

Significancia clínica, significancia estadística y valores p

usos y abusos



Tecnológico de Monterrey
Escuela de Medicina

Carlos A. Cuello

Monterrey NL
Julio 2016

Declaración de conflictos de interés

- Ningún liga con industria farmacéutica
- He recibido o recibo apoyo económico de:
 - ▶ La organización mundial de la alergia
 - ▶ la universidad de McMaster
 - ▶ Cochrane
 - ▶ CONACYT

Agenda

- Qué es el valor P
- Cuáles son sus funciones, críticas y concepciones erróneas
- Intervalos de confianza: definiciones, usos, cómo exponerlos
- Entender términos usados comúnmente como significancia
"estadística" vs "clínica"

valor p

la anécdota de Motyl et al.

- Estudio sobre como las personas con visión política de centro podían discernir entre los tonos de gris mejor que las personas de izquierda o derecha
- P valor en 0.01
- Le invitan a que replique el estudio
- P valor en 0.59



el valor P

- “The P value was never meant to be used the way it’s used today”

anónimo

el valor P

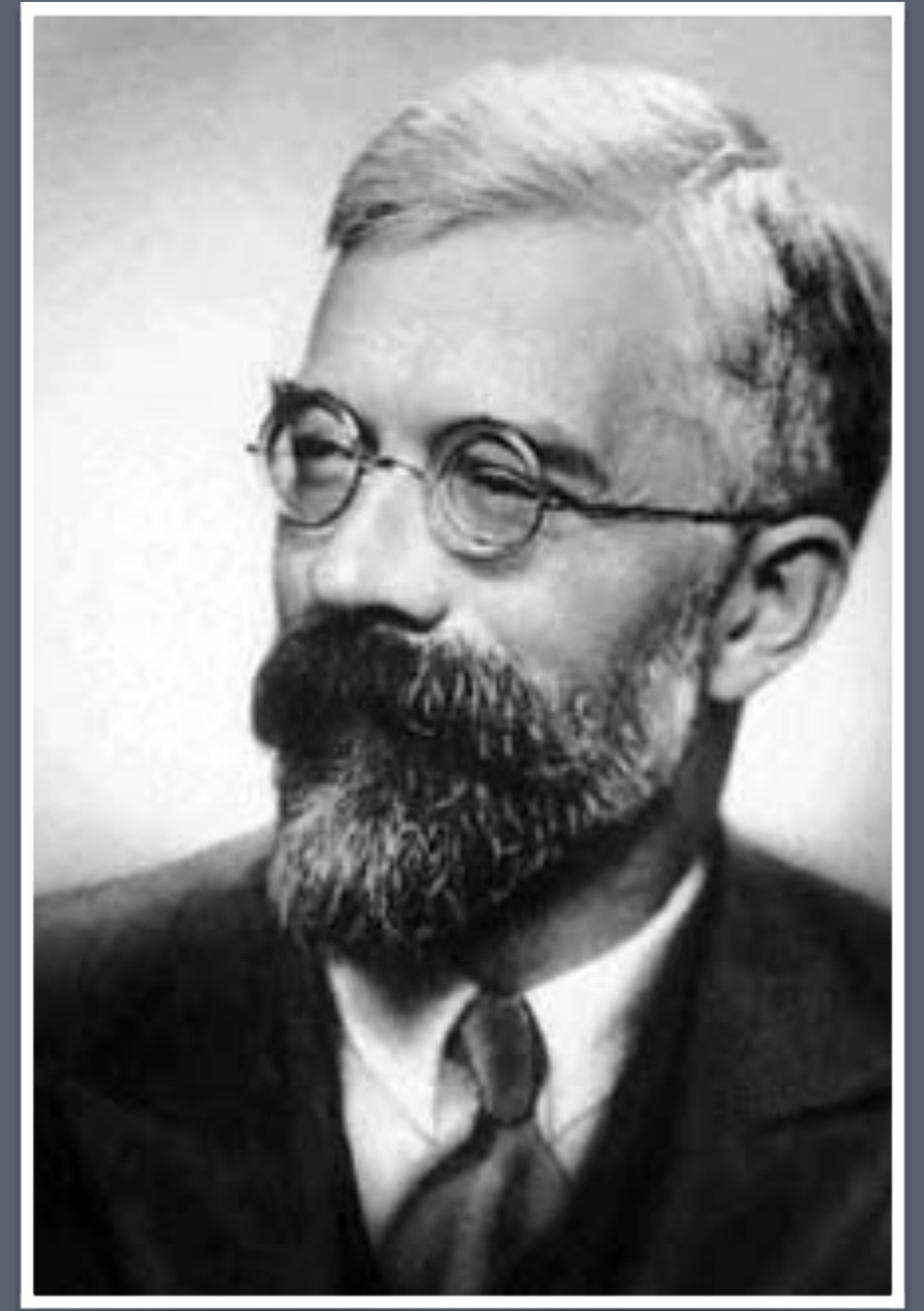
- “El valor p es como un mosquito, incómodo y difícil de ahuyentar”
- “Una herramienta estéril que no deja progenie”
- “Debería llamarse: “Statistical Hypothesis Inference Testing, o bien usar el acrónimo”

el valor P

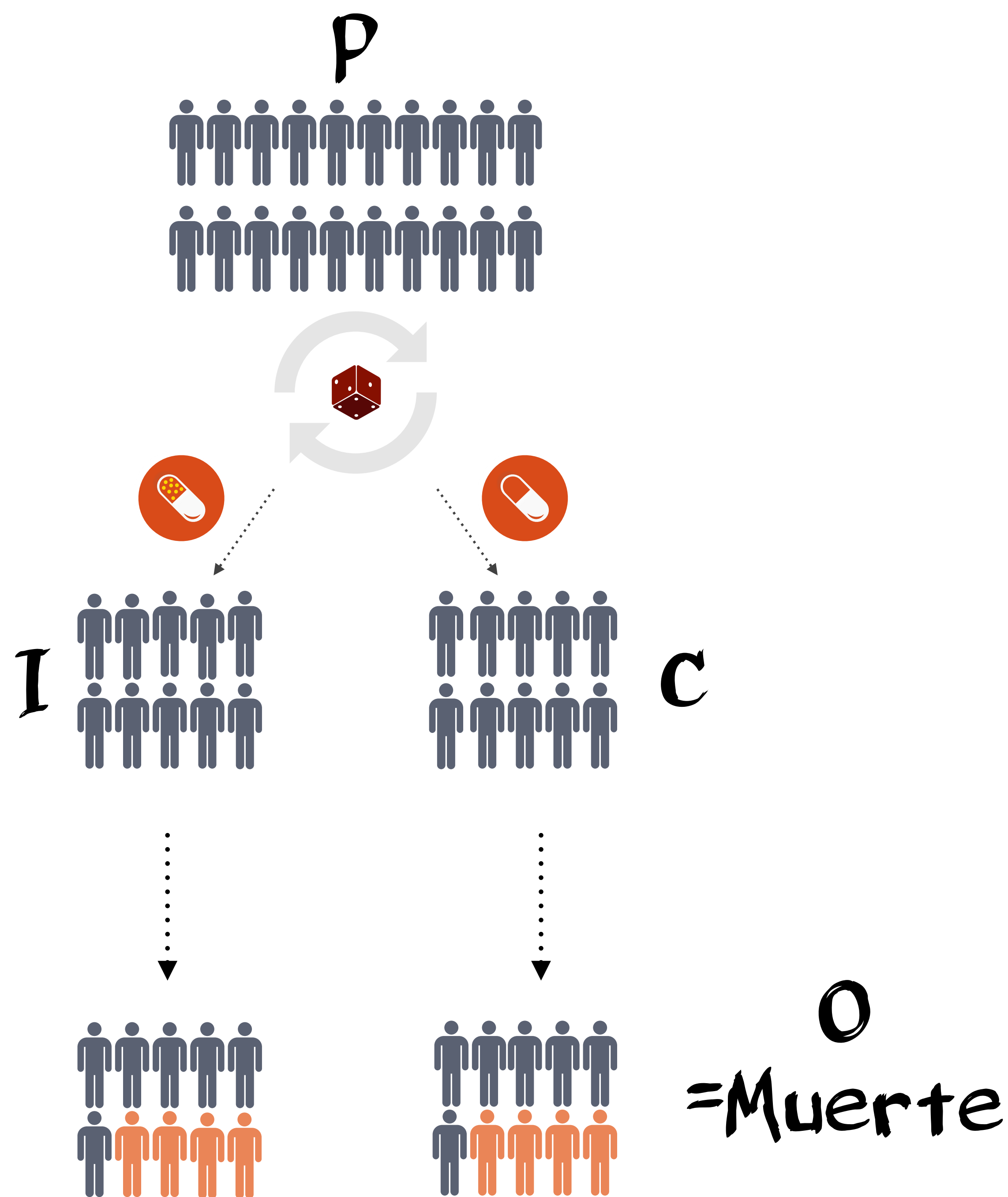
- “El valor p es como un mosquito, incómodo y difícil de ahuyentar”
- “Una herramienta estéril que no deja progenie”
- “Debería llamarse: “**S**tatistical **H**ypothesis **I**nference **T**esting, o bien usar el acrónimo”

Ronald Aylmer Fisher

- Lo creó como una medida informal para ver si valía la pena darles una segunda mirada a los resultados y repetir el estudio

