



Maria Pia Sormani

# **METODI DI ANALISI E DISEGNI DI STUDIO**

# Disegni di superiorità

---

Il disegno classico dei trial clinici è un disegno di **superiorità**:

Si vuole dimostrare se il farmaco A è meglio del farmaco B

- Definizione di «meglio» su che endpoint (endpoint primario dello studio)
- Definizione di cosa si intende per «meglio» 

Differenza minima di interesse clinico



## Disegni di superiorità

---

Esempio: trial nella SM che valuta se il farmaco A è superiore al placebo

- Definizione di «meglio» su che endpoint (endpoint primario dello studio)
  - Il farmaco A è meglio del placebo se riduce il relapse rate in un anno rispetto al placebo
- Definizione di cosa si intende per «meglio»
  - Differenza minima di interesse clinico: Riduzione del relapse rate almeno del 30%



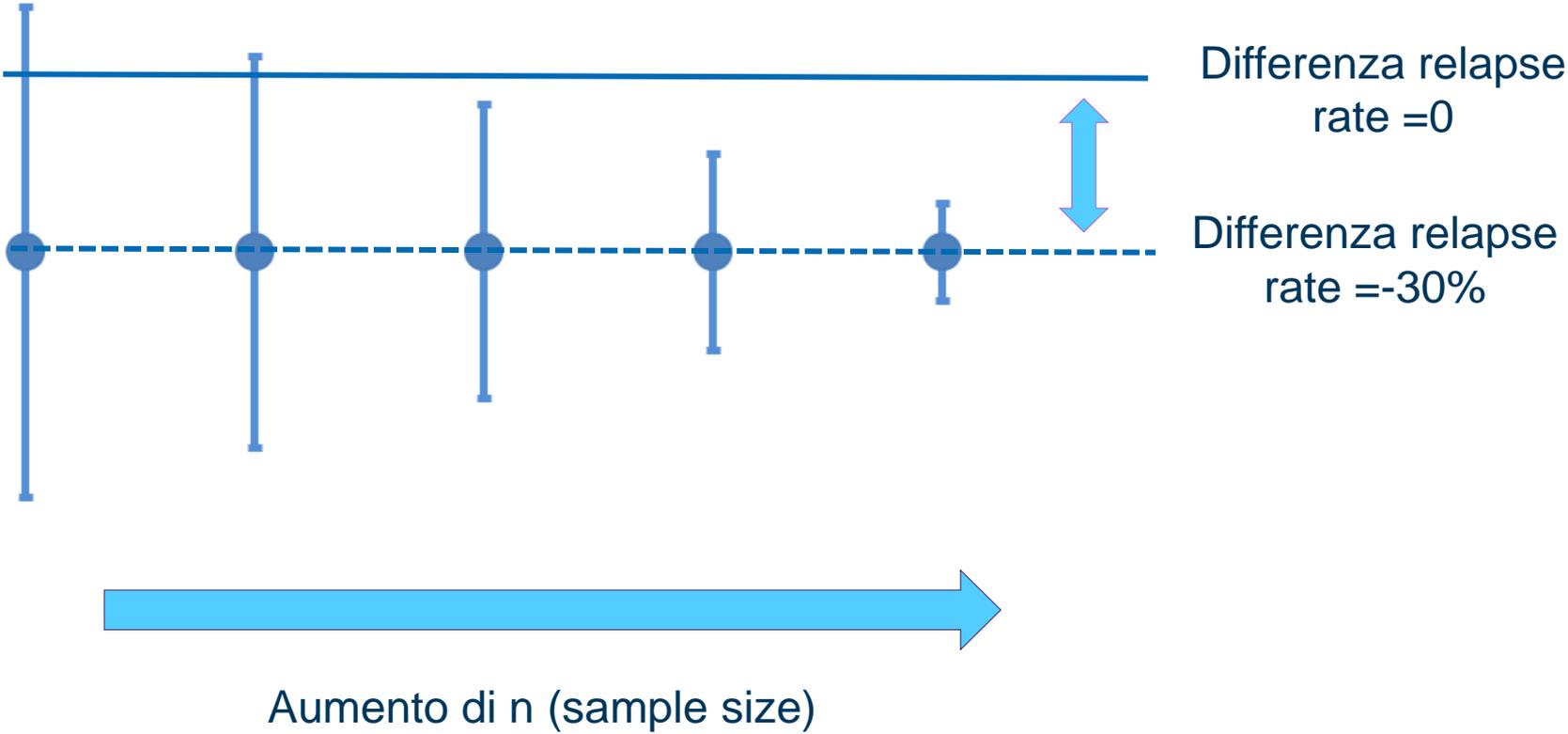
## Disegni di superiorità

---

Esempio: trial nella SM che valuta se il farmaco A è superiore al placebo

- Definizione di «meglio» su che endpoint (endpoint primario dello studio)
  - Il farmaco A è meglio del placebo se riduce il relapse rate in un anno rispetto al placebo
- Definizione di cosa si intende per «meglio»
  - Differenza minima di interesse clinico: Riduzione del relapse rate almeno del 30%

# Disegni di superiorità – sample size



# Disegni di superiorità – interpretazione del risultato

- ▶ Diminuzione del relapse rate osservata: - 29% (95% CI= -12%, -42%)

P=0.01

- ▶ La probabilità che questa diminuzione osservata nel braccio trattato, data la dimensione del campione e le caratteristiche della variabile in studio, sia casuale è minore dell'1%
- ▶ Rifiuto l'ipotesi nulla che il farmaco A sia uguale al placebo e dico che il farmaco A ha ridotto il relapse rate

- ▶ Diminuzione del relapse rate osservata: - 29% (95% CI= +6%, -52%)

P=0.16

- ▶ La probabilità che questa diminuzione osservata nel braccio trattato, data la dimensione del campione e le caratteristiche della variabile in studio, sia casuale è del 16%
- ▶ Non posso rifiutare l'ipotesi nulla che il farmaco A sia uguale al placebo

# LAQUINIMOD effect

ALLEGRO Study:

Placebo (556 pts) vs Laquinimod (550 pts)

