

4. *Special designs*

4.1.- *Meta-analysis*

Medical Statistics

José Antonio González y Erik Cobo

Abril 2015

Med. Stat: Metaanálisis



Análisis de resultados obtenidos en ensayos clínicos individuales con la finalidad de integrarlos (Glass 1976).

Análisis de resultados con el propósito de integrar los hallazgos (DerSimonian y Laird 1986).

Evaluación cualitativa y cuantitativa, síntesis e integración estructurada de información médica procedente, en forma de resultados, de diferentes fuentes de información o de estudios independientes sobre un mismo tema (Janieck 1989).

"When a number of quite independent tests of significance have been made, it sometimes happens that although few or none can be claimed individually significant, the aggregate gives an impression that the probabilities are, on the whole, lower than would have been obtained by chance.

It is sometimes desired, taking account only of these probabilities, not of the detailed composition of the data from which they are derived, which may be of very different kinds, to obtain a single test of the significance of the aggregate"

R.A. Fisher. Statistical Methods for Research Workers, pp99

Finalidades

- Combinación de evidencia obtenida por diversas fuentes.
- Aumentar precisión \Rightarrow potencia estadística
"All trials are too small" (Peto).
- Confirmar hipótesis.
- Desarrollar nuevas hipótesis.
- Esclarecer discrepancias.
- Protección ante multiplicidad.

Perspectivas

- Sintética: Estimación del efecto.
- Analítica: Identificación de diferencias entre estudios.

Objetivos concretos:

- Obtener un mejor estimador global.
- Valorar la consistencia entre-estudios.
- Valorar la calidad de la metodología.
- Facilitar el cálculo del tamaño muestral de futuros estudios.

Ventajas y Limitaciones

- Ventajas:

Fácil de entender y sencillo de implementar.
Más riguroso que las revisiones "narrativas".
Útil en análisis de estudios multicéntricos.
Permite valorar la heterogeneidad del efecto.

- Peligros:

Combinar efectos no homogéneos.
Combinación de manzanas, naranjas y algún limón.
Combinar estudios de diferente calidad.
Combinar de forma inapropiada los resultados.

metaanalysis is here to stay. Epidemiologists, statisticians, and clinicians should all be aware of the uses and limitations of the technique (Spector y Thompson 1991).

Método.

- 1.- Establecimiento del objetivo.
- 2.- Definición de las características de los estudios a ser incluidos.
- 3.- Recolección exhaustiva (¡sesgo de publicación!) de todos los estudios.
- 4.- Estudio conjunto (*pooled*) de toda la información.

La estimación conjunta (*pooled*) se puede obtener mediante:

- Modelado de datos originales
- Combinación de resultados agregados:
 - P-values
 - Valores 'Z'
 - Medidas específicas del efecto:
 - Odds ratio
 - Riesgo relativo
 - Diferencia tipificada entre medias
 - Coeficiente de correlación,...

I. Combinación de niveles de significación (P-valores)

Método de Fisher: $\sum_{c=1,2,\dots,C} (-2 \log P_c) \rightarrow \chi^2_{(2C)}$

Método de Edington: $P = (\sum_{c=1,2,\dots,C} P_c)^C / C!$

II. Combinación de estadísticos z, t.

Valores 'Z' $\sum_{c=1,2,\dots,C} Z_c / \sqrt{C} \rightarrow N(0,1)$

Valores 't' $\sum_{c=1,2,\dots,C} t_c / \sqrt{\sum [g_i / (g_i - 2)]} \rightarrow N(0,1)$

Ambos métodos de combinación:

- No consideran la magnitud de la asociación.
- No permiten estudiar la heterogeneidad.
- Además, los basados en p, no consideran el sentido de la asociación.

III. Combinación de estimadores del efecto.

a) Respuesta binaria:

Grupo	Evento		Total
	Si	No	
Tratamiento	a	b	n_1
Control	c	d	n_2

$$\cdot RR = (a/n_1)/(c/n_2) \quad V[\ln(RR)] = (1-p_2)/n_2p_2 + (1-p_1)/n_1p_1$$

$$\cdot OR = (ad)/(bc) \quad V[\ln(OR)] = (1/a) + (1/b) + (1/c) + (1/d)$$

$$\cdot RA = (a/n_1) - (c/n_2) \quad V[RA] = p_2(1-p_2)/n_2 + p_1(1-p_1)/n_1$$

$$\cdot NNT = 1/RA$$

$$\text{con: } p_1 = a/n_1 \text{ y } p_2 = c/n_2$$

b) Respuesta continua:

Grupo	Tamaño muestral	Respuesta Media	Desviación estándar
Tratamiento	n_1	m_1	s_1
Control	n_2	m_2	s_2

$$MD = m_1 - m_2 \quad se(MD) = \sqrt{(sd_1^2/n_1) + (sd_2^2/n_2)}$$

$$SMD = (m_1 - m_2)/s \quad se(SMD) = \sqrt{((1/n_1) + (1/n_2))} = \sqrt{((n_1 + n_2)/(n_1 n_2))} = \sqrt{N/(n_1 n_2)}$$

$$\text{donde } s = \sqrt{[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2]/(n_1 + n_2 - 2)}$$

Coef. correlación r: Fisher propuso la transformación, asintóticamente normal

$$Z = \frac{1}{2} \log[(1+r)/(1-r)] \quad \text{donde} \quad E(Z) = \frac{1}{2} \log[(1+r)/(1-r)]$$

$$V(Z) = 1/(n-3)$$

c) Resultados ajustados (1)

El ajuste consiste simplemente en:

- a) obtener estimadores ajustados por las covariantes en cada estudio.
Por ejemplo, mediante regresión logística o mediante el OR de Mantel-Haenszel y luego ponderar (W) por la inversa de su SE de estimación.
- b) realizar un modelado completo.

Respuesta continua: Regresión Lineal: β , se_{β}

Respuesta binaria:

- | | | |
|---------------|--|----------------------------|
| R. Poisson: | $RR = \exp(\beta) \rightarrow \ln(RR) = \beta$, | $se[\ln(RR)] = se_{\beta}$ |
| R. Logística: | $OR = \exp(\beta) \rightarrow \ln(OR) = \beta$, | $se[\ln(OR)] = se_{\beta}$ |
| R. Cox: | $HR = \exp(\beta) \rightarrow \ln(HR) = \beta$, | $se[\ln(HR)] = se_{\beta}$ |

c) Resultados ajustados (2).

La existencia de factores desequilibrados entre los grupos de los estudios, por ejemplo, diferente proporción de fumadores, o diferente edad o tiempo de evolución promedios, podrían ser fuentes de heterogeneidad entre los estudios.

Las estimaciones ajustadas por las posibles covariantes:

- disminuirán la heterogeneidad entre los estudios
- disminuirán la varianza residual (en el caso de respuestas cuantitativas)
- disminuirán las 'n' de los subgrupos (en el caso de respuestas cualitativas)

Debe recordarse que este ajuste, sólo tiene sentido en ausencia de interacción respuesta-tratamiento-covariante, es decir: si la covariante no es modificadora del efecto.

IV. Método de Wolf, para combinar estimadores.

Sean C estudios o ensayos Clínicos $c = 1, 2, \dots, C$

Sean Y_c el efecto en el ensayo c

$V(Y_c)$ la varianza de la estimación Y_c

$W_c = 1 / V(Y_c)$ la ponderación asignada al estudio c

El estimador del efecto común, ponderando por el inverso de la varianza, es:

$$\underline{Y} = \Sigma_c W_c Y_c / \Sigma_c W_c$$

A) Modelo de efectos fijos:

Considera C como el conjunto de todos los estudios a los que se quiere hacer inferencia o bien que todos tienen un efecto común constante.

⇒ Queremos ver si se ha demostrado un efecto.

$$V(\underline{Y}) = (1/\Sigma W_c)^2 \cdot \Sigma W_c^2 V(Y_c) = (1/\Sigma W_c)^2 \cdot \Sigma W_c^2 / W_c = (1/\Sigma W_c)^2 \cdot \Sigma W_c = (1/\Sigma W_c)$$

$$S.E.(\underline{Y}) = (\Sigma W_c)^{-1/2}$$

Permite realizar la inferencia habitual, sea IC o PH, sobre \underline{Y} .



Origen de diferentes estimaciones puntuales del efecto entre estudios

- a) Variabilidad aleatoria del muestreo. Es función de n y σ^2 .
- b) Errores no aleatorios en el proceso de estimación (por ejemplo, EC de baja calidad metodológica).
- c) Heterogeneidad: el efecto que se estima no es constante.