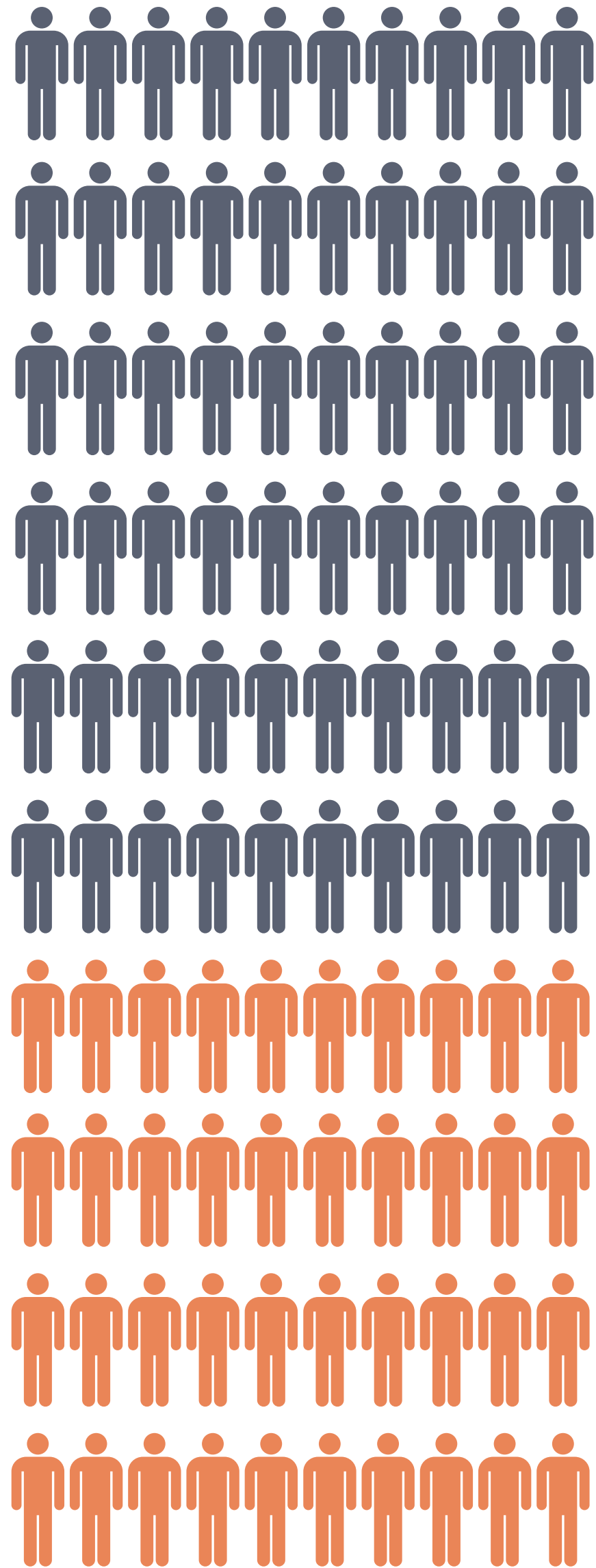


HIPÓTESIS NULA

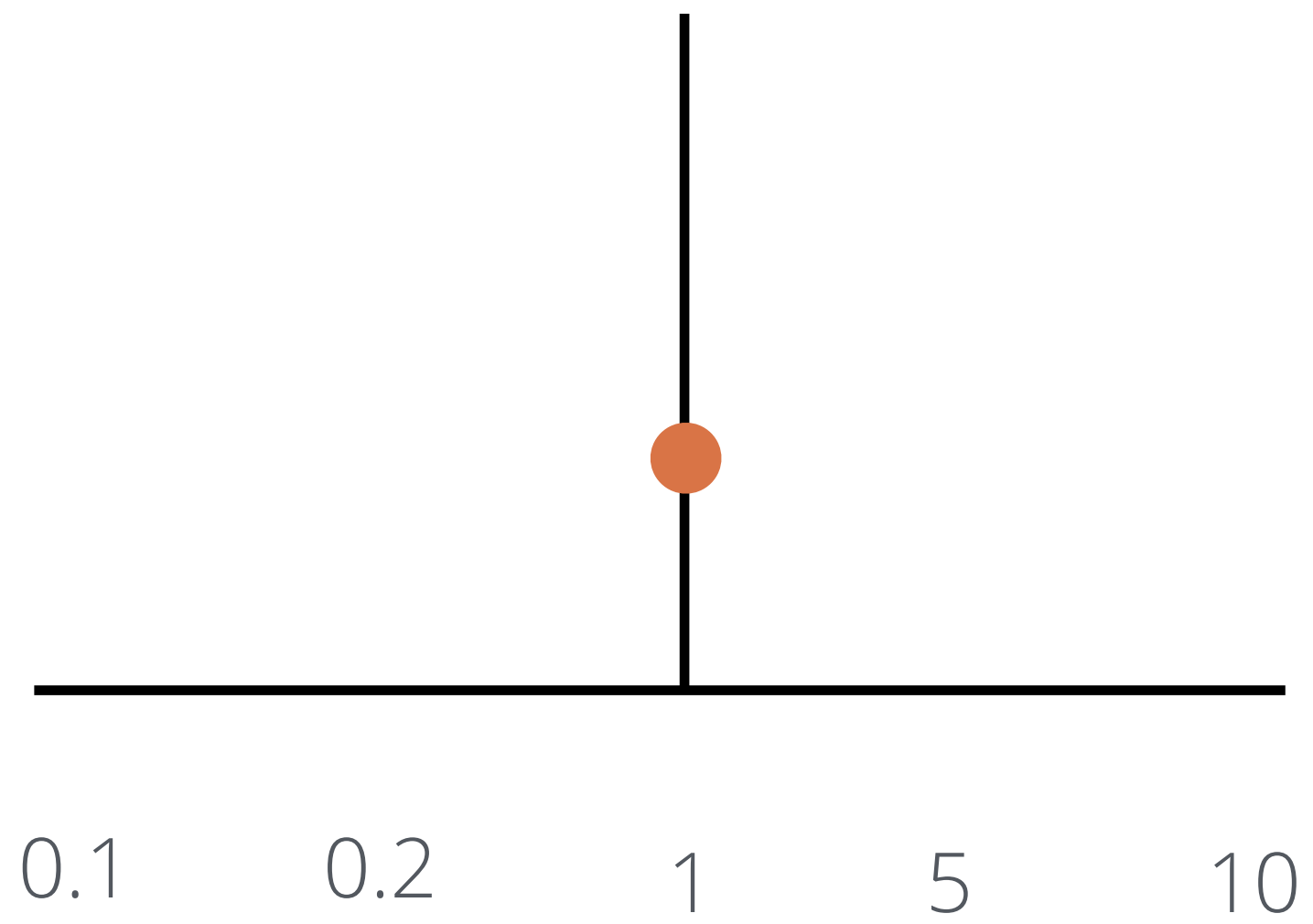
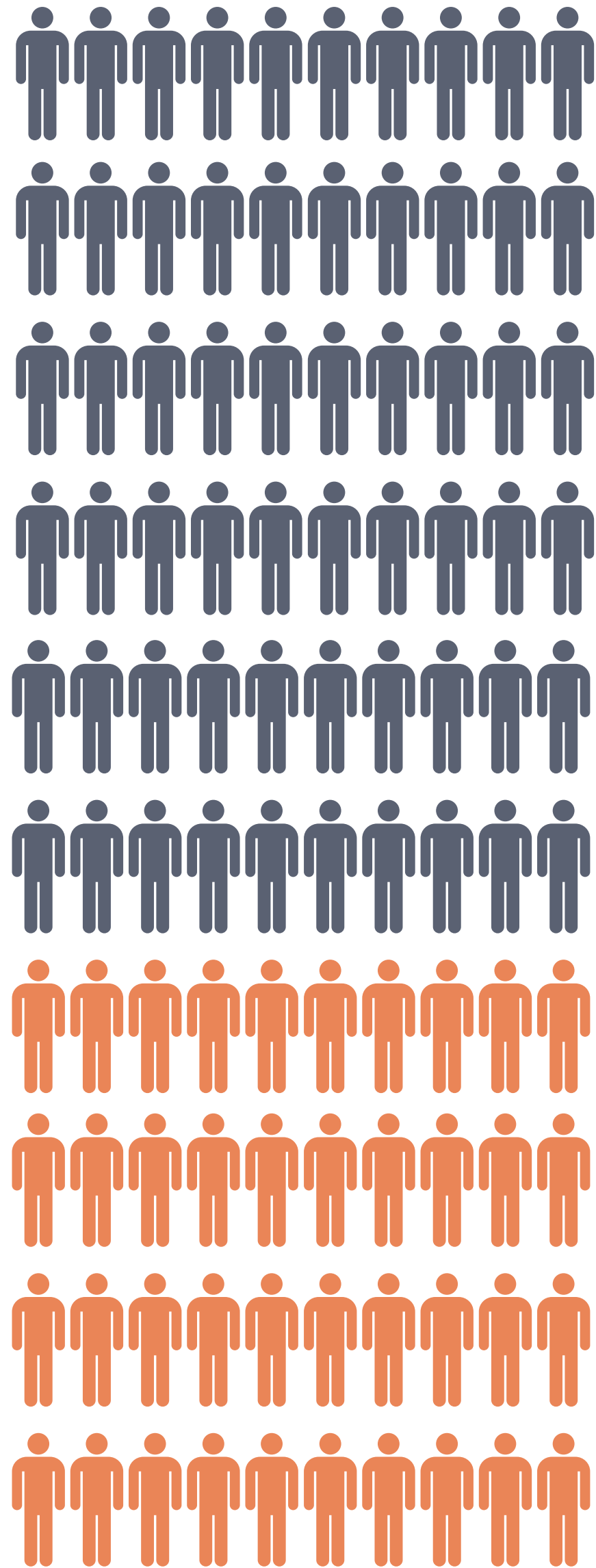


HIPÓTESIS NULA

med



placebo



RIESGO RELATIVO

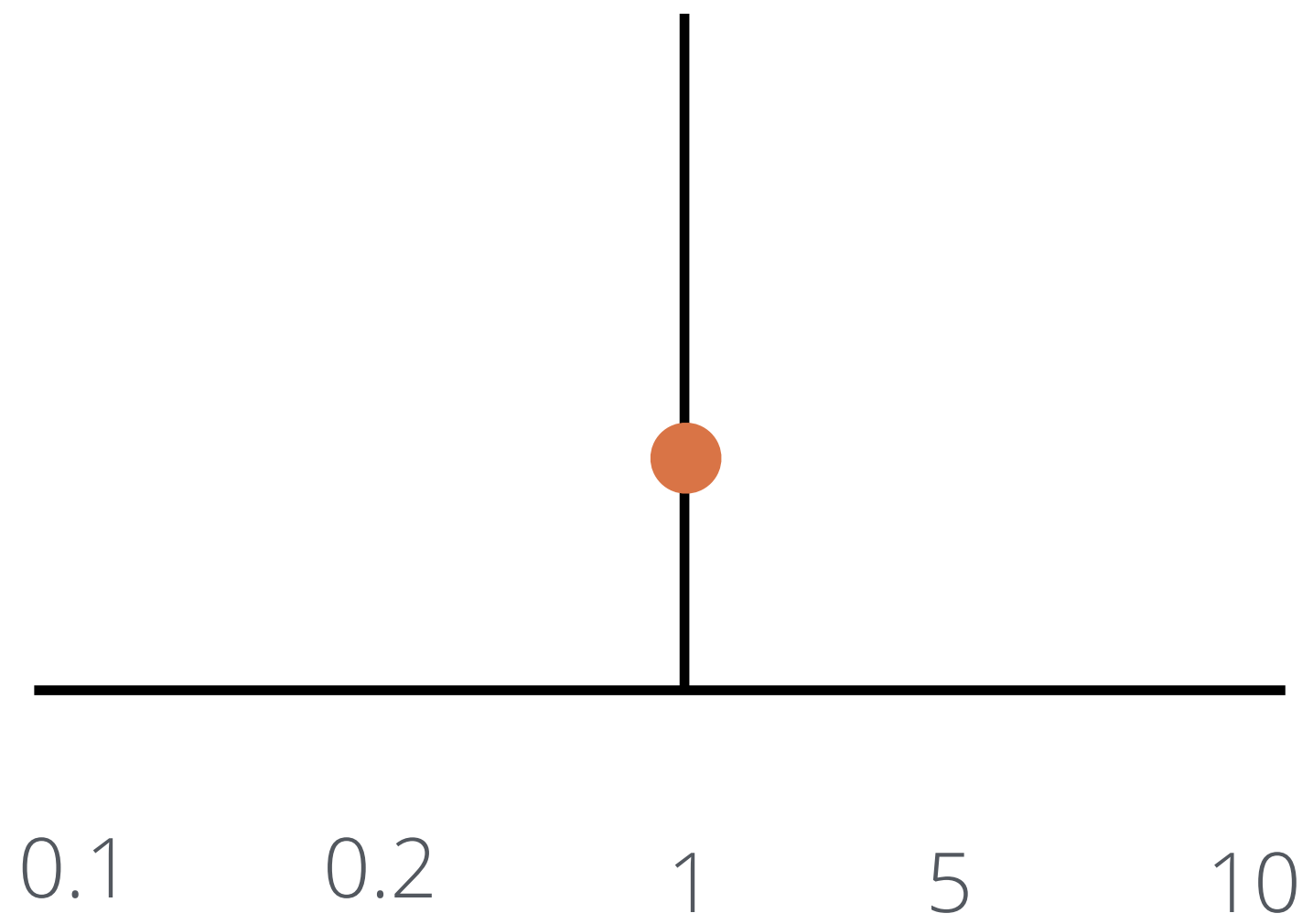
med placebo

$$\frac{40/100}{40/100} = 1$$

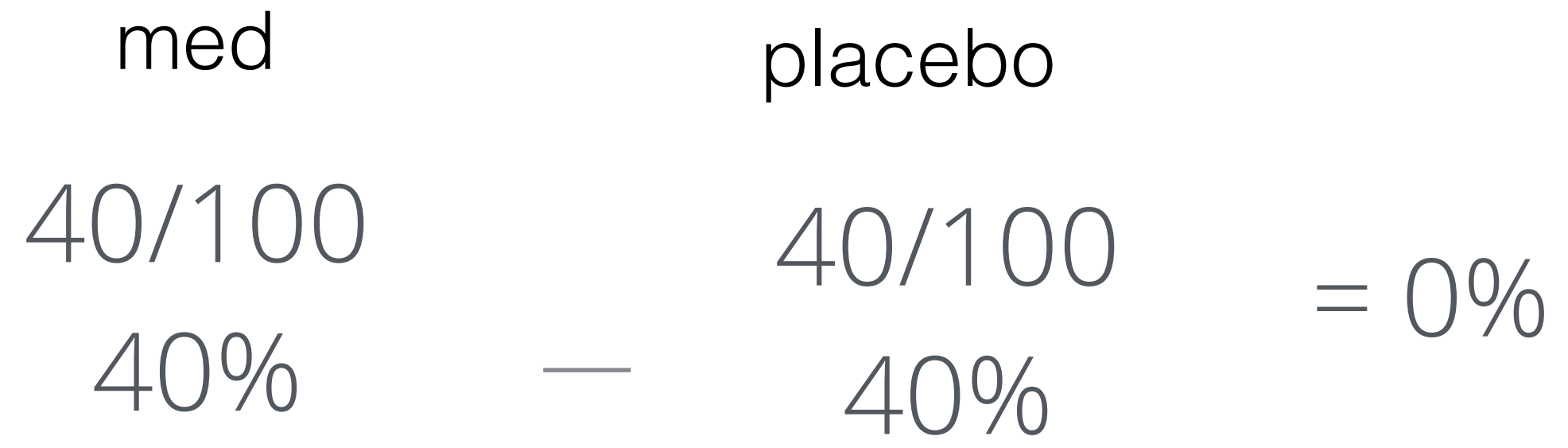
40% 40%

Riesgo de muerte

HIPÓTESIS NULA

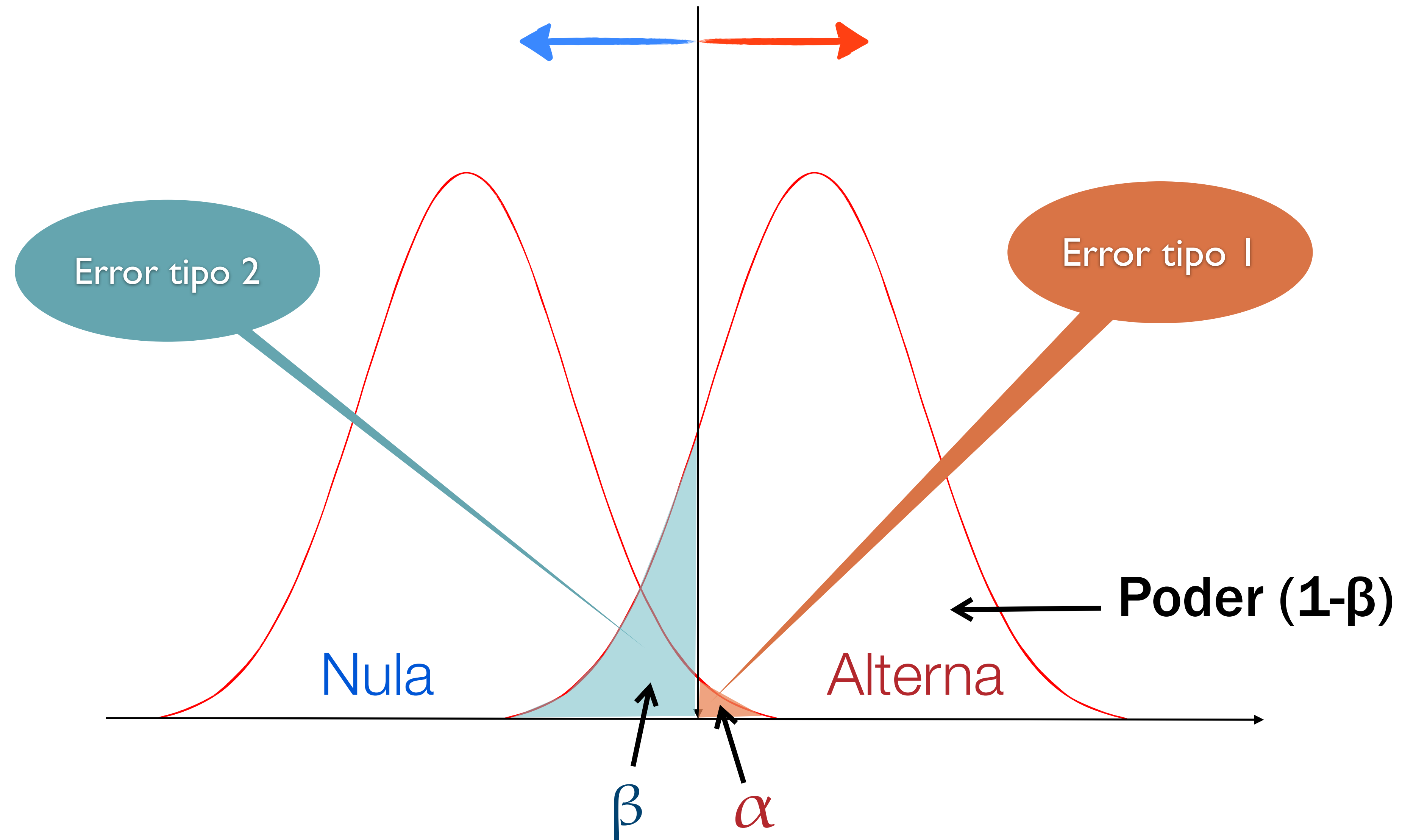


DIFERENCIA DE RIESGOS



Riesgo de muerte

Aceptar nula Rechazar nula



P valor ó valor P

1. Asume que la hipótesis nula es cierta
2. Nos dá la probabilidad de haber obtenido este resultado o uno más extremo

P valor ó valor P

- En el ejemplo de Motyl ¿un p valor de 0.01 significa que hay un 1% de probabilidad de que el resultado sea una falsa alarma?

volados

pareja	nuevo tx	placebo	probabilidad bajo la H0
1			0.5
2			0.5
3			0.5
4			0.5
5			0.5

$p=0.03$

~~“solo hay 3% de probabilidad de que ambos sean iguales, por lo tanto un 97% de que el nuevo tratamiento es mejor”~~

volados

pareja	nuevo tx	placebo	probabilidad bajo la H0
1			0.5
2			0.5
3			0.5
4			0.5
5			0.5

$p=0.03$

significa que la probabilidad de este resultado (todos ganar) es del 3%, **BAJO LA H0** de 0.5

PROBABLE CAUSE

A *P* value measures whether an observed result can be attributed to chance. But it cannot answer a researcher’s real question: what are the odds that a hypothesis is correct? Those odds depend on how strong the result was and, most importantly, on how plausible the hypothesis is in the first place.