Primary endpoint = Relapse rate over 2 years

Placebo: ARR=0.39

Laquinimod: : ARR=0.30

**Laquinimod reduces relapse rate of 23%, p=0,002**

**POSITIVE STUDY**

---

RR=0.77, 95% C.I. for RR: [0.65, 0.91]

# LAQUINIMOD effect

BRAVO Study:

Placebo (450 pts) vs Laquinimod (434 pts)

Primary endpoint = Relapse rate over 2 years

Placebo: ARR=0.34

Laquinimod: : ARR=0.28

**Laquinimod reduces relapse rate of 18%, p=0,075**

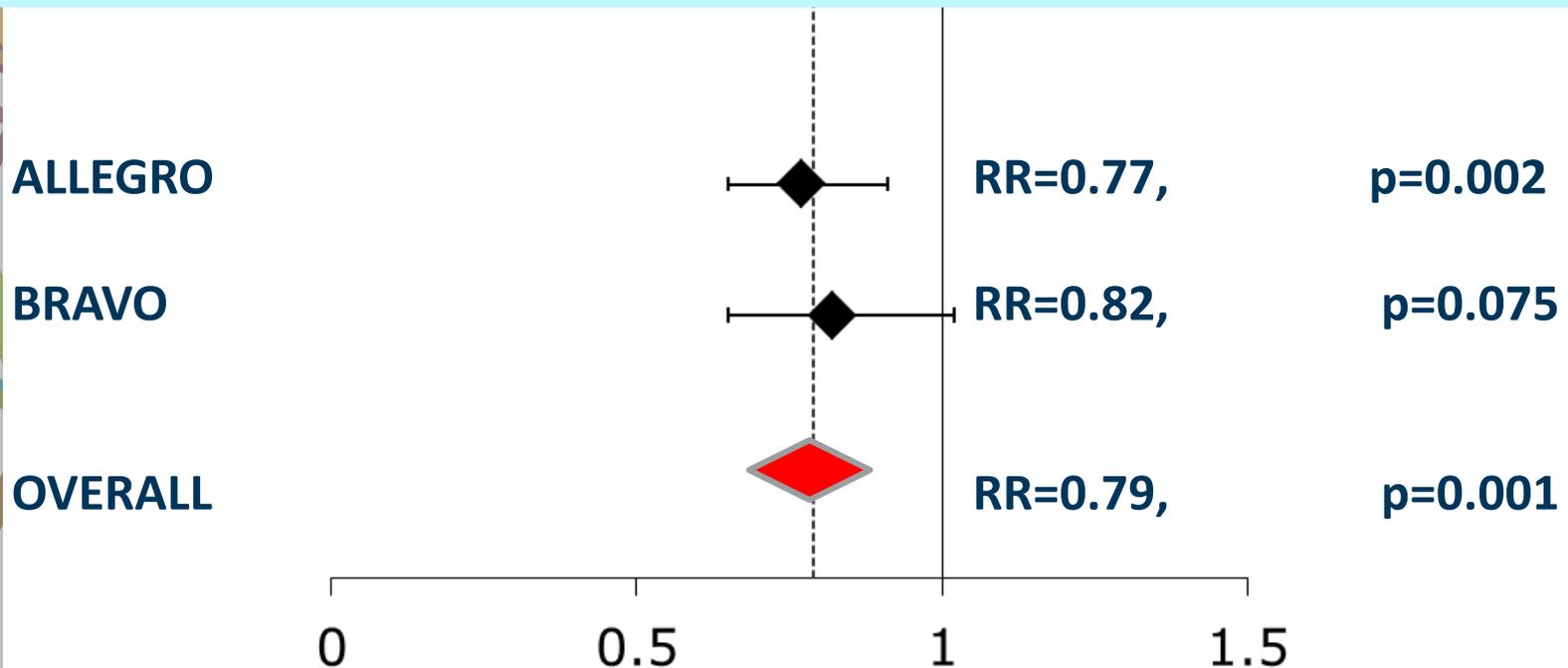
**NEGATIVE STUDY**

---

RR=0.82, 95% C.I. for RR: [0.66, 1.02]

# LAQUINIMOD effect

**LAQUINIMOD EFFECT = RELAPSE REDUCTION OF 21%,  
p=0.001**



**EFFECT DIFFERENCE (interaction test), p=0.80**



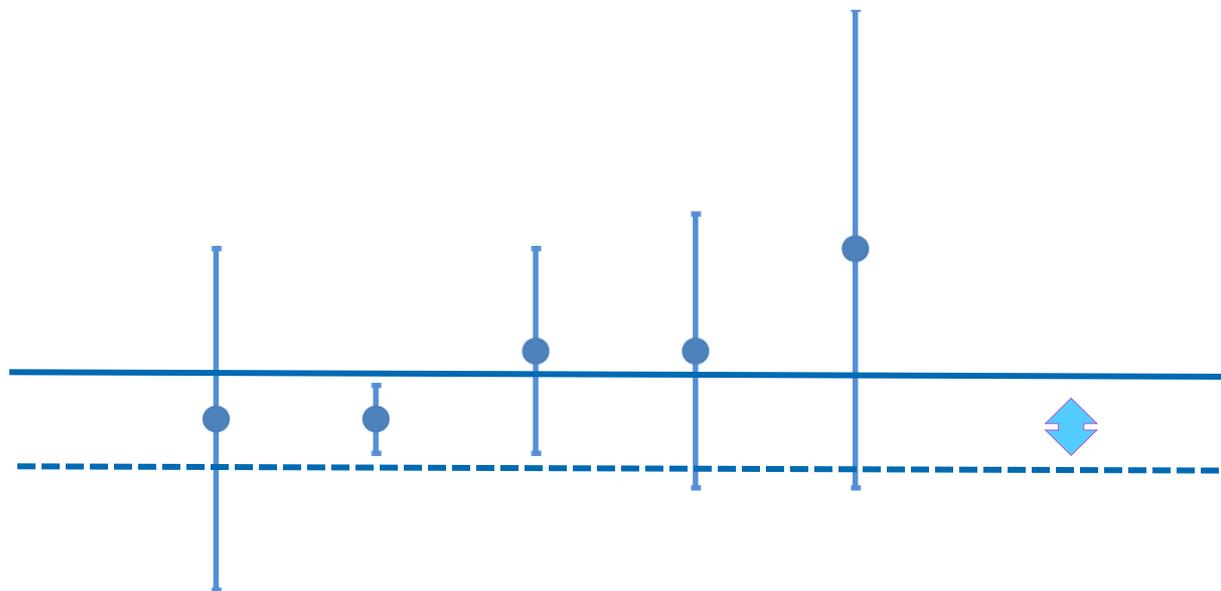
## Disegni di non-inferiorità

---

Il disegno di **non-inferiorità** è utile quando per immettere un farmaco nella pratica clinica basta dimostrare che non è meno efficace dello standard