Homogeneidad del efecto:

La medida común del efecto \underline{Y} tendrá pleno sentido (incluso para estudios futuros) cuando el efecto Y_c sea el mismo en todos los estudios.

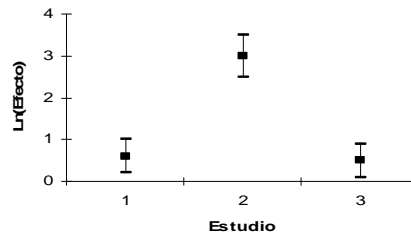
Lo que se conoce como ausencia de interacción (ensayo-tratamiento-respuesta), y "combinabilidad"

Para ponerla a prueba, puede utilizarse el estadístico

$$Q = \Sigma_c W_c (Y_c - \underline{Y})^2 \quad \text{que sigue una distribución } \chi^2_{C-1}$$



Ejemplo:



Estudio	θ_i	se_i	V_i	w_i	$\theta_i \cdot w_i$	%
1	0.6	(0.21)	0.045	22	13.2	32.8
2	3.0	(0.26)	0.067	15	45.0	22.4
3	0.5	(0.18)	0.033	30	15.0	44.8
Σ				67	73.2	100.0

Prueba de heterogeneidad : $P(\chi^2_{2} > 70.4) < 0.001$

Conclusiones: Investigar estudio 2.

¿Y si no se encuentra explicación para la heterogeneidad?

B) Modelo de efectos aleatorios:

C es una muestra de todos los estudios a los que se quiere hacer inferencia y el efecto varía de uno a otro \Rightarrow Se quiere aplicar el efecto a todas las posibles situaciones.

Asume cierta variabilidad del efecto entre los estudios, alrededor de su valor central y, por tanto, la heterogeneidad es, aquí, axiomática y no tiene sentido la prueba de interacción.

Ahora, incluso para muestras infinitas en cada estudio, los Y_c variarán.

Sean $Q = \sum_c W_c (Y_c - \bar{Y})^2$ (igual que antes)

σ^2_D = el componente de la varianza interestudios -desconocida- del efecto Y_c

Siendo $\bar{W} = \sum_c W_c / C$ la media de los pesos y $S^2_W = (\sum_c W_c^2 - C\bar{W}^2) / (C-1)$

La estimación D (Dersimonian and Laird) de σ^2_D es

$D = (Q - (C-1)) / U$ si $Q > C-1$ Donde $U = (C-1)(\bar{W} - S^2_W / C\bar{W})$

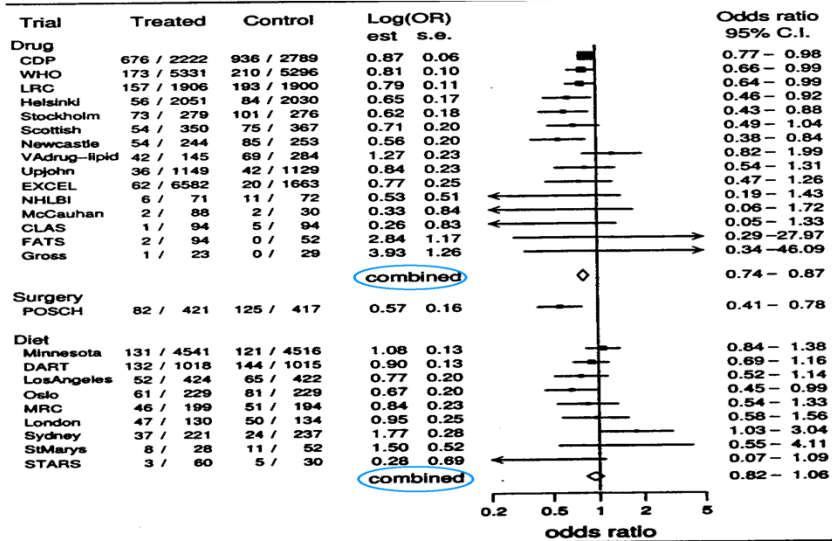
$D = 0$ si $Q \leq C-1$

Se estima el efecto: $\bar{Y}^* = \sum_c W^*_c Y_c / \sum_c W^*_c$ Donde $W^*_c = (D + W_c)^{-1}$

Graficos para el meta análisis

Forest plot (grafico del bosque, global,...)