- Chance of real effect
- Chance of no real effect

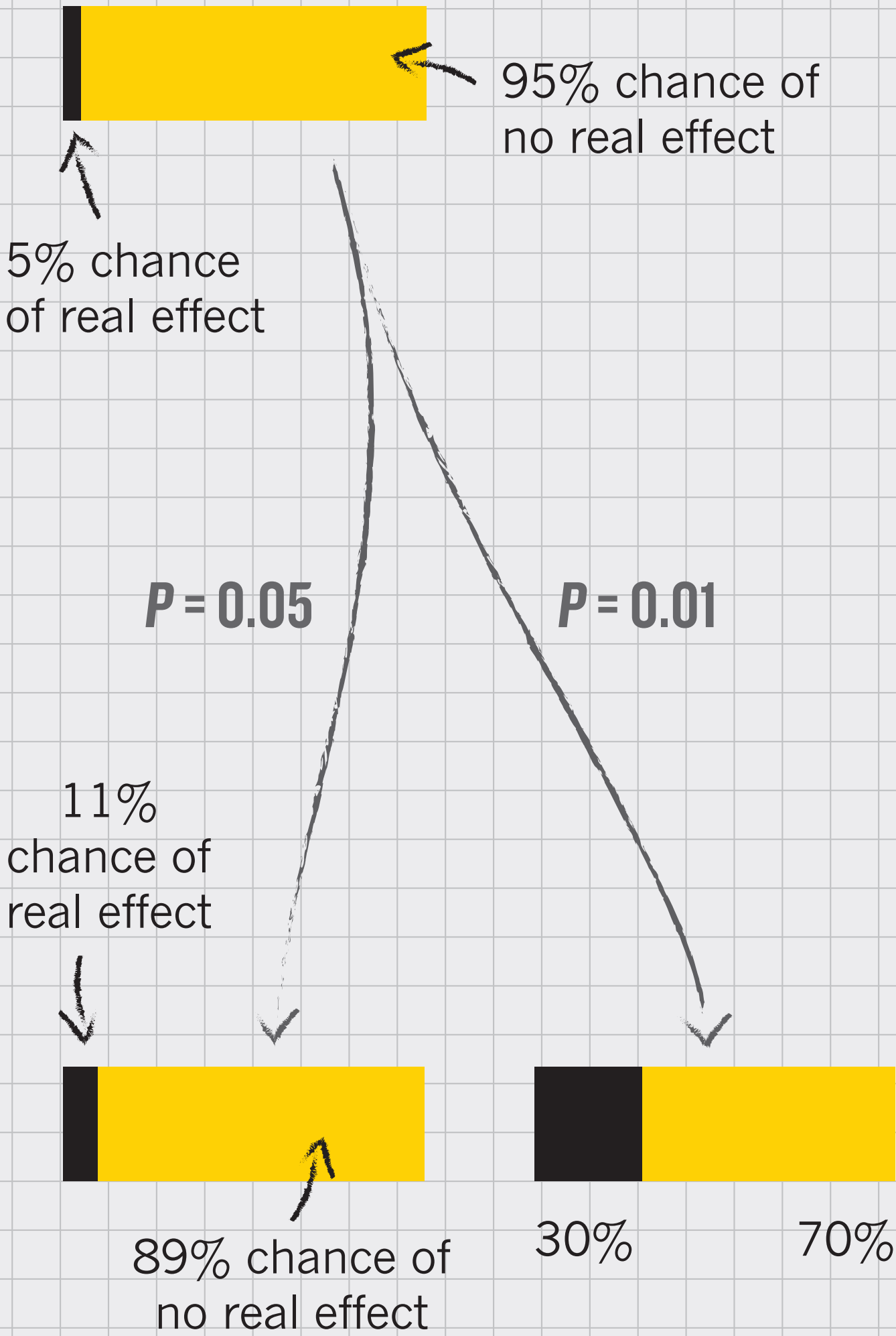
Before the experiment
The plausibility of the hypothesis — the odds of it being true — can be estimated from previous experiments, conjectured mechanisms and other expert knowledge. Three examples are shown here.

The measured *P* value
A value of 0.05 is conventionally deemed ‘statistically significant’; a value of 0.01 is considered ‘very significant’.

After the experiment
A small *P* value can make a hypothesis more plausible, but the difference may not be dramatic.

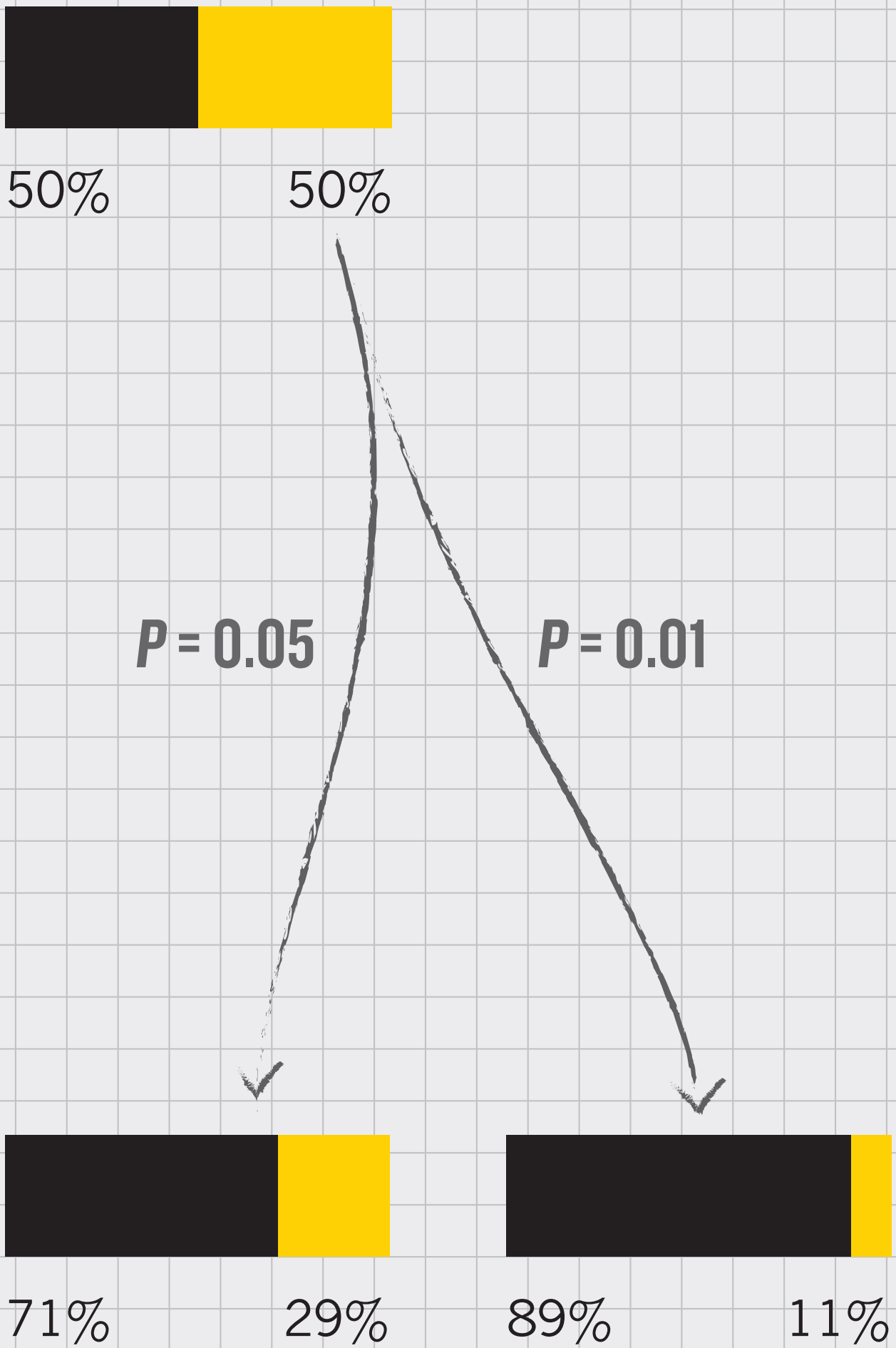
THE LONG SHOT

19-to-1 odds against



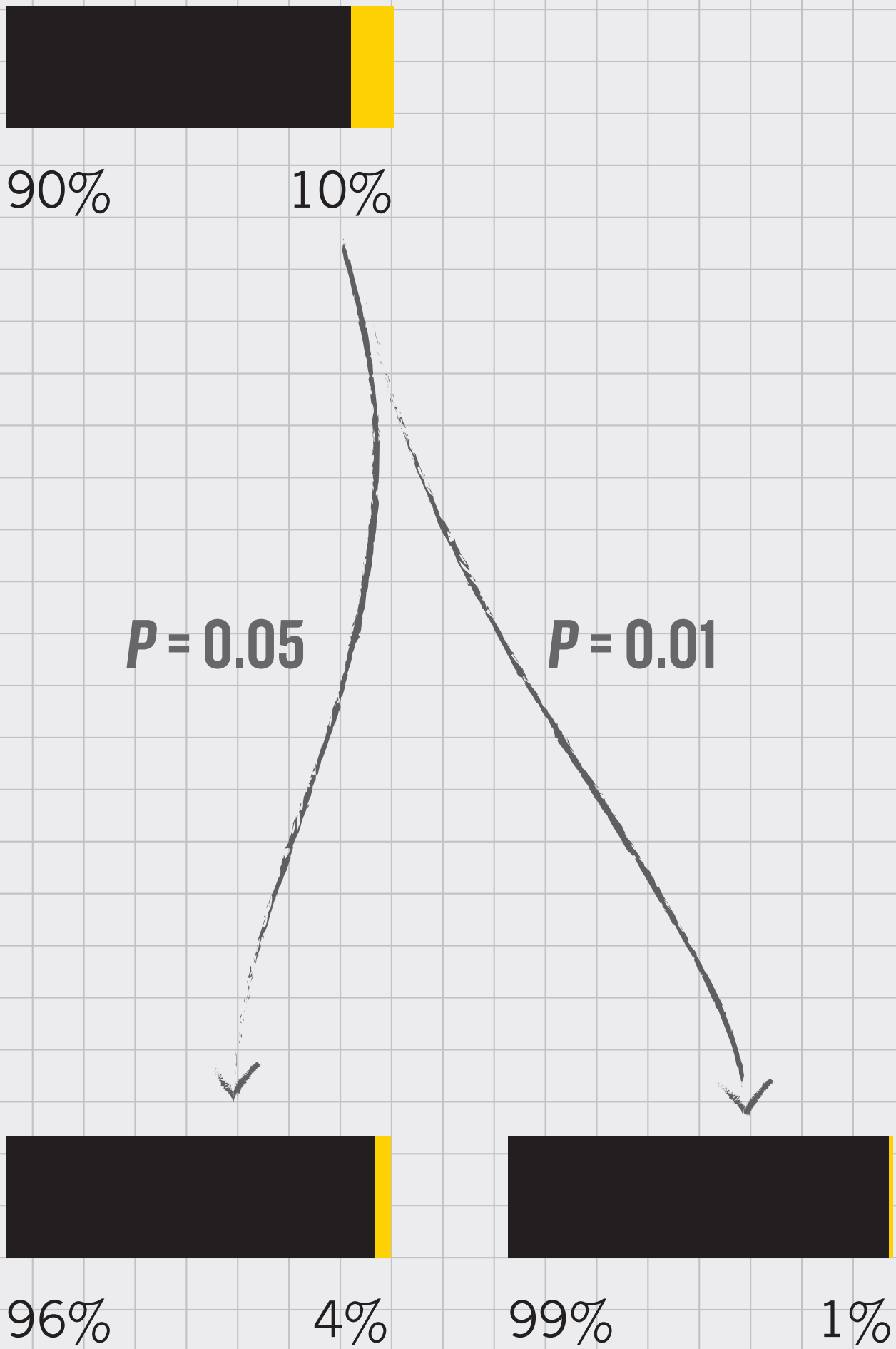
THE TOSS-UP

1-to-1 odds



THE GOOD BET

9-to-1 odds in favour



P HACKING

JELLY BEANS
CAUSE ACNE!

SCIENTISTS!
INVESTIGATE!

BUT WE'RE
PLAYING
MINECRAFT!
... FINE.



WE FOUND NO
LINK BETWEEN
JELLY BEANS AND
ACNE ($p > 0.05$).



THAT SETTLES THAT.

I HEAR IT'S ONLY
A CERTAIN COLOR
THAT CAUSES IT.

SCIENTISTS!

BUT
MIIINECRAFT!



20 veces después...



News

GREEN JELLY
BEANS LINKED
TO ACNE!

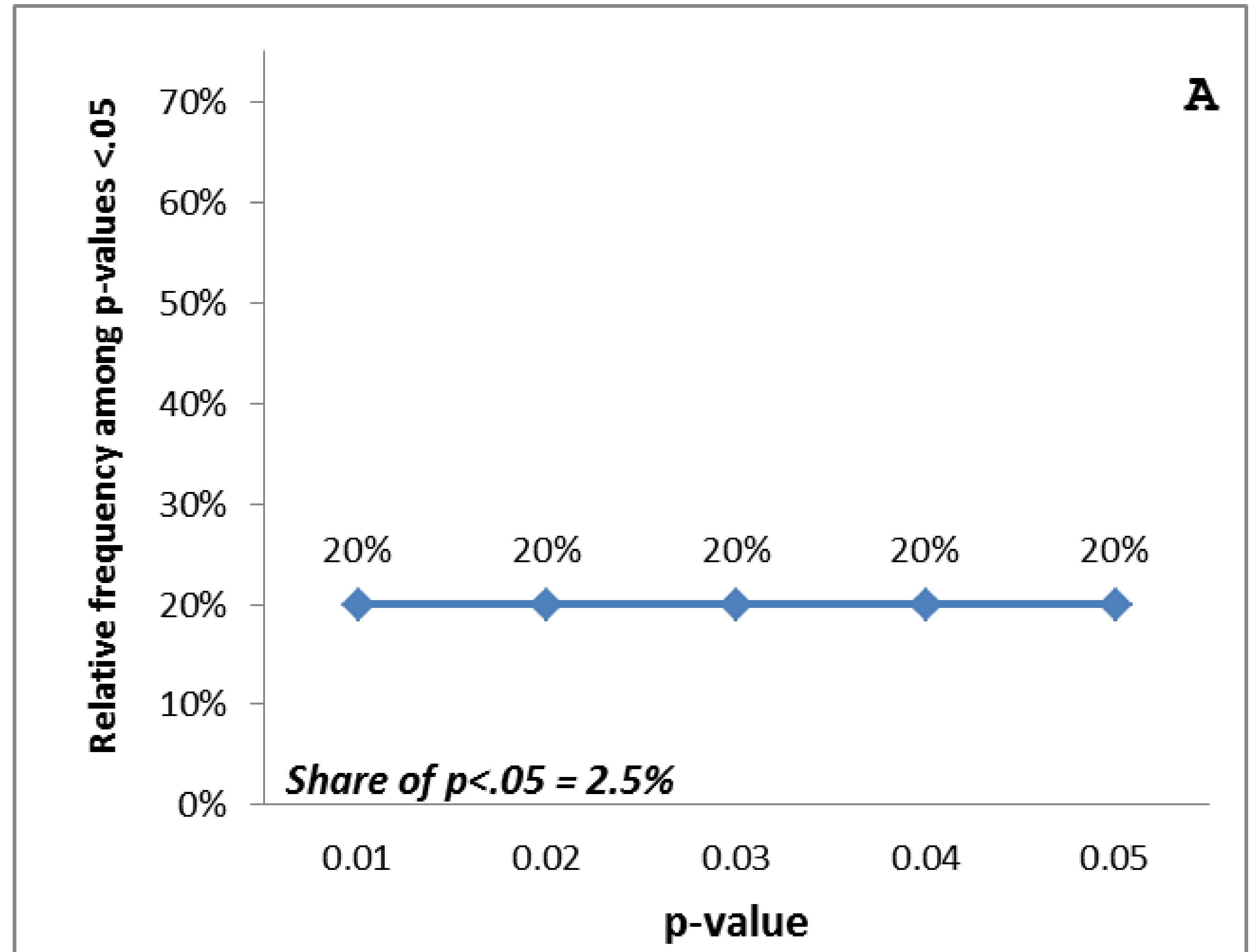
95% CONFIDENCE

ONLY 5% CHANCE
OF COINCIDENCE!

