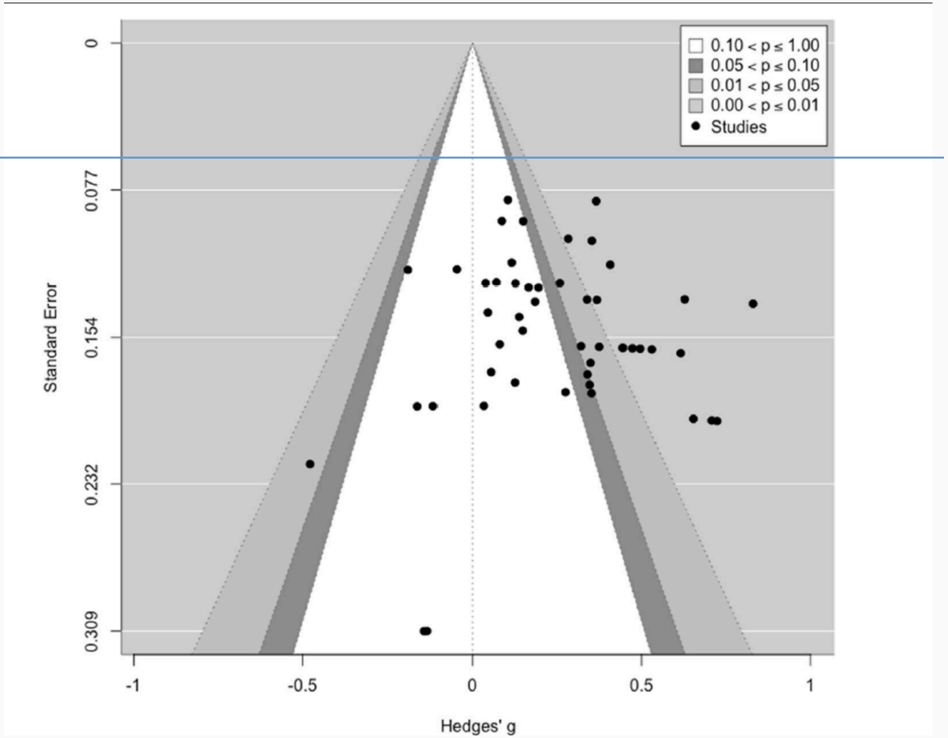
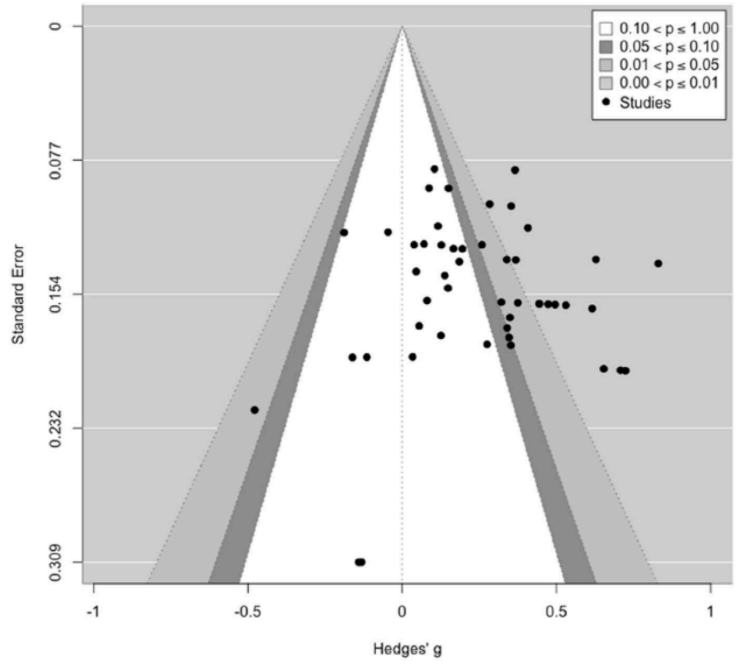
# Statistical Theory for Data Science