- Ha meno effetti collaterali
- Ha un miglior metodo di assunzione
- Ha dei vantaggi diversi dall'efficacia (costi)

# Disegni di non inferiorità



Differenza relapse  
rate = 0

Differenza relapse  
rate = margine di  
non-inferiorità

# Disegni di non-inferiorità – interpretazione del risultato

---

- ▶ Diminuzione del relapse rate osservata: - 7% (95% CI= +5%, -14%)
  - ▶ L'intervallo di confidenza mi dice che la differenza vera sta tra +5% e -14%. Siccome -14% non è ritenuta una differenza clinicamente rilevante dico che il farmaco A non è inferiore al farmaco B
- ▶ Diminuzione del relapse rate osservata: - 7% (95% CI= +12%, -22%)
  - ▶ L'intervallo di confidenza mi dice che la differenza vera sta tra +12% e -22%. Siccome -22% è ritenuta una differenza clinicamente rilevante non posso escludere che il farmaco A non sia inferiore al farmaco B



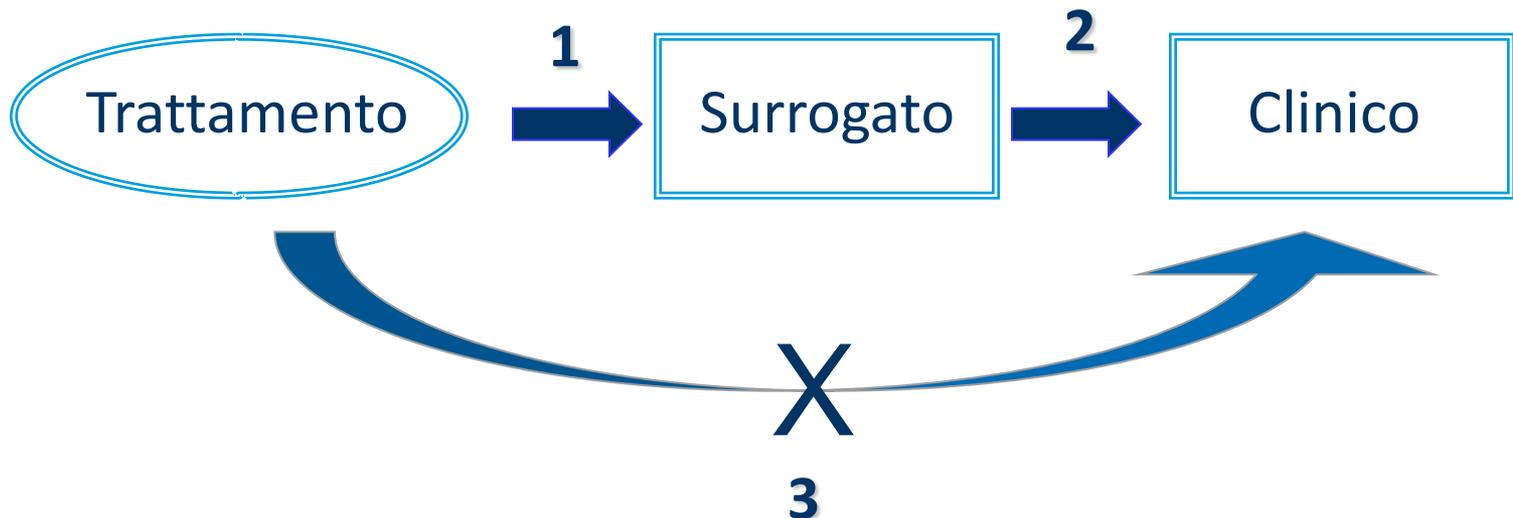
## *Gli endpoints surrogati*

---

Una variabile clinica, strumentale o di laboratorio che puo' essere utilizzata come endpoint primario in uno studio clinico poiche' permette di valutare/stimare gli effetti del trattamento sull'endpoint primario 'naturale

# *Gli endpoint surrogati*

L'effetto del trattamento sull'endpoint clinico deve essere interamente assorbito dall'endpoint surrogato



# MRI as a surrogate endpoint

---

The correlation between conventional MRI markers and the clinical course of the disease is poor (“clinico-radiological paradox”)

The role of conventional MRI lesions as surrogate markers for monitoring the anti-inflammatory effect of treatments is still controversial

# MRI as a surrogate outcome

---

## Low correlation MRI lesions-relapses:

Held et al, Neurology 2005:

- MRI lesions did not result as predictors of relapses in a multivariate model (SLC database, 306 pts)