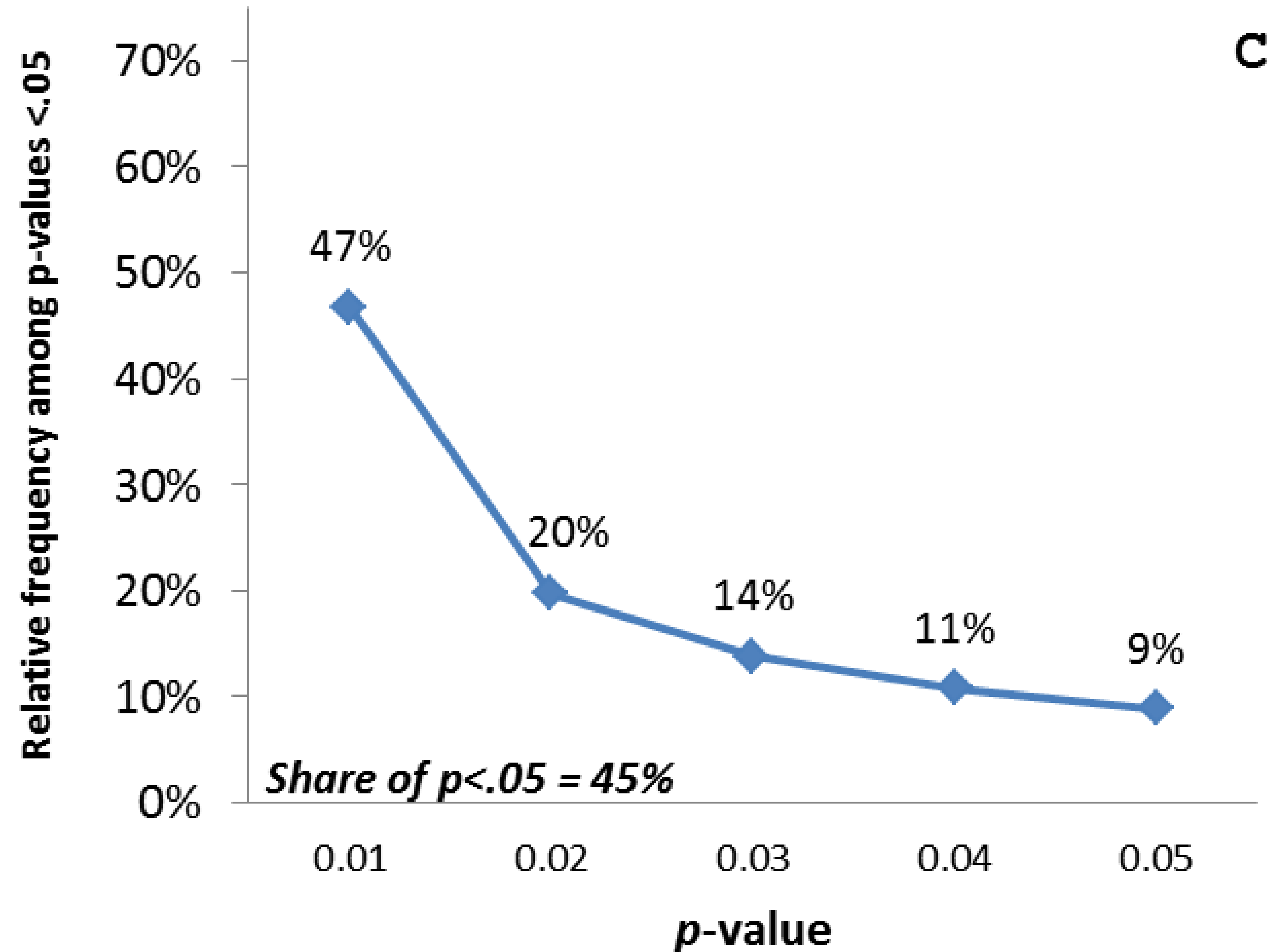


SCIENTISTS...

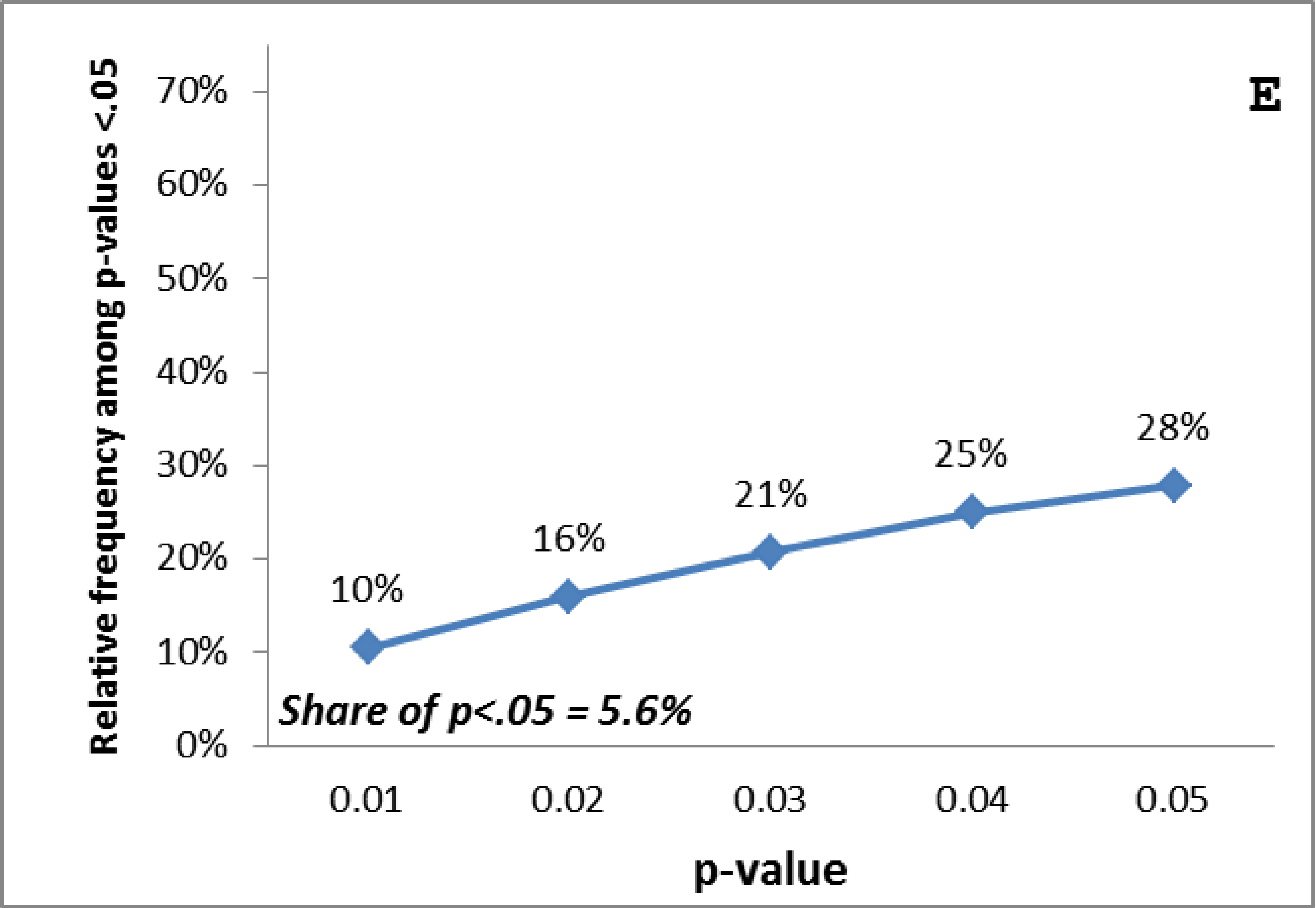
Distribución de valores p en estudios de una intervención sin efecto



Distribución de
valores p en
estudios de una
intervención
CON efecto



Distribución de valores p en estudios de una intervención sin efecto
con **p-hacking**

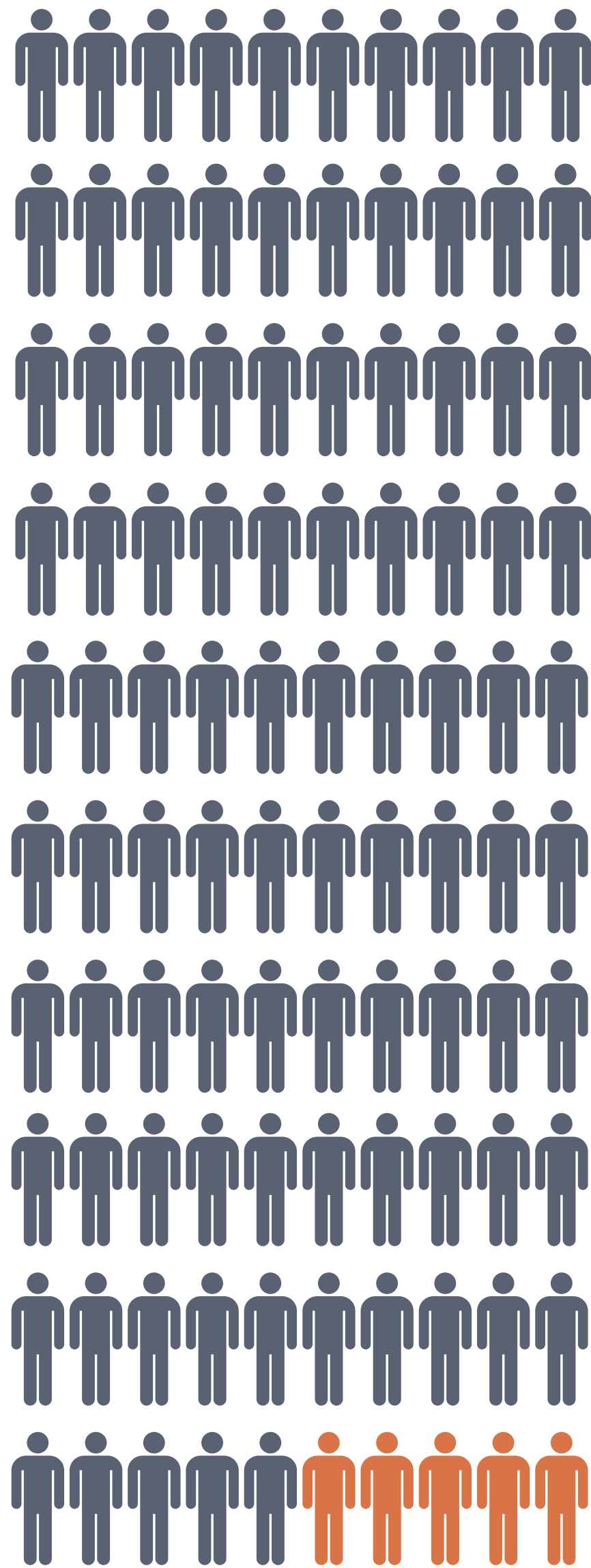


**A VECES EL VALOR P
DESVÍA LA ATENCIÓN DE LO
IMPORTANTE, EL TAMAÑO
DEL EFECTO**

Un estudio de más de 19 mil

- Aquellos que conocen a sus parejas online tienen menos probabilidad de terminar divorciados
 - $p < 0.002$
 - Efectos:
 - ▶ Divorcio en personas off-line= 7.67%
 - ▶ Divorcio en personas on-line= 5.96%
- Diferencia de 1.5%
- 
- A hand-drawn orange bracket connects the two items in the 'Efectos' list, pointing towards the text 'Diferencia de 1.5%'. The bracket is drawn with a simple, sketchy line.

10,000
on-line



10,000
off-line



favorece online

favorece offline

100 50 10 0 10 50 100

DIFERENCIA DE RIESGOS

off-line

on-line

$$\frac{500}{10,000} \quad \text{—} \quad \frac{700}{10,000} = 2\%$$

5% — 7%

Riesgo de divorcio

“Significancia”

- Trata de no usar las palabras “estadísticamente significativo”, mejor puedes decir:
 - ▶ “no hay evidencia que soporte (o rechace) la hipótesis nula”
 - ▶ “un efecto no puede ser confirmado o refutado por la evidencia disponible”
- siempre presentar los IC95%, independientemente de si usas el valor p o no
- considera cuál nivel de cambio sería clínicamente importante
- absence of evidence $\neq$ evidence of absence

American Statistics Association



AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®

732 North Washington Street, Alexandria, VA 22314 • (703) 684-1221 • Toll Free: (888) 231-3473 • www.amstat.org • [www.twitter.com/AmstatNews](https://twitter.com/AmstatNews)

AMERICAN STATISTICAL ASSOCIATION RELEASES STATEMENT ON STATISTICAL SIGNIFICANCE AND P-VALUES

*Provides Principles to Improve the Conduct and Interpretation of Quantitative
Science*

March 7, 2016

American Statistics Association

1. Valores p pueden indicar que tan incompatible son los datos con un modelo específico
2. No miden la probabilidad de que la hipótesis de estudio sea verdadera, o la probabilidad de que los datos sean producidos por el azar solamente
3. Conclusiones científicas ... no deben basarse solo en nivel de significancia especificado (el clásico 0.05)