STA2212H S LEC9101



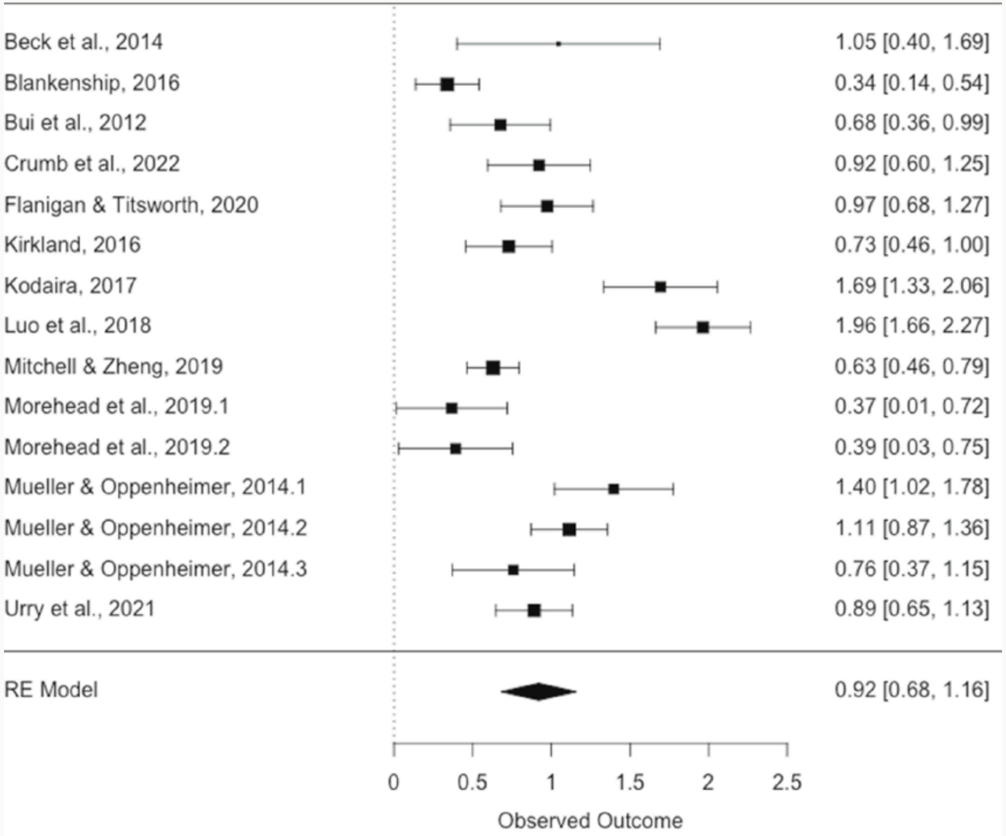

Week 7

February 24 2026




***Note.*** A positive standardized effect size indicates that students who handwrote their notes had higher achievement than those who typed their notes.

**Fig. 1** Funnel plot depicting Hedges' $g$ and precision in assessing achievement: handwritten vs. typed note-taking



# Today

1. Midterm 2 March 10
2. Recap: Multiple testing, reproducibility
3. Papers re project      Google sheet
4. Inference with missing data
5. Meta-analysis

guidelines needed

# Recap: Multiple testing

- $p$-values $p_1, \ldots, p_m$, typically arising from many similar tests on the same data
- e.g. two-sample tests for expressions levels of each of $m$ genes/proteins/test scores/...
- family-wise error rate (fwer) controlled by ensuring

$$\mathrm{pr}(\text{any true null rejected}) \leq \alpha$$

- Bonferroni correction: reject $H_{0i}$ if $p_i \leq \alpha/m$

$\alpha = 0.05$ is conventional

# Recap: Multiple testing

- $p$-values $p_1, \ldots, p_m$, typically arising from many similar tests on the same data
- e.g. two-sample tests for expressions levels of each of $m$ genes/proteins/test scores/...
- family-wise error rate (fwer) controlled by ensuring

$$\mathrm{pr}(\text{any true null rejected}) \le \alpha$$

- Bonferroni correction: reject $H_{0i}$ if $p_i \le \alpha/m$ — $\alpha = 0.05$ is conventional

- false discovery rate (fdr) controlled using the BH method — Benjamini & Hochberg 1995
- order the $p$-values $p_{(1)} \le \cdots \le p_{(m)}$: reject $H_{0i}$ for all $i$ where $p_{(i)} \le \frac{i}{m} q$

