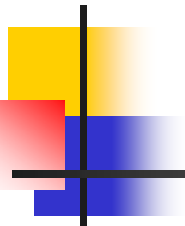
Outcomes of the 7 studies





Study	HR	OR	RR	RLS
Cunningham	1.82 [1.19-2.78]	1.82 [1.00-3.31]	1.16 [0.74-1.83]	1.38 [1.01-1.89]
Hunt	1.54 [0.74-3.23]	1.60 [0.39-6.50]	1.09 [0.55-2.13]	1.25 [0.64-2.46]
Allen Marsh	1.54 [0.88-2.70]	2.44 [0.59-10.03]	1.09 [0.36-3.27]	1.41 [0.83-2.37]
Glimelius	2.22 [0.63-7.69]	3.00 [0.11-82.40]	1.09 [0.07-16.5]	1.49 [0.48-4.62]
Nordic group	1.25 [0.84-1.85]	3.07 [0.60-15.63]	1.05 [0.27-4.12]	1.39 [0.88-2.20]
Hafstrom	2.17 [1.03-4.76]	4.67 [0.21-101.56]	1.06 [0.07-16.7]	1.57 [0.69-3.57]
Rougier	1.39 [0.92-2.13]	2.63 [0.50-14.97]	1.04 [0.26-4.09]	1.33 [0.83-2.15]
<i>Pooled results</i>	<i>1.54</i> <i>[1.32-1.79]*</i>	<i>2.04</i> <i>[1.29-3.24]</i>	<i>1.06</i> <i>[1.02-1.10]</i>	<i>1.39</i> <i>[1.20-1.60]</i>

*: résultat dans l'article



■ ORs et RRs des études

- Les ORs (RRs) augmentaient (diminuaient) significativement avec la proportion de décès dans le groupe chimiothérapie
- Négativement corrélés (peut être un moyen de savoir s'il faut utiliser le RLS)

■ Résultat poolé

- Toutes les méta-analyses aboutissent à des résultats significatifs
- Mais avec des effets très différents: $RR=1.06$ et $OR=2.04$
- Le RLS poolé est 1.39, relativement proche du HR poolé (1.54)
- Absence d'hétérogénéité (I-square=0%)



Conclusions

- Méta-analyses de ORs et de RRs dans un contexte de données de survie
 - Hétérogénéité causée par des longueurs de suivi différentes (artefact) lorsque l'effet est fort (HR=2 ou plus): artefact dû à la définition des ORs et des RRs
 - Cette hétérogénéité artificielle peut perturber la détection des vrais facteurs d'hétérogénéité
 - Biais causé par des longueurs de suivi différentes même lorsque l'effet est faible
 - Sur-estimation de l'effet dans les méta-analyses de ORs
 - Sous-estimation de l'effet dans les méta-analyses de RRs
 - Si les RRs et les ORs sont négativement corrélées (Spearman: $\rho = -0.43$), alors on peut suspecter un effet des longueurs de suivi (outil diagnostique pratique)



- Avantages de la méta-analyse du RLS

- Adaptée aux données de survie puisque RLS est un estimateur du HR
- Non sensible aux différences dans les longueurs de suivi:
 - Absence de biais, absence d'hétérogénéité
- A partir de tableaux 2*2
- Ne nécessite pas d'explicitier les longueurs de suivi
- Simple d'utilisation

- Hypothèse du modèle:

- Données de survie à hasard proportionnel