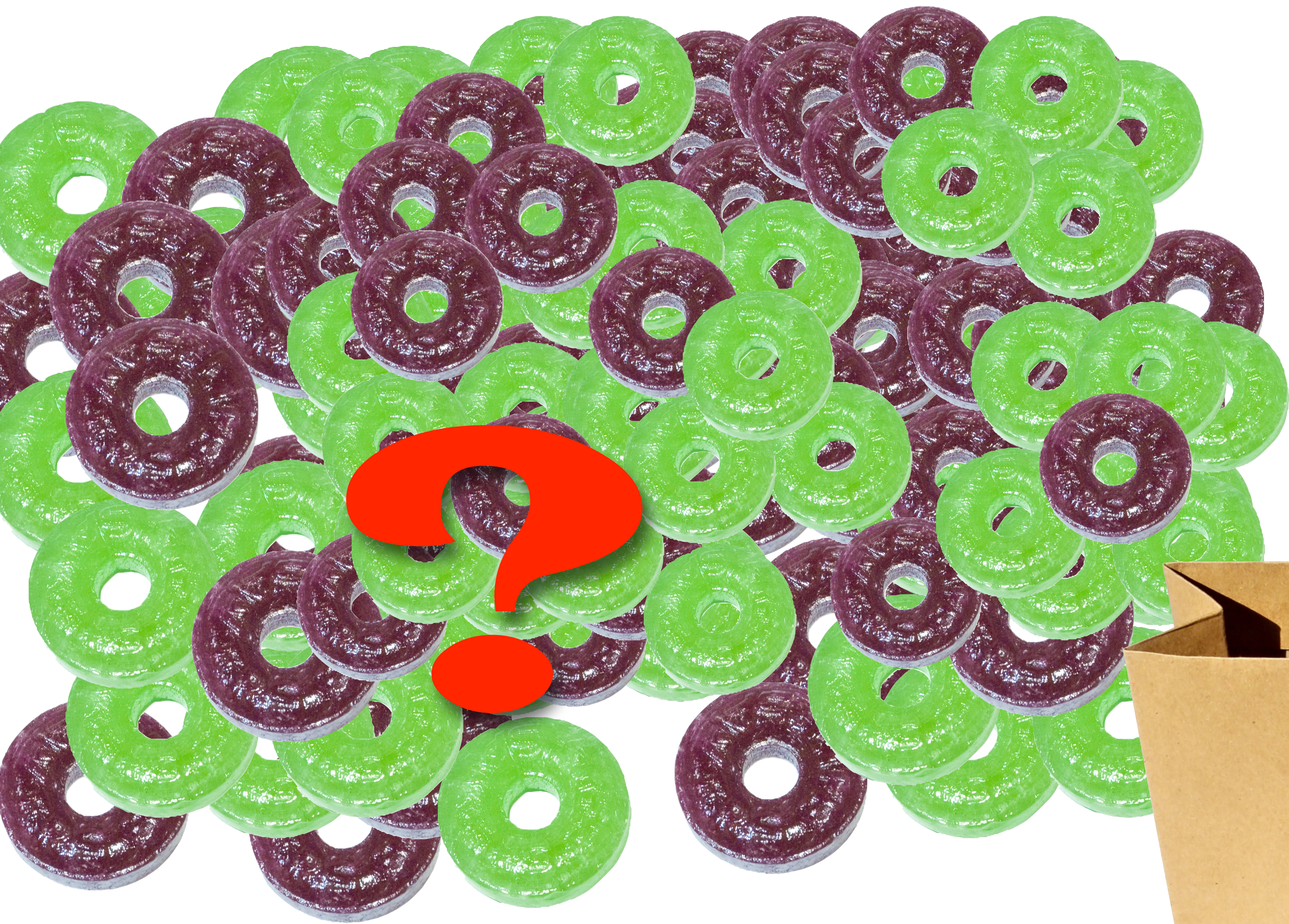
American Statistics Association

4. Inferencias apropiadas requieren un reporte completo y transparente
5. Un valor p “estadísticamente significativo” NO mide el tamaño del efecto o la importancia del resultado
6. Por sí solo, el valor p no provee una medida adecuada de la evidencia en cuanto al modelo o la hipótesis

INTERVALOS DE

CONFIANZA

PRECISIÓN



ADIVINEMOS

- Sin ver ni SACAR ninguno ¿Cuántos caramelos verdes hay en la bolsa?

Mínimo

Máximo

0

?

30



Saca 1

- ¿Cuántos caramelos verdes hay en la bolsa?

Mínimo

Máximo

1

30



Saca 2

- ¿Cuántos caramelos verdes hay en la bolsa?

Mínimo

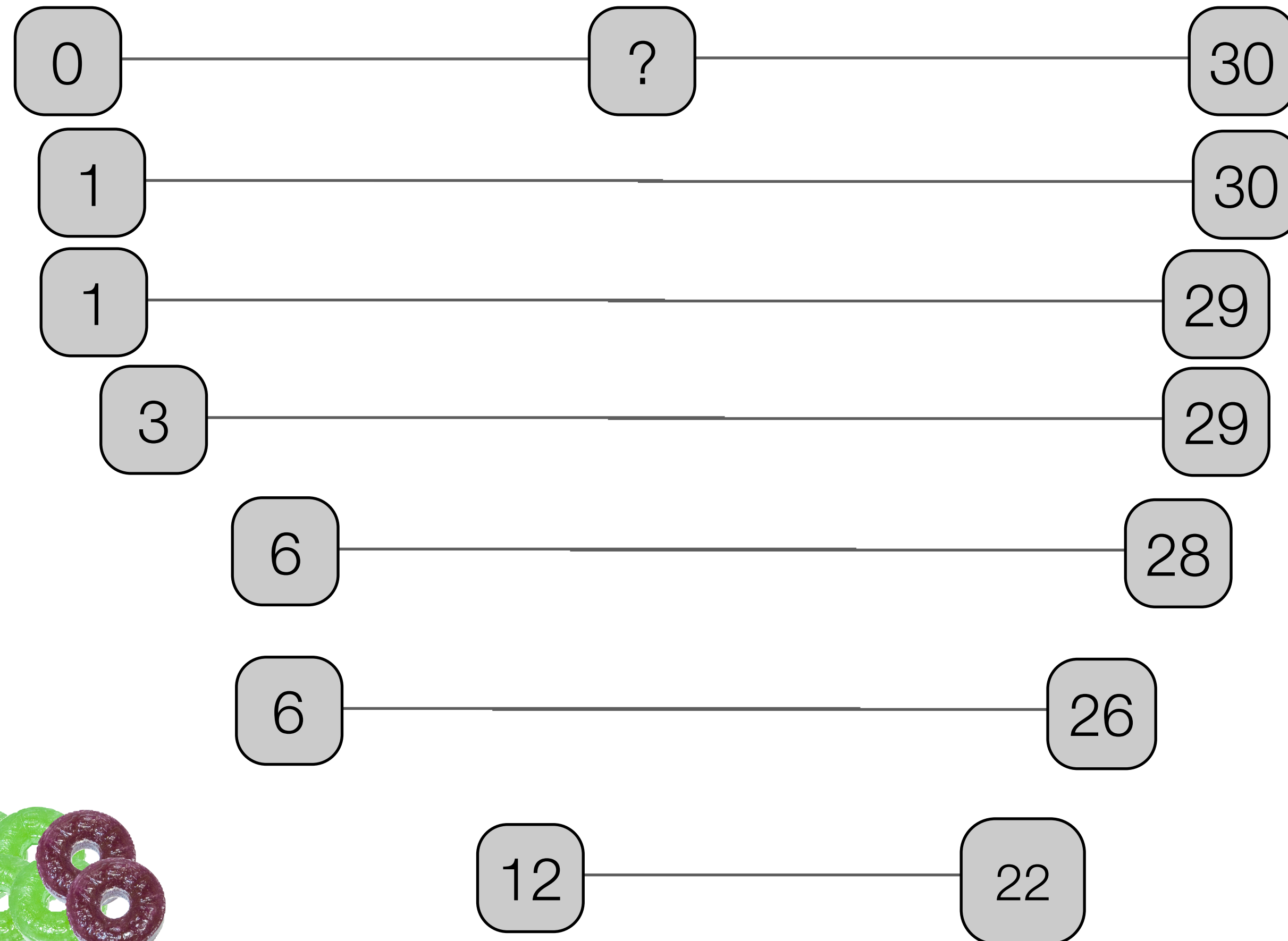
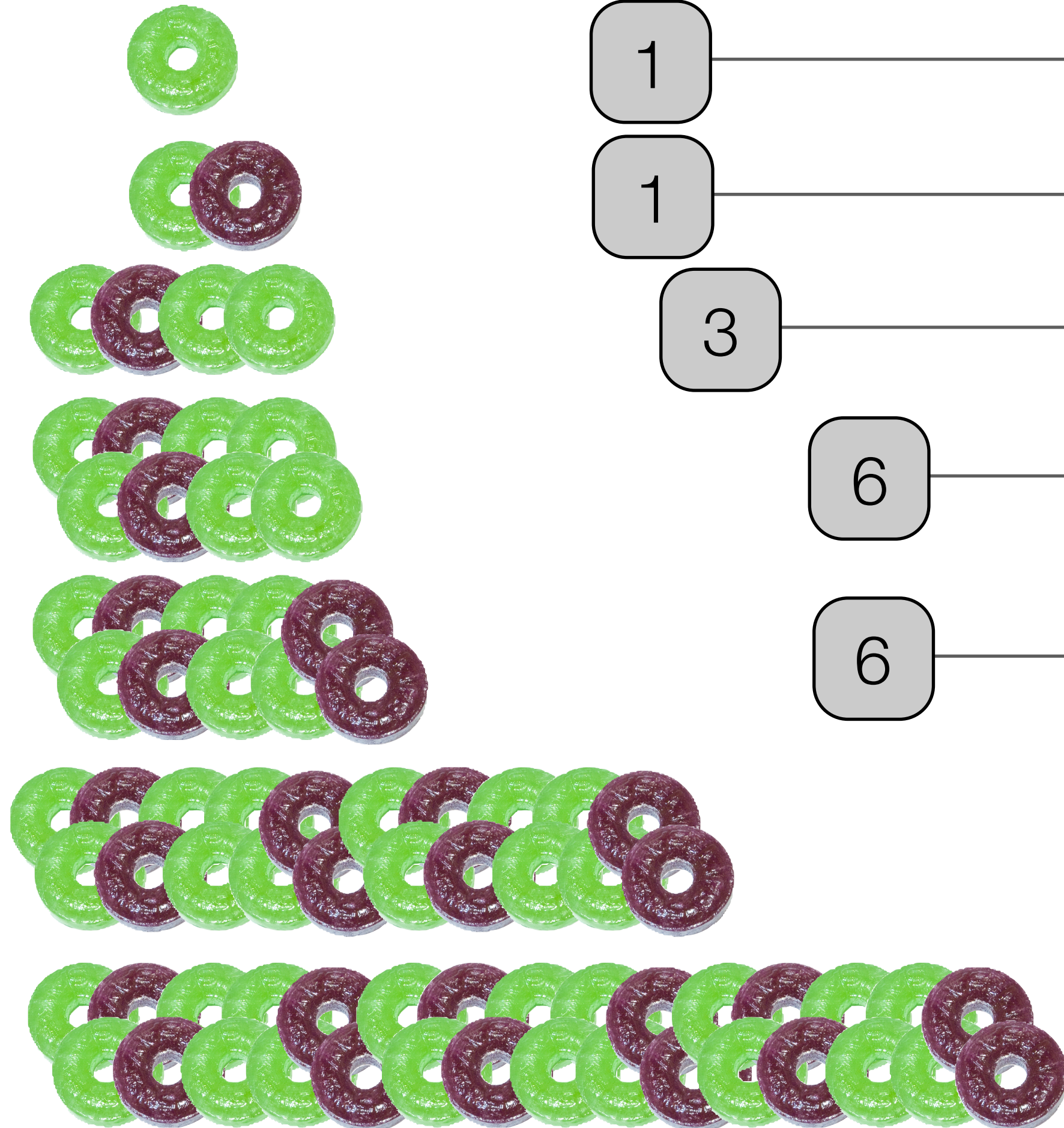
Máximo

1

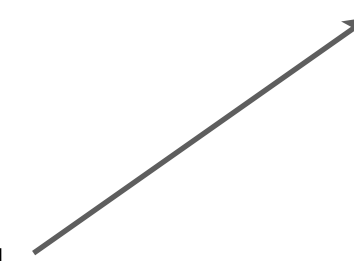
29



N = 0



La verdad



18

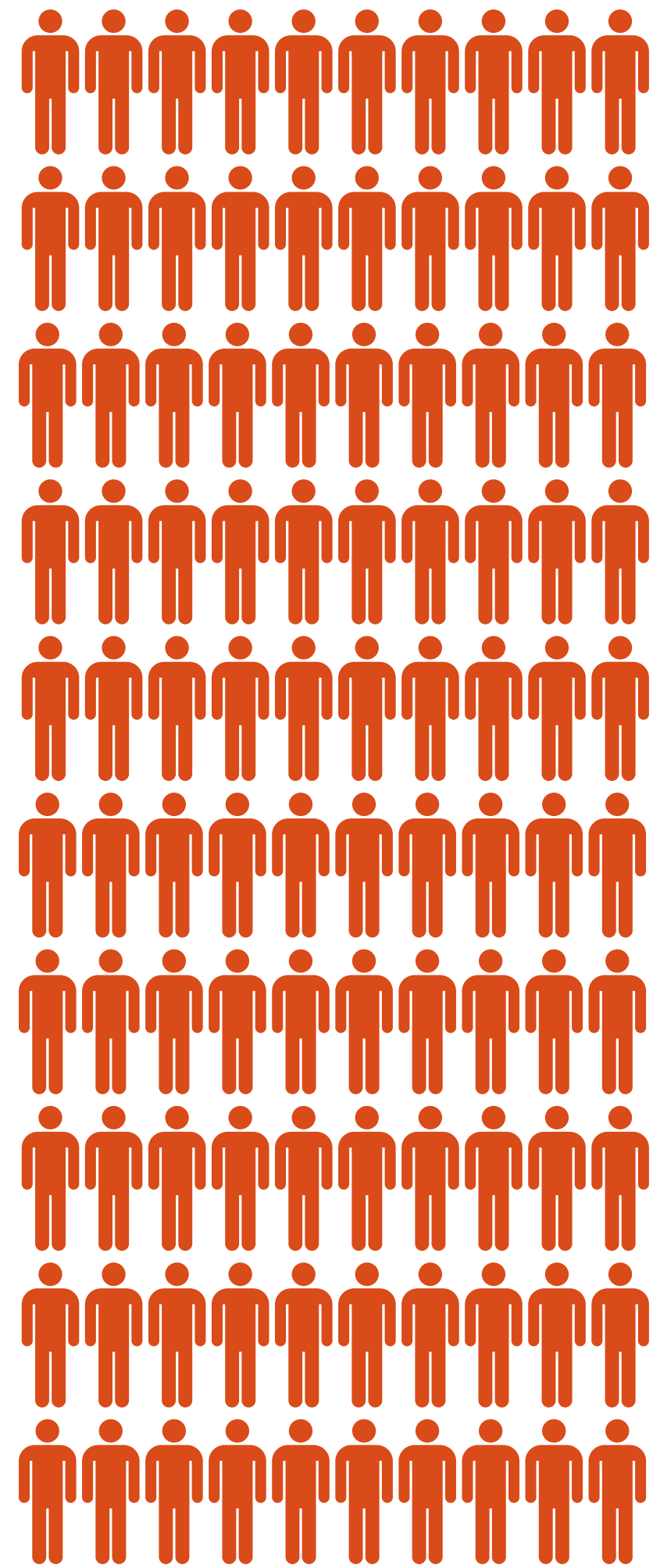


imagina

EN TODO MÉXICO

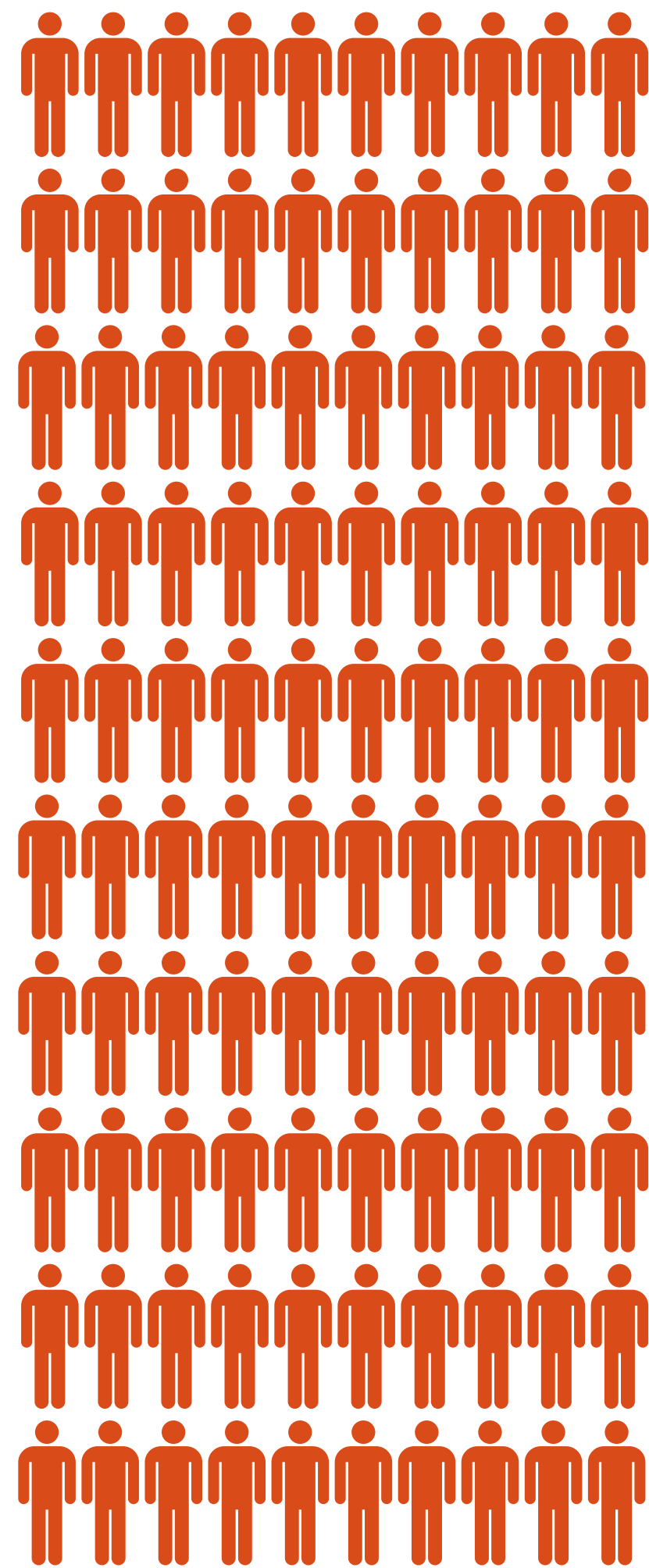
**IQ promedio es de 100 puntos
(desviación estándar = 10)**





haces un
muestreo
aleatorio de 100
personas y
obtienes...