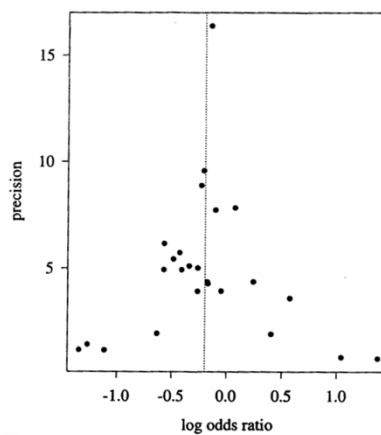
1. Ordenar los intervalos por longitud
2. Reducir el espacio ocupado por los intervalos
3. Añadir texto: etiquetas e información numérica
4. Idealmente, debería tener:
 - Un IC para cada estudio (usualmente, al 95%)
 - Una estimación puntual marcada con un cuadrado de área proporcional al peso (e inversamente proporcional a la variancia)
 - Una línea vertical donde $OR=1$
 - La estimación combinada (resultado global del MA), marcada con un rombo, centrado en la estimación puntual, y cuya amplitud coincide con los límites del IC.
 - Cortar intervalos muy amplios, para evitar que se "coman" el gráfico.



Funnel plot (Gráfico en embudo)

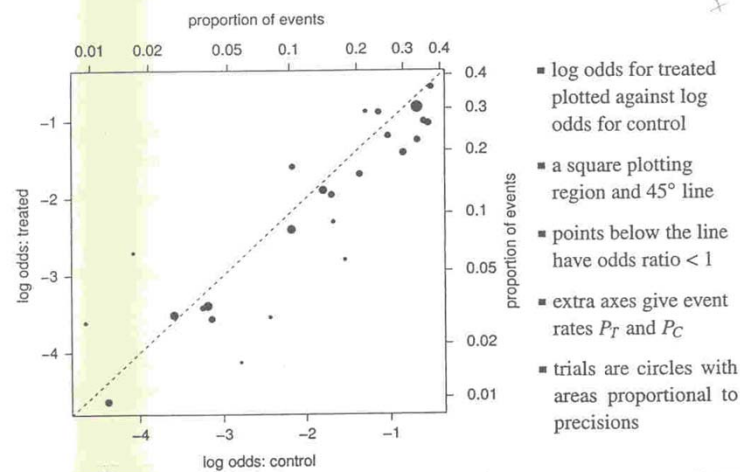
Representa: en abscisas la “señal” y en ordenadas el “ruido”

1. Se representa el tamaño de muestra frente al efecto para cada estudio
 - Los efectos para muestras pequeñas se dispersan más que aquellos de muestras grandes
 - En ausencia de sesgos (y con un efecto del tratamiento constante), los puntos deberían disponerse simétricamente respecto una línea vertical, posiblemente en forma de embudo
 - Asimetría –especialmente, ausencia de puntos con baja precisión y efectos “negativos”– pueden indicar un sesgo de selección.
2. Mejor representar la raíz cuadrada del tamaño muestral o 1/se en vez del tamaño muestral

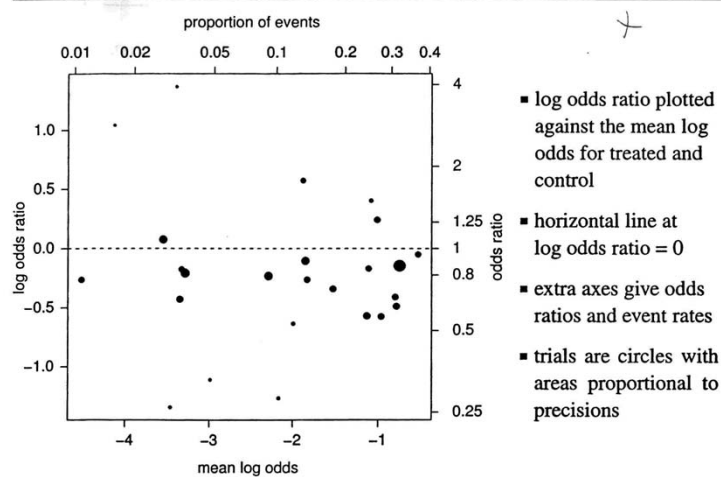
Funnel plot of the IHD data

- $\frac{1}{se(\hat{\eta})}$ plotted against $\hat{\eta}$
where $\eta = \log \text{ odds ratio}$
- vertical line shows the combined estimate
- low precision points scatter more
- here the scatter looks quite symmetrical about the vertical line