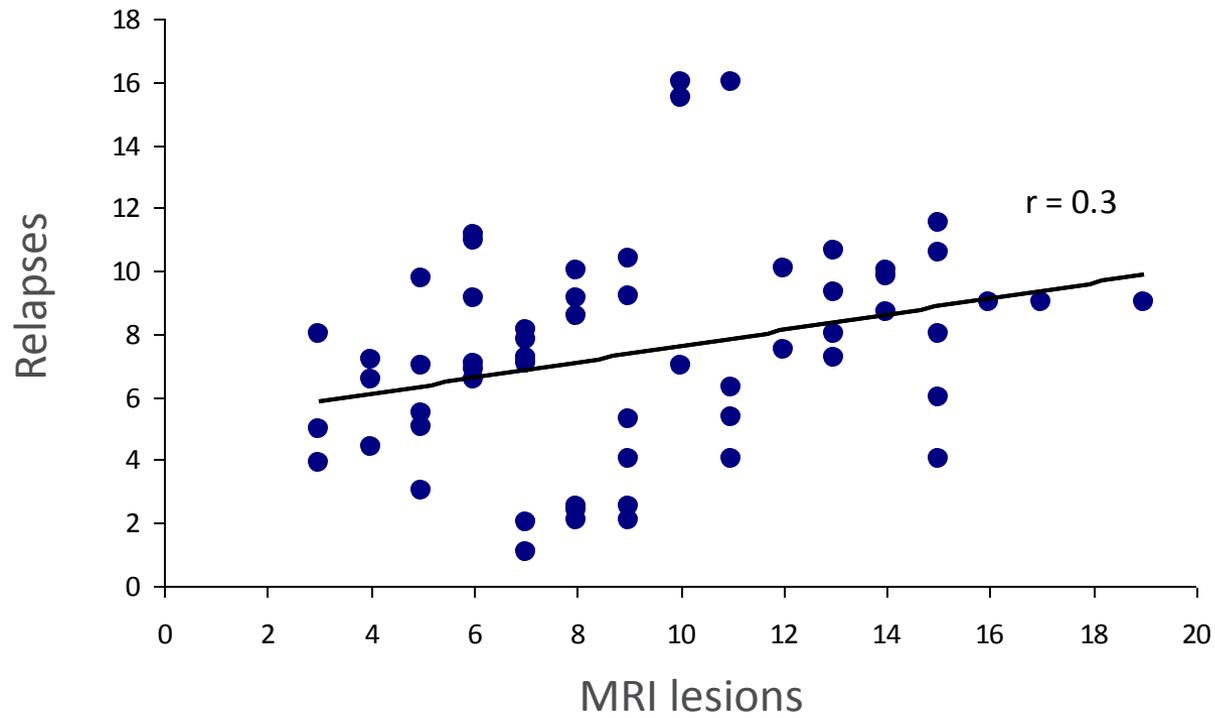
Petkau et al, Mult Scler 2008:

- the project of validating MRI lesions as surrogate for relapses was stopped since the pre-requisite of correlation between MRI lesions and relapses was not met (SLC database, 208 pts)

Daumer et al, Neurology 2009:

- A retrospective analysis on 31 placebo arms of RCT showed no correlation between Gad lesions and relapses (SLC database, 409 pts)

# Correlations



MRI center 1

MRI center 2

# Validazione degli endpoint surrogati

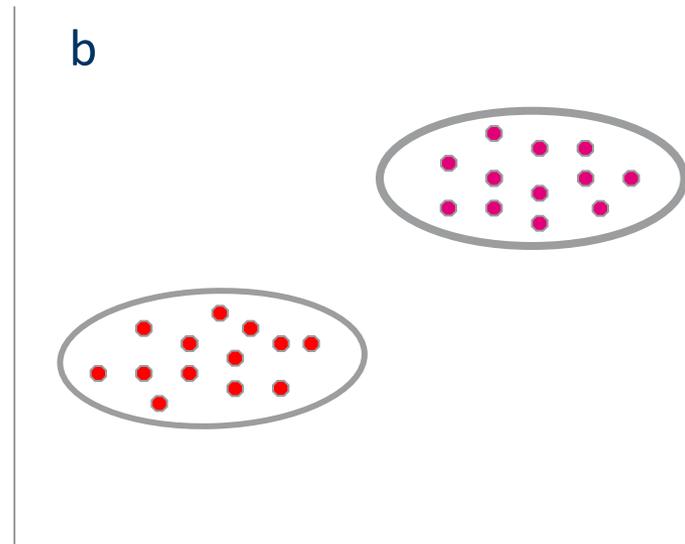
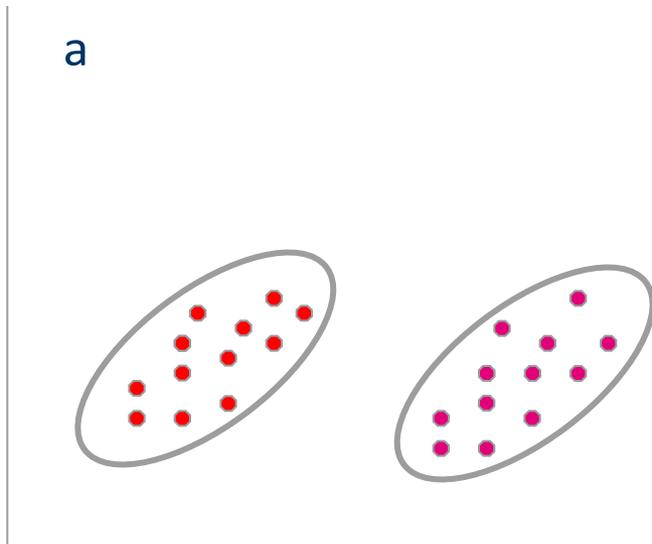
Single patient perspective