  $q = 0.1$ is conventional

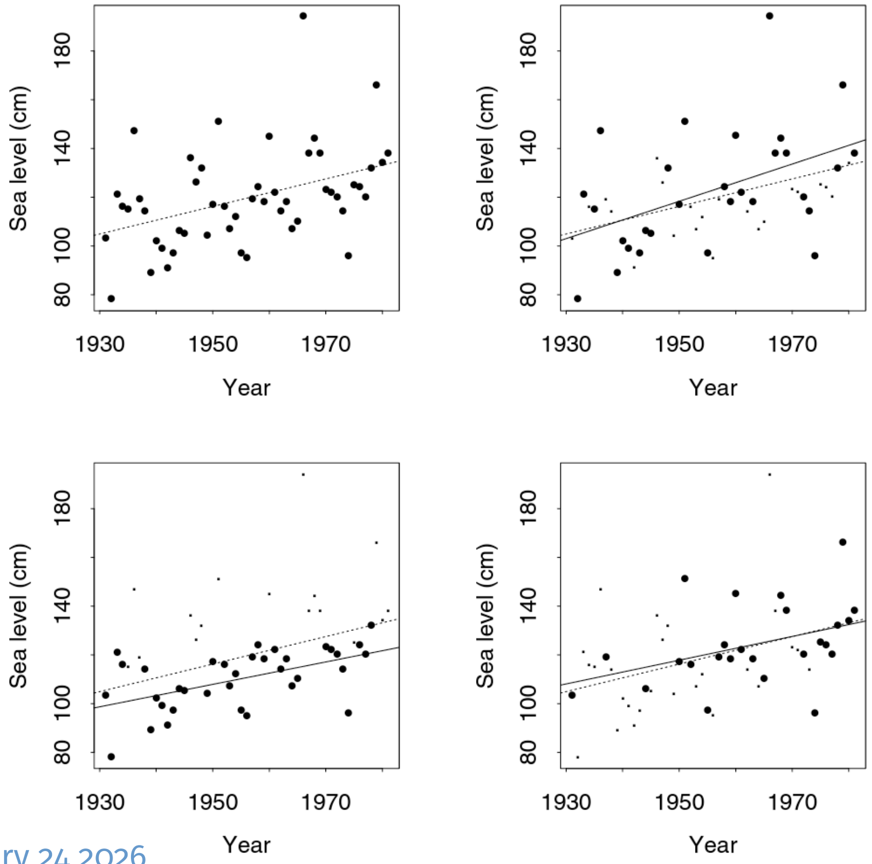
- a conservative correction available for dependent $p$-values

# Missing data





**Figure 5.12** Missing data in straight-line regression for Venice sea-level data. Clockwise from top left: original data, data with values missing completely at random, data with values missing at random — missingness depends on $x$ but not on $y$, and data with non-ignorable non-response — missingness depends on both $x$ and $y$. Missing values are represented by a small dot. The dotted line is the fit from the full data, the solid lines those from the non-missing data.



# Models with missing data



- context: independent observations $(y_i, x_i), i = 1, \dots, n$ — $x_i$ could be a vector
- model $f(\boldsymbol{y} \mid \boldsymbol{x}; \theta)$ or sometimes $f(\boldsymbol{y}, \boldsymbol{x}; \theta)$ — linear regression; glm; etc
- some observations on $\boldsymbol{y}$ may be missing

# Models with missing data

- context: independent observations $(y_i, x_i), i = 1, \ldots, n$ — $x_i$ could be a vector
- model $f(\boldsymbol{y} \mid \boldsymbol{x}; \theta)$ or sometimes $f(\boldsymbol{y}, \boldsymbol{x}; \theta)$ — linear regression; glm; etc
- some observations on $\boldsymbol{y}$ may be missing
- e.g. clinical trial, $x_i$ covariate(s) measured at baseline,
  $y_i$ response after treatment, or after some time has elapsed
- observation on subject $i$ becomes $(y_i, x_i, R_i)$, $R_i = 1$ for complete observation
  $R_i = 0$ for incomplete observation

# Models with missing data

- context: independent observations $(y_i, x_i), i = 1, \dots, n$ — $x_i$ could be a vector
- model $f(\boldsymbol{y} \mid \boldsymbol{x}; \theta)$ or sometimes $f(\boldsymbol{y}, \boldsymbol{x}; \theta)$ — linear regression; glm; etc
- some observations on $\boldsymbol{y}$ may be missing
- e.g. clinical trial, $x_i$ covariate(s) measured at baseline,
  $y_i$ response after treatment, or after some time has elapsed
- observation on subject $i$ becomes $(y_i, x_i, R_i)$, $R_i = 1$ for complete observation
  $R_i = 0$ for incomplete observation
- contribution to likelihood function from complete observation

$$f(y_i, x_i, R_i; \theta) = \mathrm{pr}(R_i = 1 \mid x_i, y_i) f(y_i \mid x_i; \theta) f(x_i; \theta)$$

- contribution to likelihood function from incomplete observation — no $\theta$

$$f(x_i, R_i; \theta) = \int \mathrm{pr}(R_i = 0 \mid x_i, y_i) f(y_i \mid x_i; \theta) f(x_i; \theta) dy_i$$

in usual regression settings, $f(x_i)$

# Types of missing data

- contribution to likelihood function from incomplete observation $R_i = 0$

$$f(x_i, R_i; \theta) = \int \mathrm{pr}(R_i = 0 \mid x_i, y_i) f(y_i \mid x_i; \theta) f(x_i; \theta) dy_i$$