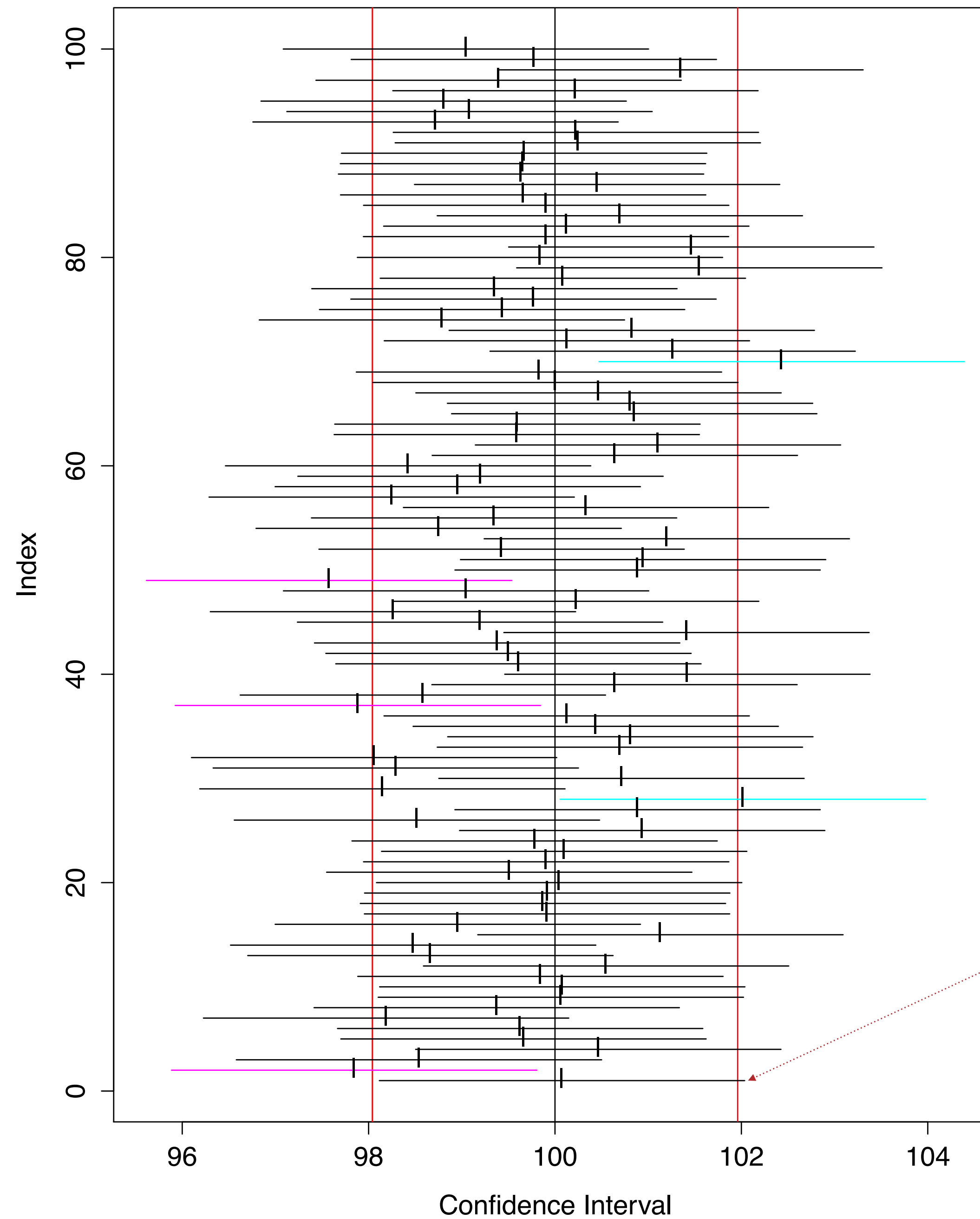


Media=100.02

I.C. 95% de 98.02 a 102.01

¿Qué significa?

Confidence intervals based on z distribution



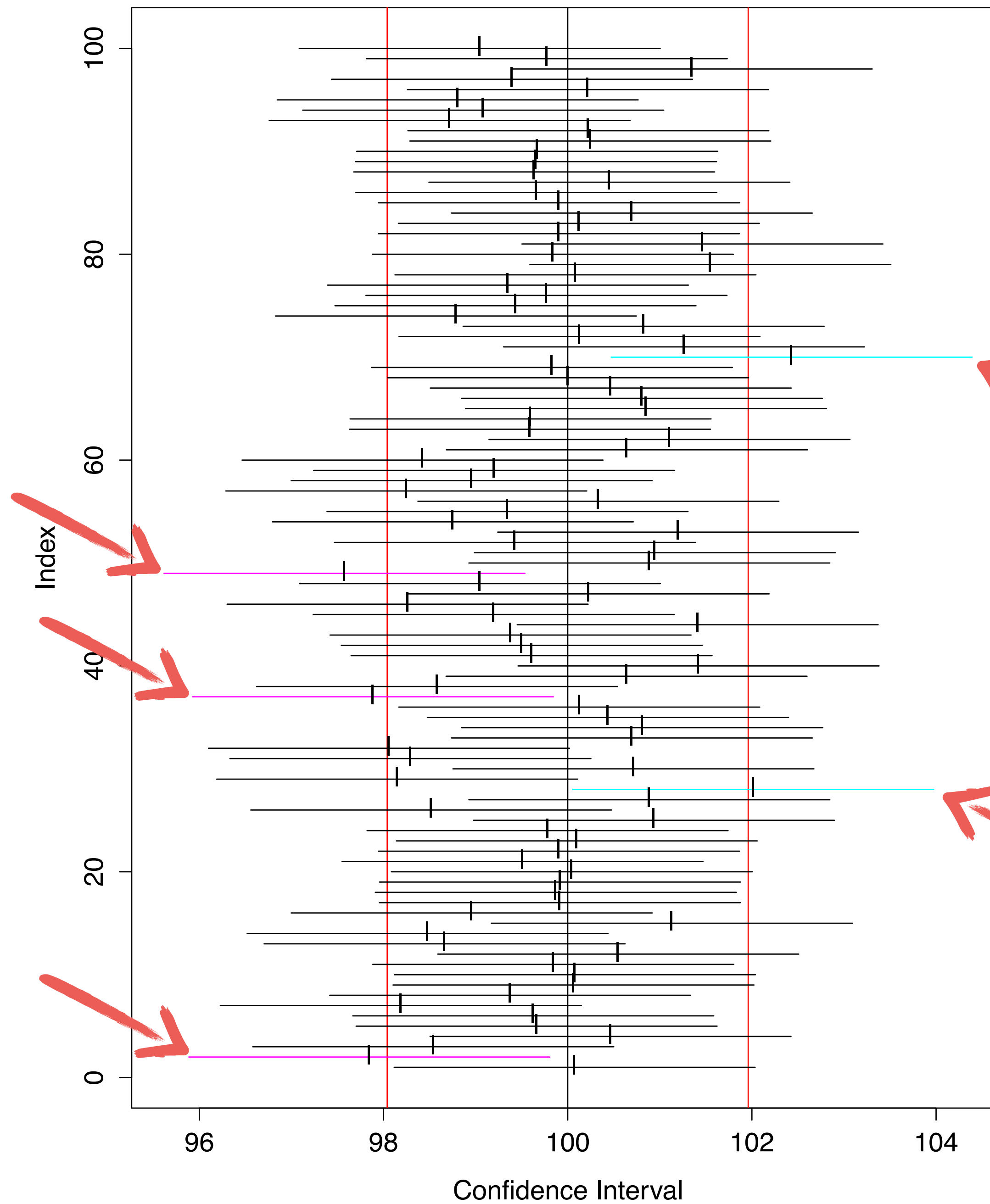
IC95% de 100
estudios
 $\mu = 100 (\pm 10)$

mi estudio

I.Q. = 100.02

I.C. 95% de 98.02 a 102.01

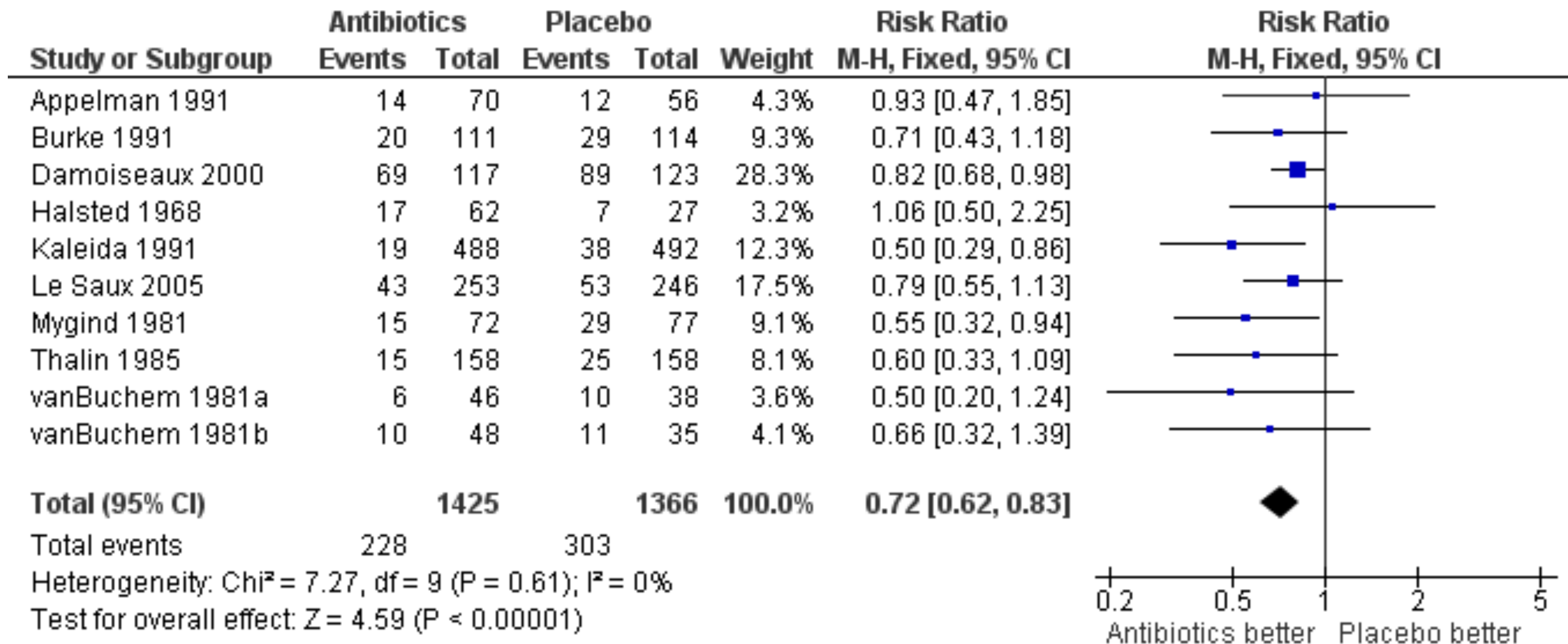
Confidence intervals based on z distribution



IC95% de 100
estudios
 $\mu = 100 (\pm 10)$

si repitiésemos
el estudio 100
veces, en 5
hallaríamos
que no se
incluye la
media
poblacional (el
verdadero
valor)

¿Son estos resultados clínicamente importantes (significativos)?