Clinical trial perspective

INDIVIDUAL LEVEL

TRIAL LEVEL

Effect on th clinical Y

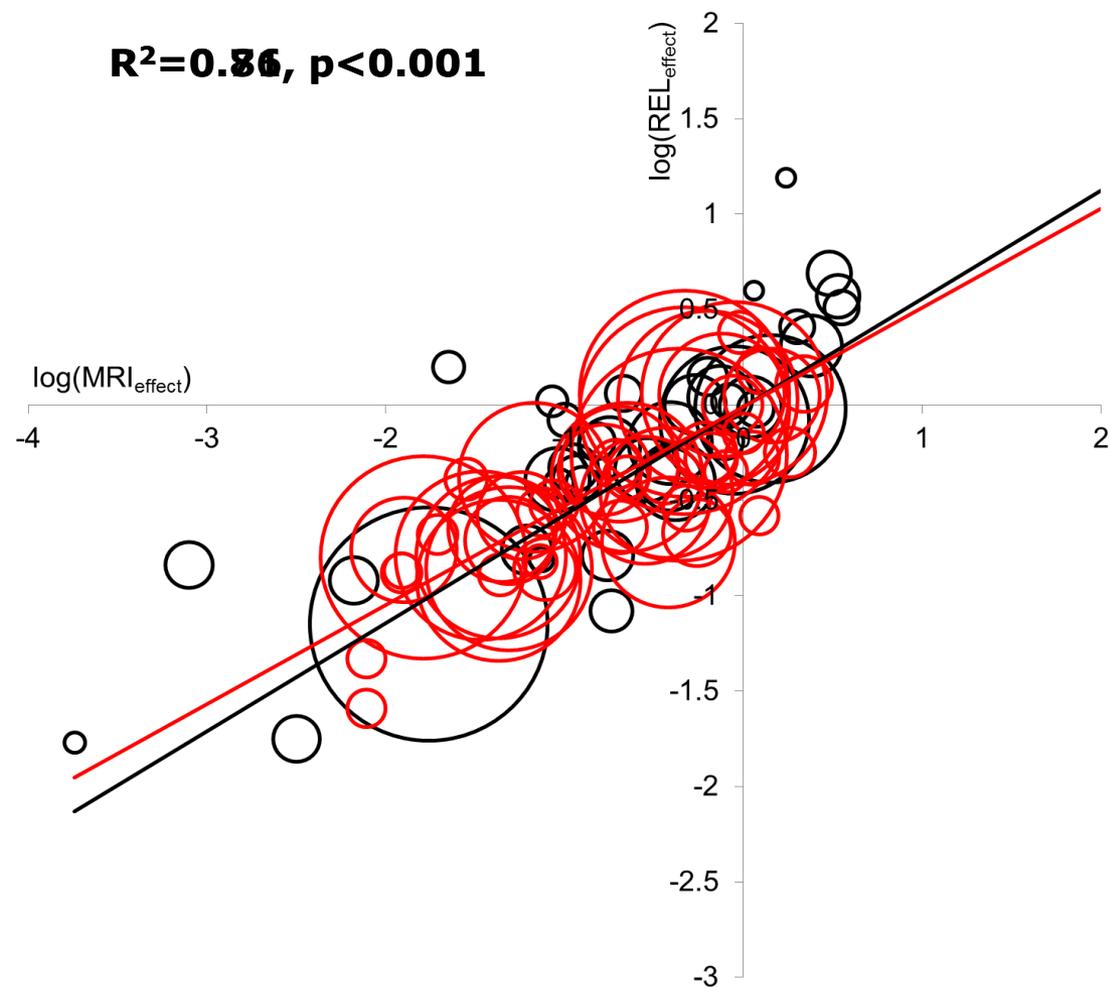


Effect on the surrogate X