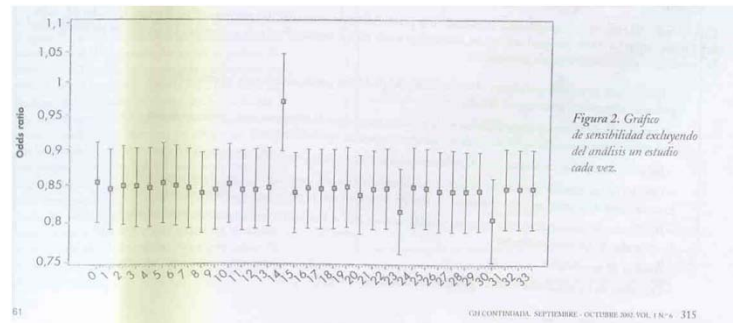
L'Abbe plot for the IHD data



Oldham plot for the IHD data



Sensibilidad



Este gráfico excluye del análisis un estudio cada vez.
 Cuando deja fuera el número 14, no hay significación estadística.
 En cambio, si se incluye, sí.
 Este metaanálisis no es robusto, ya que “descansa” en un único estudio.

2. Combinación de evidencia no experimental

Principio de comparabilidad (Colton):

“Los grupos son comparables si difieren sólo en el tratamiento asignado.”

Se ha visto que, en un entorno experimental, ensayo clínico, mediante la asignación aleatoria (que permite grupos “idénticos” al inicio) y el seguimiento y evaluación ‘a ciegas’ (que permite grupos “idénticos” durante y al final) es razonable este principio de comparabilidad pero tiene implicaciones éticas.

En Epidemiología, cuando se estudian causas de efectos negativos, como el tabaco, estas limitaciones éticas requieren un entorno no experimental. La discusión es si en ausencia de diseño experimental es correcto establecer relaciones de causa – efecto.

Fisher: en ausencia de diseño experimental, siempre puede existir un factor de confusión no controlado.

Bradford Hill: aunque no exista evidencia experimental, la observación conjunta de una serie de 'factores', hace más verosímil la explicación causal de la asociación observada.

Factores:

- *Temporality*: ordenación temporal
- *Pseudo-experiment*: suprimida la supuesta causa, baja el efecto.
- *Biological Gradient*: monotonicidad de la relación.
- *Strength*: intensidad de la relación: R^2 .
- *Specificity*: no presencia en no expuestos.
- *Consistency*: igual relación en diferentes poblaciones.
- *Plausibility, Coherence, Analogy*: coherencia teórica.

Ejemplo: discusión sobre la relación entre el tabaco y el cáncer de pulmón. No se puede experimentar con el tabaco, de forma que su relación con el cáncer de pulmón no se puede establecer como causa – efecto formalmente.

Fisher afirma que la causalidad del tabaco y el cáncer puede estar confundida por otra variable: ¿puede un gen provocar ambos fenómenos?

Bradford Hill afirma que en ciertas condiciones, lo más verosímil es la causalidad:

- Relación intensa, simple, monótona (incluso lineal).
- Consistencia en diferentes poblaciones, estudios.
- Pseudo experimentación (mujeres: fuman más y aumenta cáncer de pulmón).
- Plausibilidad biológica, coherencia teórica y parsimonia.

Pregunta: ¿Concuerdan los factores de Hill con la definición formal de causalidad y sus premisas?

Med. Stat: Metaanálisis



En Farmacoepidemiología se precisa imputar causalmente a un fármaco, cierto fenómeno no deseable:

Acontecimiento Adverso (AA): Mera observación de un acontecimiento no deseable

Reacción o Efecto Adverso: AA que puede ser imputado al fármaco en cuestión

Criterios de la Comunidad Europea:

1. Asociación en el tiempo, incluida retirada y reexposición, o lugar de exposición
2. Explicación farmacológica, niveles en sangre, conocimiento previo.
3. Presencia de fenómenos clínicos o patológicos característicos.
4. Exclusión de otras causas; ausencia de explicaciones alternativas.
5. Que los informes de los casos sean fidedignos y estén completos.
6. Medición cuantitativa del grado de contribución del fármaco al desarrollo de la reacción

La **normativa en España** distingue entre:

Improbable/Condicional/Posible/Probable/Definitiva.

Pregunta: ¿Concuerdan los criterios de la CE con los factores de Hill?