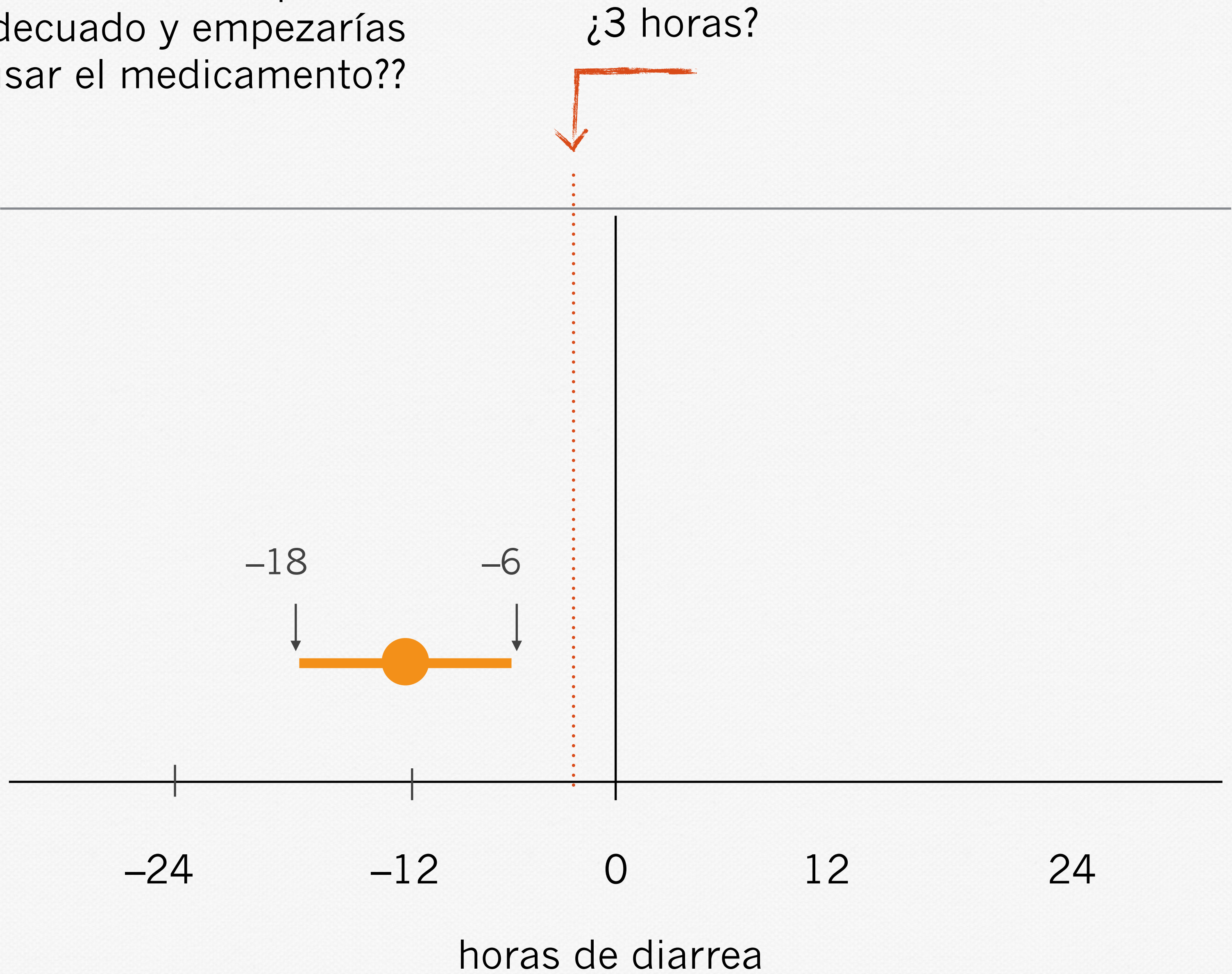
Antibióticos para otitis media aguda en niños
desenlace de mejoría clínica (ausencia de dolor de oído) a los 2-7d

ejemplo con diarrea



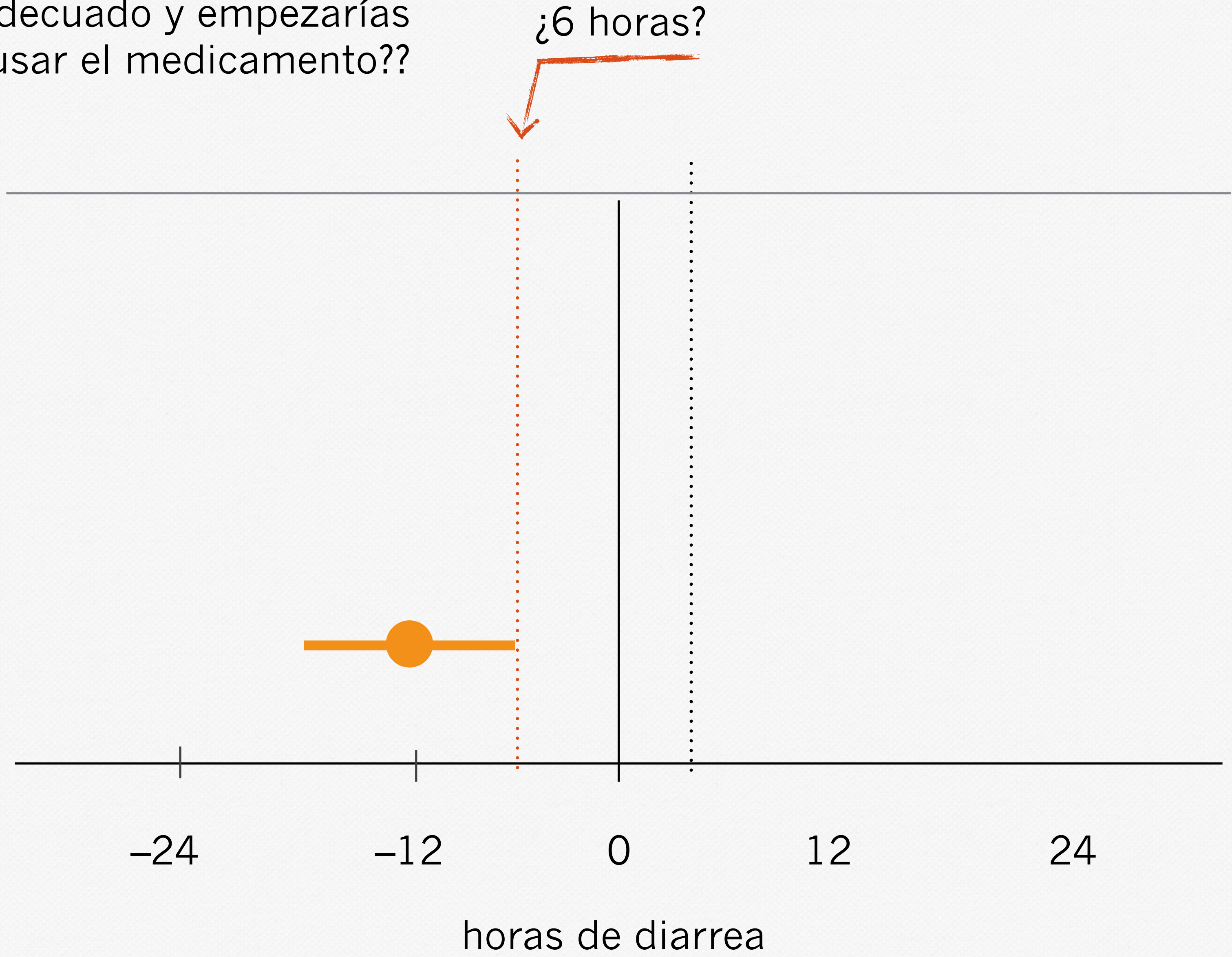
imagina medicamento que
provee disminución de 12
horas (IC95% de -18 a -6)

¿Cuántas horas menos
con diarrea crees que sea
lo adecuado y empezarías
a usar el medicamento??



ejemplo con diarrea

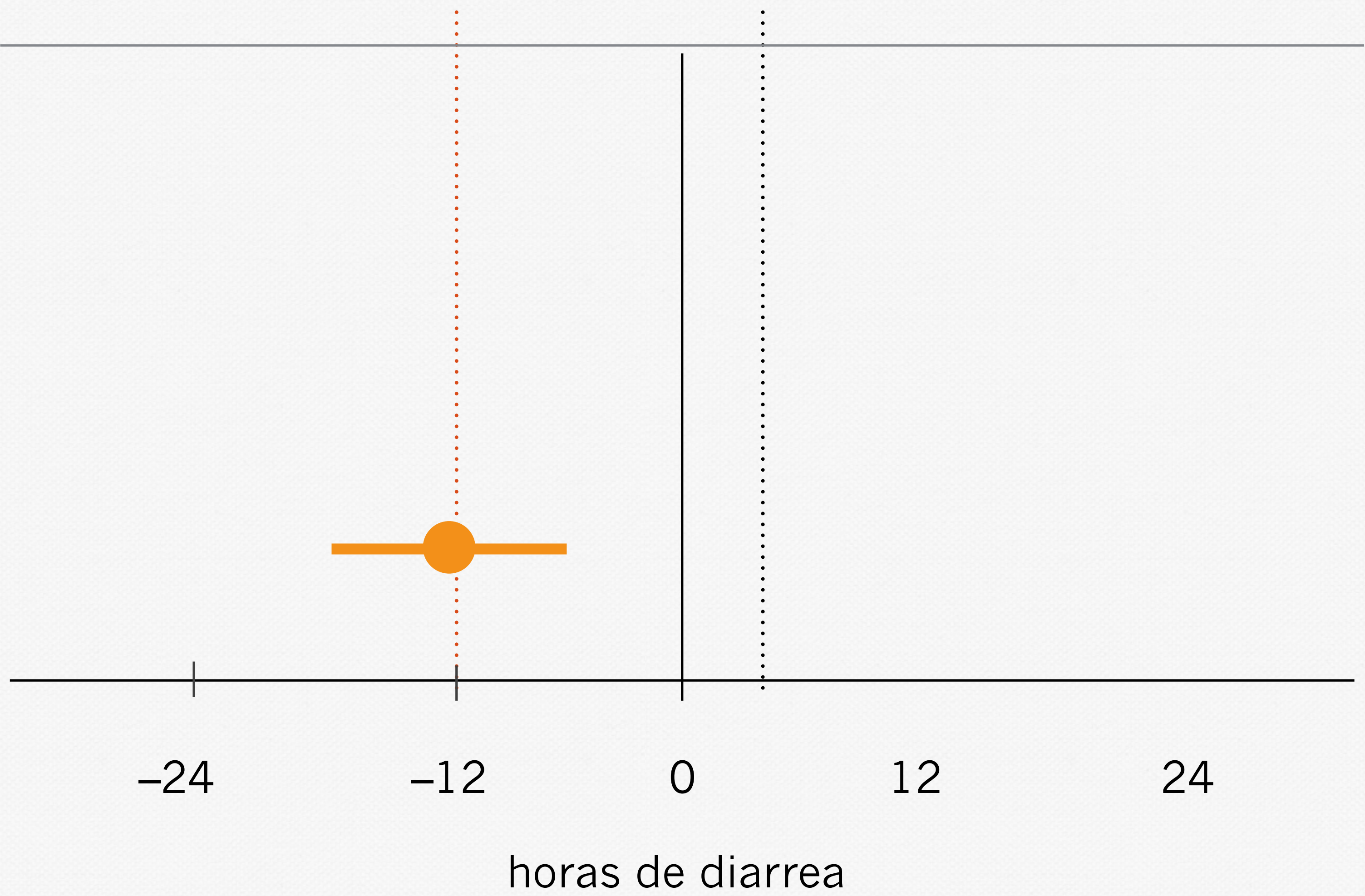
¿Cuántas horas menos
con diarrea crees que sea
lo adecuado y empezarías
a usar el medicamento??



ejemplo con diarrea

¿Cuántas horas menos
con diarrea crees que sea
lo adecuado y empezarías
a usar el medicamento??

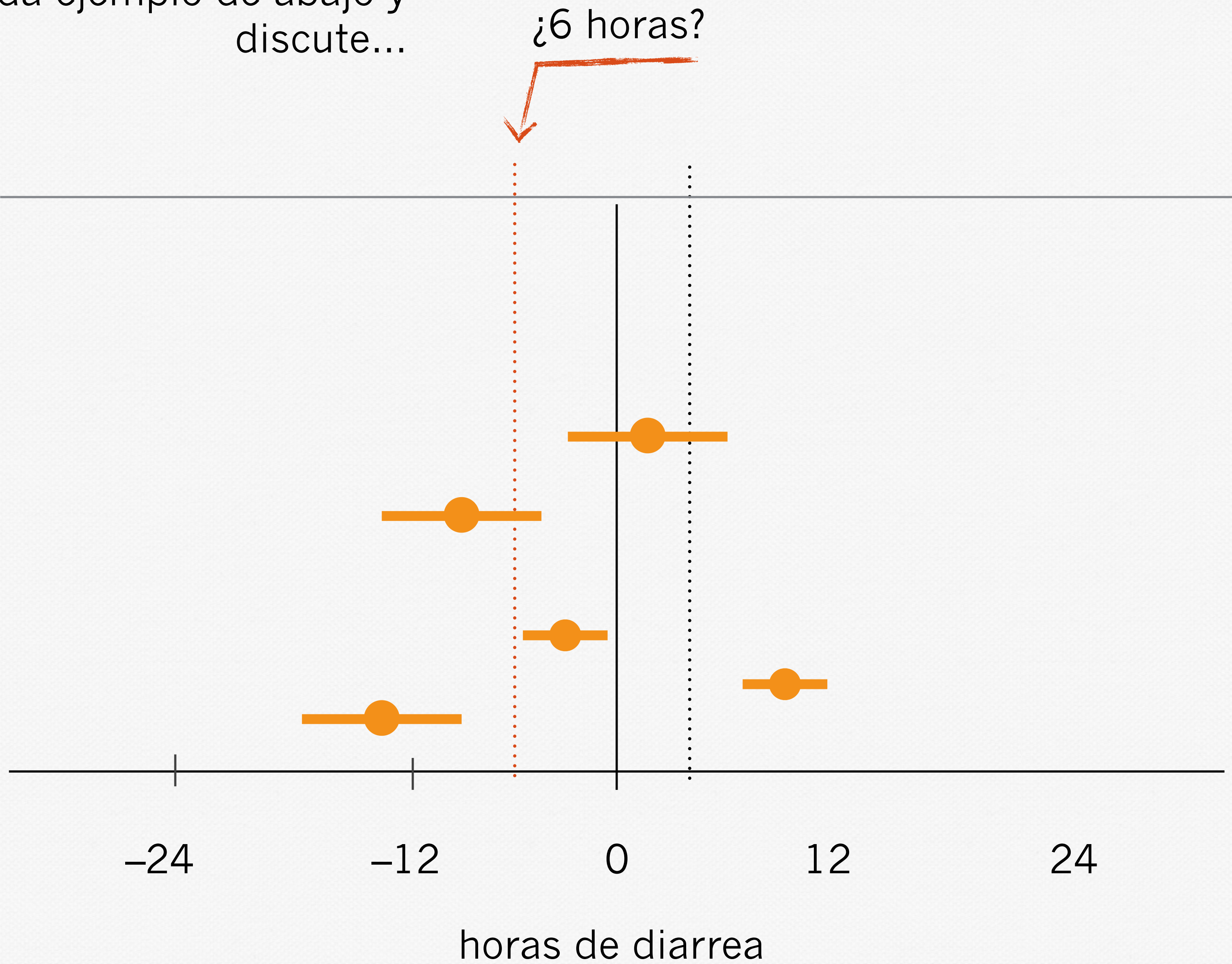
¿12 horas?