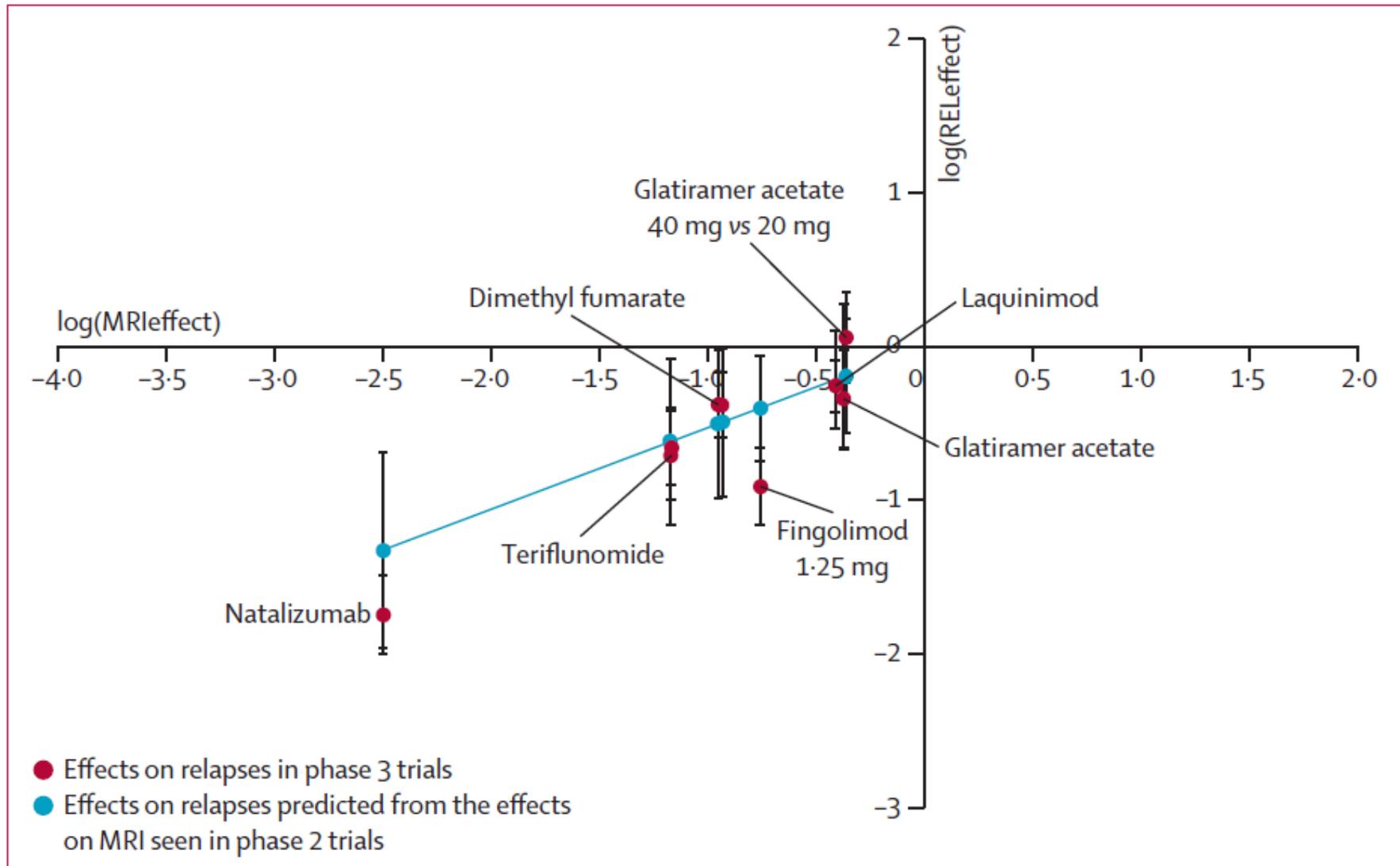
Effect on the surrogate X

- TRIAL 1
- TRIAL 2

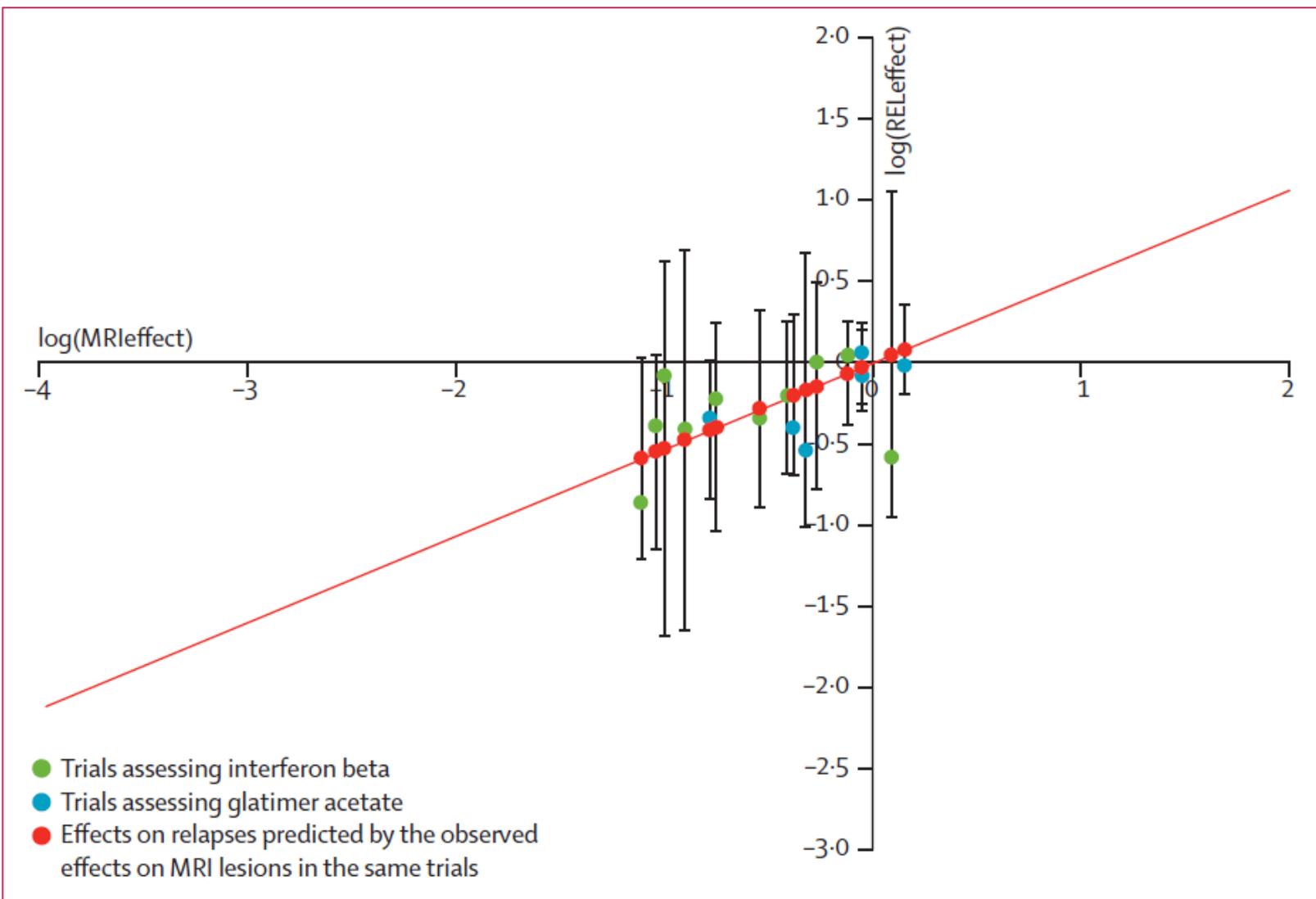
# Le lesioni MRI come surrogato per le relapse