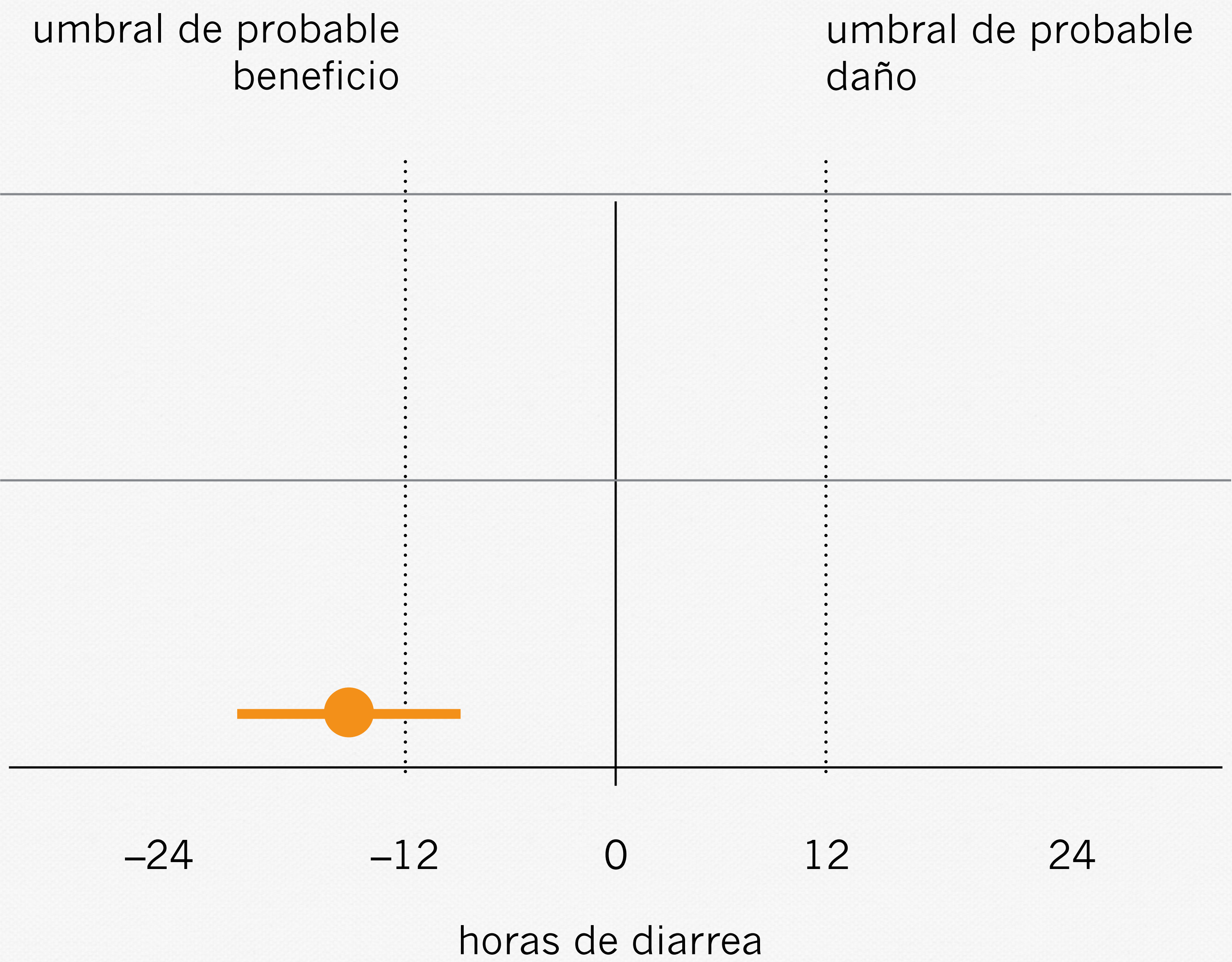
ejemplo con diarrea



con 6 horas, piensa en
cada ejemplo de abajo y
discute...

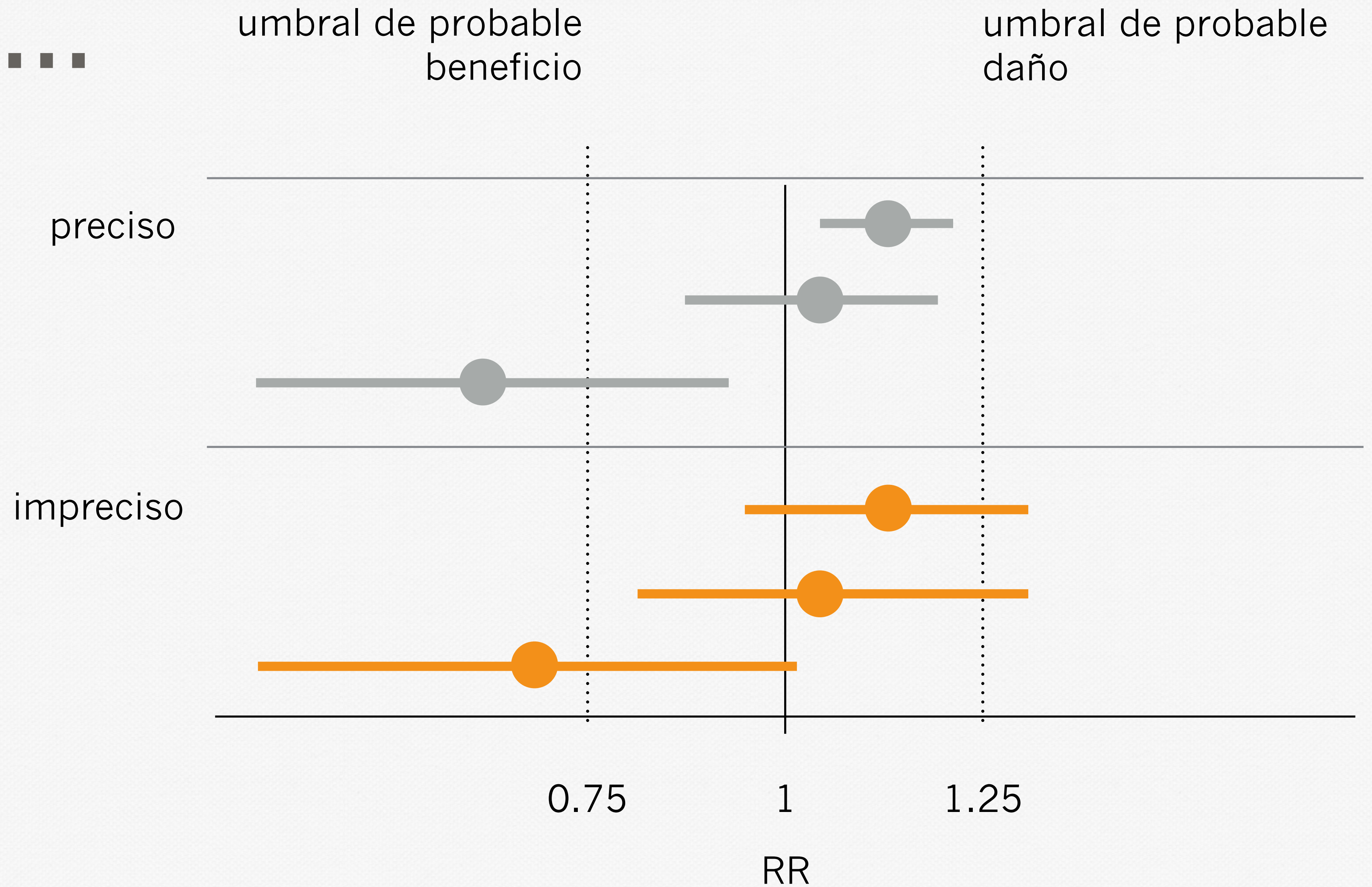


los dos umbrales



Imprecisión

en resumen...



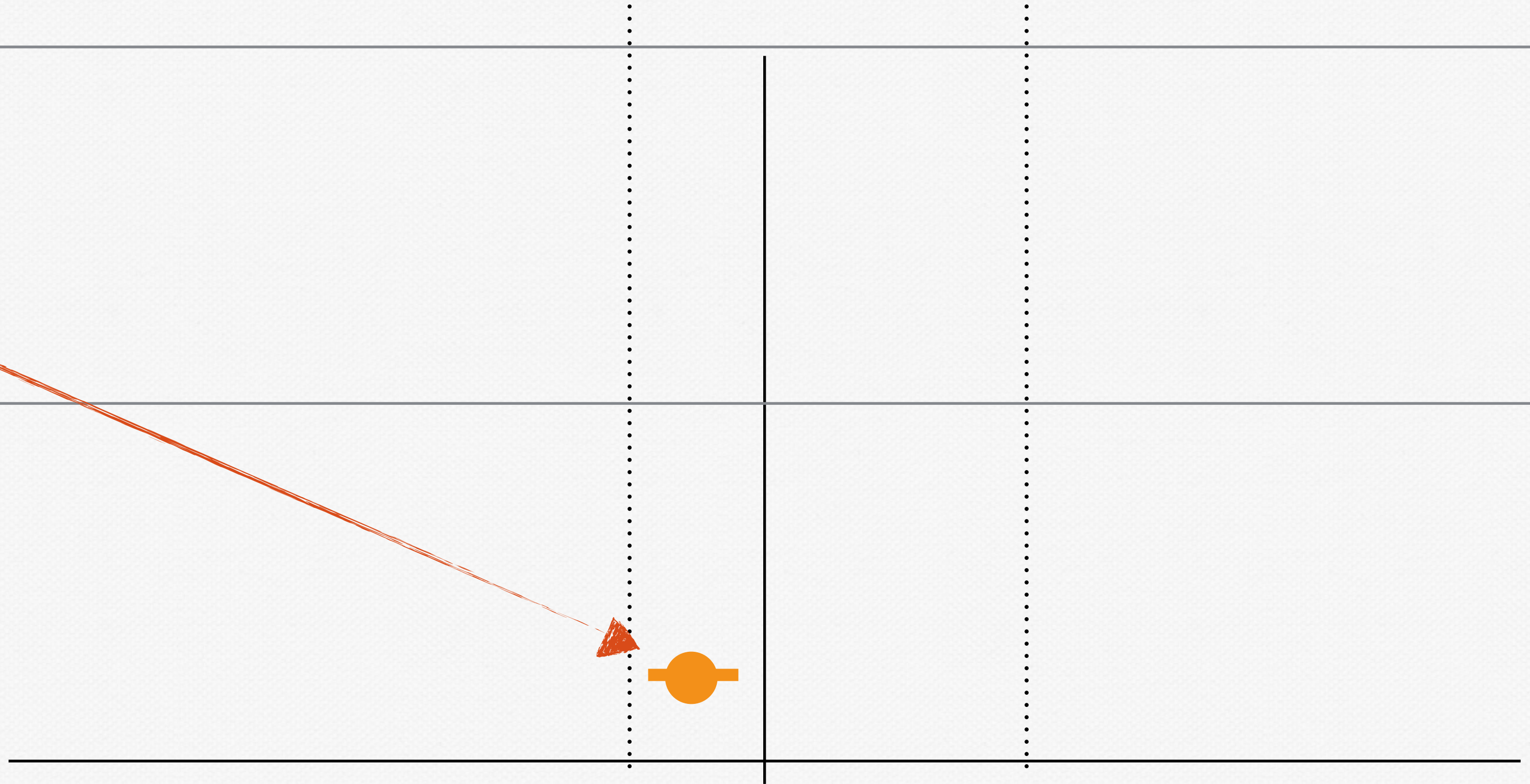
6 horas menos con
diarrea

IC95% de -9 a -2

p= 0.001

umbral de probable
beneficio

umbral de probable
daño



-24 -12 0 12 24

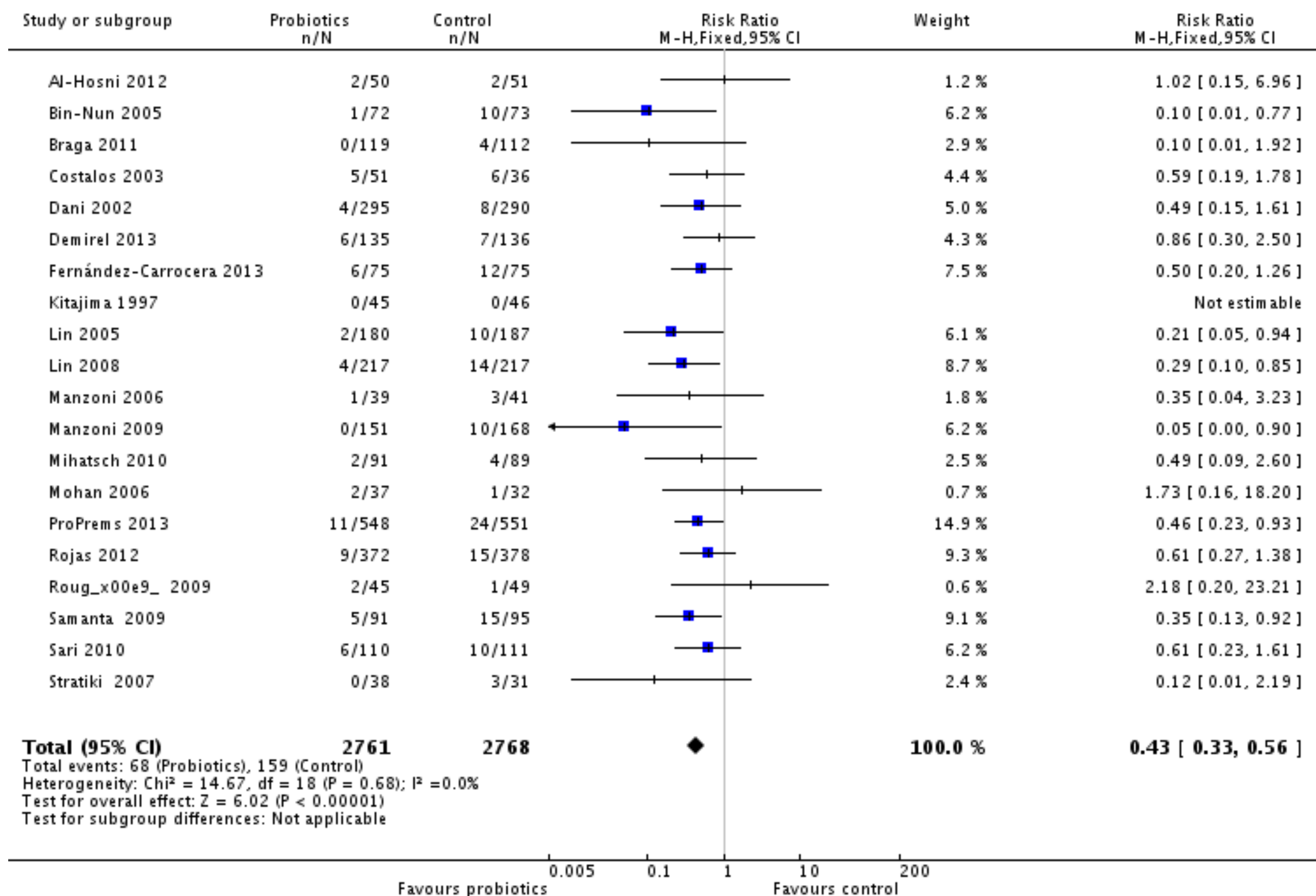
horas de diarrea

Revisión sistemática

- Mean duration of diarrhoea (mean difference 24.76 hours; 95% confidence interval 15.9 to 33.6 hours; n=4555, trials=35)
- Diarrhoea lasting ≥ 4 days (risk ratio 0.41; 0.32 to 0.53; n=2853, trials=29)
- Stool frequency on day 2 (mean difference 0.80; 0.45 to 1.14; n=2751, trials=20).

probióticos en NEC

- In a meta-analysis of trial data, enteral probiotics supplementation significantly reduced the incidence of severe NEC (stage II or more)
- Severe NEC: RR 0.43, 95% CI 0.33 to 0.56; 20 studies, 5529 infants
- Mortality: RR 0.65, 95% CI 0.52 to 0.81; 17 studies, 5112 infants



“Researchers should always report effect sizes and confidence intervals. These convey what a P value does not: the magnitude and relative importance of an effect”

“Investigadores deberían reportar el tamaño del efecto e intervalos de confianza.

Estos ofrecen lo que el valor p no puede: magnitud e importancia relativa del efecto”

¿Cuál es la evidencia?

¿Le creo a los resultados?

¿Qué hacemos ahora con estos resultados?

Los números sirven para empezar la discusión científica, no para terminarla



✉ cuelloca@mcmaster.ca

🐦 [@CharlieNeck](https://twitter.com/CharlieNeck)