# MRI lesions as surrogate for relapses (TRIAL LEVEL)

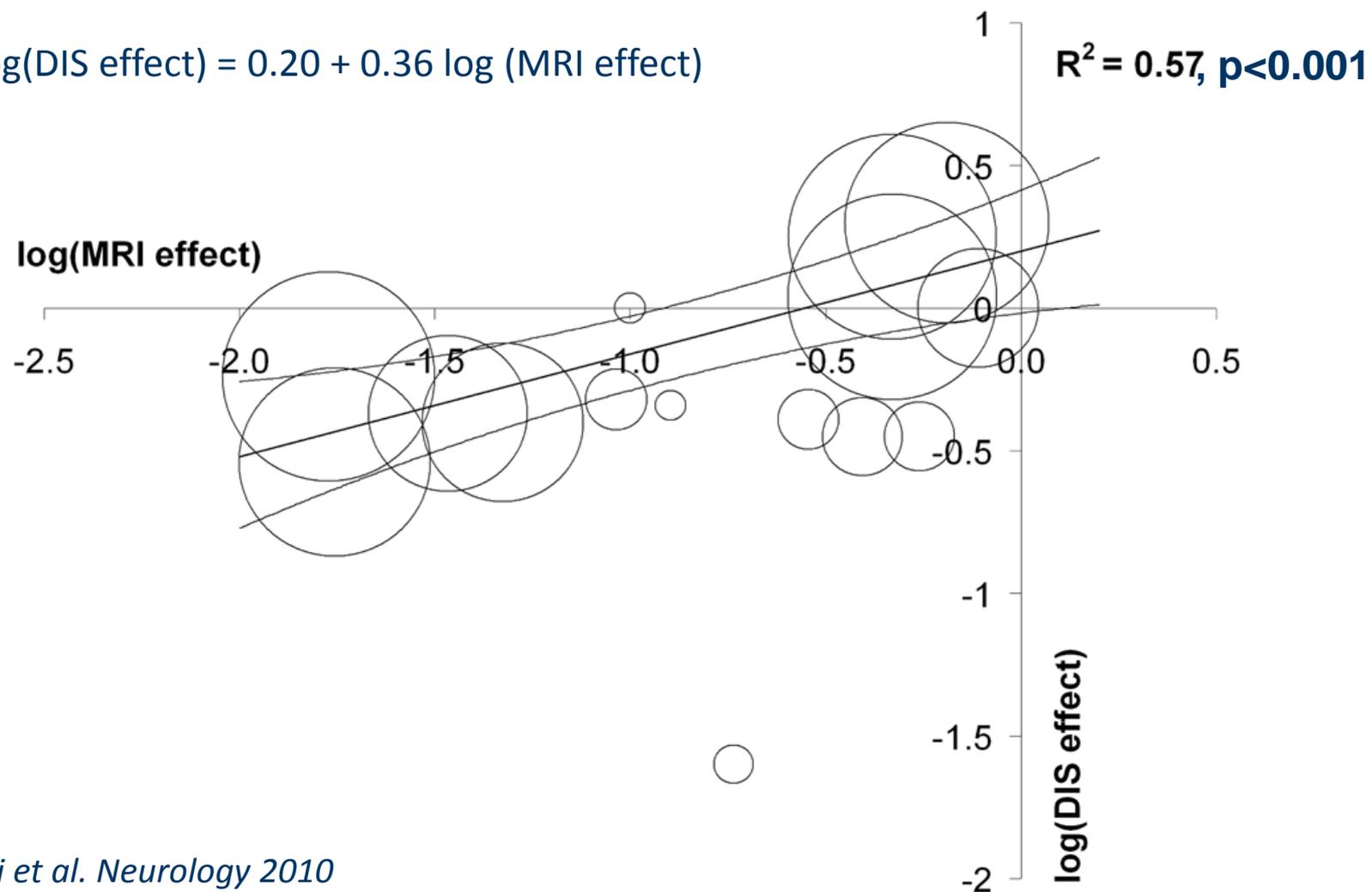


# MRI lesions as surrogate for relapses (TRIAL LEVEL)



# Le lesioni MRI come surrogato per la disability

$$\log(\text{DIS effect}) = 0.20 + 0.36 \log(\text{MRI effect})$$





## *Implicazioni*

---

- ▶ Questa osservazione fornisce una giustificazione evidence-based alla scelta degli endpoint di RM negli studi di fase II
- ▶ Questa osservazione non è sufficiente a permettere l'utilizzo degli endpoint di RM come endpoints primari negli studi di fase III
- ▶ La stima dell'effetto di un trattamento sulle lesioni di RM può dare indicazioni sul disegno ottimale degli studi di fase III
- ▶ E' necessaria una validazione individual-level per rafforzare l'idea di un rapporto di causa-effetto tra lesioni RM ed endpoint clinici



# Le analisi per sottogruppi

---

Confronto tra i due trattamenti in sottogruppi di pazienti definiti in base a particolari fattori (prognostici) pre-randomizzazione

SCOPI :

A) Studio Negativo

- ▶ valutare se esistono differenze limitate a pazienti con
- ▶ caratteristiche particolari

B) Studio Positivo

- ▶ valutare se l'efficacia del trattamento migliore è
- ▶ maggiore (presente solo) in particolari pazienti

C) Ipotesi Pre-esistente

- ▶ A priori si ritiene che l'efficacia del trattamento
- ▶ sperimentale sarà maggiore in particolari pazienti



# Le analisi per sottogruppi

---

## Problemi :

- ▶ molteplicita' dei tests
- ▶ perdita di potenza e potenza variabile nei singoli sottogruppi (studio positivo)
- ▶ difficoltà' interpretative
- ▶ analisi prestabilite vs analisi suggerite dai risultati



# Le analisi per sottogruppi

---

## **Metodo più usato (sbagliato) :**

Il test di significativita' viene ripetuto in ogni sottogruppo al livello di significativita' convenzionale



## Le analisi per sottogruppi

Esempio: trial con 120 pazienti a braccio

---

Braccio A : 60% di risposte

Braccio B : 40% di risposte ,  $p=0.002$

Tre sottogruppi:

< 30 anni:  $p=0.01$

30-50 anni:  $p=0.2$

>50 anni:  $p=0.2$

**Conclusioni?**

# Le analisi per sottogruppi

Esempio: trial con 120 pazienti a braccio

---

Braccio A : 60% di risposte

Braccio B : 40% di risposte ,  $p=0.002$

Tre sottogruppi:

< 30 anni: 80 pz, 48 (60%) vs 32 (40%),  $p=0.01$

30-50 anni: 20 pz, 12 (60%) vs 8 (40%),  $p=0.2$

>50 anni: 20 pz, 12 (60%) vs 8 (40%),  $p=0.2$

**Conclusioni?**

# Le analisi per sottogruppi

Esempio: trial con 120 pazienti a braccio

---

Braccio A : 60% di risposte

Braccio B : 40% di risposte ,  $p=0.002$

Tre sottogruppi:

< 30 anni:  $p=0.07$

30-50 anni:  $p=0.07$

> 50anni:  $p=0.07$

**Conclusioni?**

# Le analisi per sottogruppi

Esempio: trial con 120 pazienti a braccio

---

Braccio A : 60% di risposte

Braccio B : 40% di risposte ,  $p=0.002$

Tre sottogruppi:

< 30anni: 40 pz,      24 (60%) vs 16 (40%),  $p=0.07$

30-50 anni: 40 pz,      24 (60%) vs 16 (40%),  $p=0.07$

>50anni: 40 pz,      24 (60%) vs 16 (40%),  $p=0.07$

**Conclusioni?**

# Le analisi per sottogruppi

## Metodo più usato (sbagliato) :

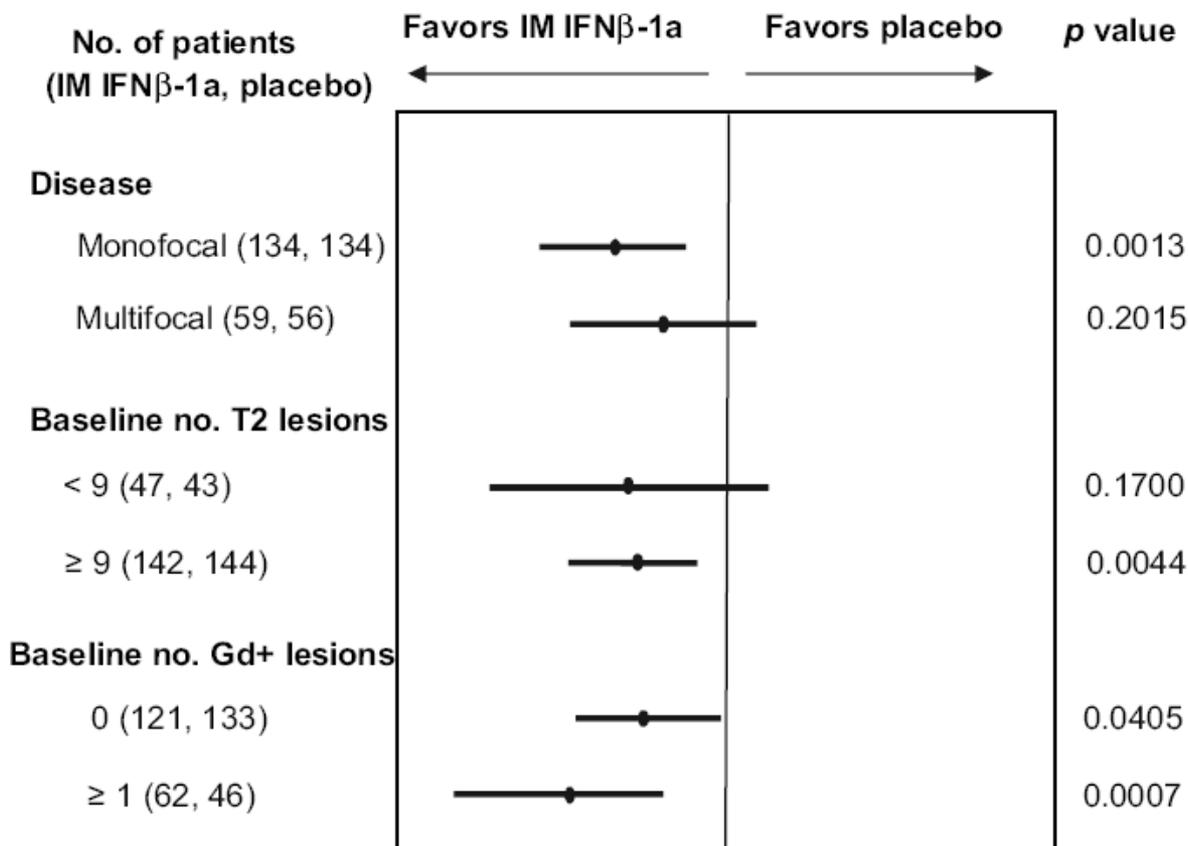
Il test di significativita' viene ripetuto in ogni sottogruppo al livello di significativita' convenzionale

	n	Risk for CDMS in placebo <sup>a</sup>	Risk reduction by IFNB-1b <sup>b</sup>	Hazard ratio $\pm$ 95% CI for treatment effect	p-value
<b>Age</b>					
<30 years	242	60%	51%		0.009
$\geq$ 30 years	226	33%	48%		0.001
<b>Sex</b>					
Male	137	47%	52%		0.017
Female	331	45%	45%		0.002
<b>Steroids</b>					
No steroids	136	38%	62%		0.006
Steroids	332	48%	41%		0.004
<b>CSF<sup>c</sup></b>					
CSF negative	47	36%	28%		0.522
CSF positive	267	49%	45%		0.004
<b>Presenting syndrome<sup>d</sup></b>					
Optic neuritis	116	42%	58%		0.016
Brainstem/cerebellum syndrome	74	43%	68%		0.013
Spinal syndrome	89	51%	51%		0.048

Treatment with placebo better      Treatment with IFNB-1b better

# Le analisi per sottogruppi

## Studio CHAMPS : HR=0.56, p=0.0016 (189 vs 187 pz)



by clinical criteria as having multifocal disease at baseline. IM IFNβ-1a initiated at a first demyelinating attack delayed CDMS in monofocal patients ( $P = 0.0013$ ), patients with or without gadolinium-enhancing lesions ( $P = 0.0007$ ,  $P = 0.0405$ ) and patients with at least nine T2 lesions at baseline ( $P = 0.0044$ ). These data confirm that IM IFNβ-1a delays conversion to CDMS in patients with CIS.



## Le analisi per sottogruppi

---

- ▶ Viene definita come analisi primaria l'analisi sull'intero gruppo
- ▶ Viene introdotta come analisi secondaria l'analisi per alcuni sottogruppi predefiniti se c'è significatività nell'analisi primaria. Questa si basa su un test di interazione
- ▶ Il test di significatività nei vari sottogruppi è privo di senso e produce risultati assurdi



# Le analisi per sottogruppi

## Metodo di analisi corretto: test di interazione

---

Nuova ipotesi nulla: omogeneità dell'effetto in tutti i sottogruppi (l'effetto è lo stesso?)

In qualche sottogruppo esiste evidenza 'significativa' di un effetto diverso?

Scarsa potenza statistica

Suggerimento per nuovi studi o meta-analisi



# Conclusioni

---

- ▶ La lettura dei risultati di un trial clinico va sempre fatta criticamente
- ▶ Un occhio particolare deve andare all'entità degli effetti riportati più che alla loro significatività statistica
- ▶ Un effetto sulle lesioni di RM indica un probabile effetto anche sulle ricadute cliniche e sulla progressione di disabilità a breve termine
- ▶ Le analisi per sottogruppi non vanno condotte analizzando il p value in ogni sottogruppo