- missing completely at random: $\mathrm{pr}(R_i = 0 \mid x_i, y_i) = \mathrm{pr}(R_i = 0)$ MCAR
- missing at random: $\mathrm{pr}(R_i = 0 \mid x_i, y_i) = \mathrm{pr}(R_i = 0 \mid x_i)$ MAR
- non-ignorable non-response $\mathrm{pr}(R_i = 0 \mid x_i, y_i)$ no simplification

- contribution to likelihood function from incomplete observation $R_i = 0$

$$f(x_i, R_i; \theta) = \int \mathrm{pr}(R_i = 0 \mid x_i, y_i) f(y_i \mid x_i; \theta) f(x_i; \theta) dy_i$$

- missing completely at random: $\mathrm{pr}(R_i = 0 \mid x_i, y_i) = \mathrm{pr}(R_i = 0)$ MCAR
- missing at random: $\mathrm{pr}(R_i = 0 \mid x_i, y_i) = \mathrm{pr}(R_i = 0 \mid x_i)$ MAR
- non-ignorable non-response $\mathrm{pr}(R_i = 0 \mid x_i, y_i)$ no simplification

- likelihood function for sample $(y_i, x_i, R_i), i = 1, \ldots, n$

$$L(\theta; \mathbf{R}, \mathbf{x}, \mathbf{y}) = \prod_{i \in M} \int \mathrm{pr}(R_i = 0 \mid x_i, y_i) f(y_i \mid x_i; \theta) f(x_i; \theta) dy_i$$

$$\times \prod_{i \notin M} \mathrm{pr}(R_i = 1 \mid x_i, y_i) f(y_i \mid x_i; \theta) f(x_i; \theta)$$

# Likelihood function

- likelihood function for sample $(y_i, x_i, R_i), i = 1, \ldots, n$

$$L(\theta; \mathbf{R}, \mathbf{x}, \mathbf{y}) = \prod_{i \in \mathcal{M}} \int \mathrm{pr}(R_i = 0 \mid x_i, \cancel{y_i}) f(y_i \mid x_i; \theta) \underline{f(x_i; \theta)} dy_i \times$$

$$\prod_{i \notin \mathcal{M}} \mathrm{pr}(R_i = 1 \mid x_i, y_i) f(y_i \mid x_i; \theta) f(x_i; \theta)$$

(i)

$$\boxed{Pr(R_i = 0 \mid y_i, x_i) = Pr(R_i = 0 \mid x_i)}$$

$$\boxed{f(x_i; \theta) = f(x_i)}$$

$$= \prod_{i \in M} \underbrace{pr(R_i = 0 \mid x_i) f(x_i; \cancel{\theta})}_{} \underbrace{\int f(y_i \mid x_i; \theta) dy_i}_{\equiv 1}$$

$$\times \prod_{i \notin M} \{f(y_i \mid x_i; \theta)\} f(x_i; \cancel{\theta}) \underline{Pr(R_i = 1 \mid x_i; \cancel{y_i})}$$

$$= \prod_{i \in M} c_i \times \prod_{i \notin M} c_i \cdot f(y_i \mid x_i; \theta) \, Pr(R_i = 1 \mid x_i)$$

$$\propto \prod^{n} \{f(y_i \mid x_i; \theta)\}^{r_i}$$

# Likelihood function



i=

- likelihood function for sample $(y_i, x_i, R_i), i = 1, \ldots, n$

$$L(\theta; \boldsymbol{R}, \boldsymbol{x}, \boldsymbol{y}) = \prod_{i \in \mathcal{M}} \int \mathrm{pr}(R_i = 0 \mid x_i, y_i) f(y_i \mid x_i; \theta) f(x_i; \theta) dy_i \times$$
$$\prod_{i \notin \mathcal{M}} \mathrm{pr}(R_i = 1 \mid x_i, y_i) f(y_i \mid x_i; \theta) f(x_i; \theta)$$

- under MAR or MCAR ,

$$\boxed{L(\theta; \boldsymbol{R}, \boldsymbol{x}, \boldsymbol{y}) \propto \prod_{i=1}^{n} \{f(y_i \mid x_i; \theta)\}^{r_i} f(x_i; \theta)}$$

dist$^n$ of $R \mid x, y$ can't depend on $\theta$

# Likelihood function



- likelihood function for sample $(y_i, x_i, R_i), i = 1, \ldots, n$

$$L(\theta; \boldsymbol{R}, \boldsymbol{x}, \boldsymbol{y}) = \prod_{i \in \mathcal{M}} \int \mathrm{pr}(R_i = 0 \mid x_i, y_i) f(y_i \mid x_i; \theta) f(x_i; \theta) dy_i \times \prod_{i \notin \mathcal{M}} \mathrm{pr}(R_i = 1 \mid x_i, y_i) f(y_i \mid x_i; \theta) f(x_i; \theta)$$

- under MAR or MCAR ,

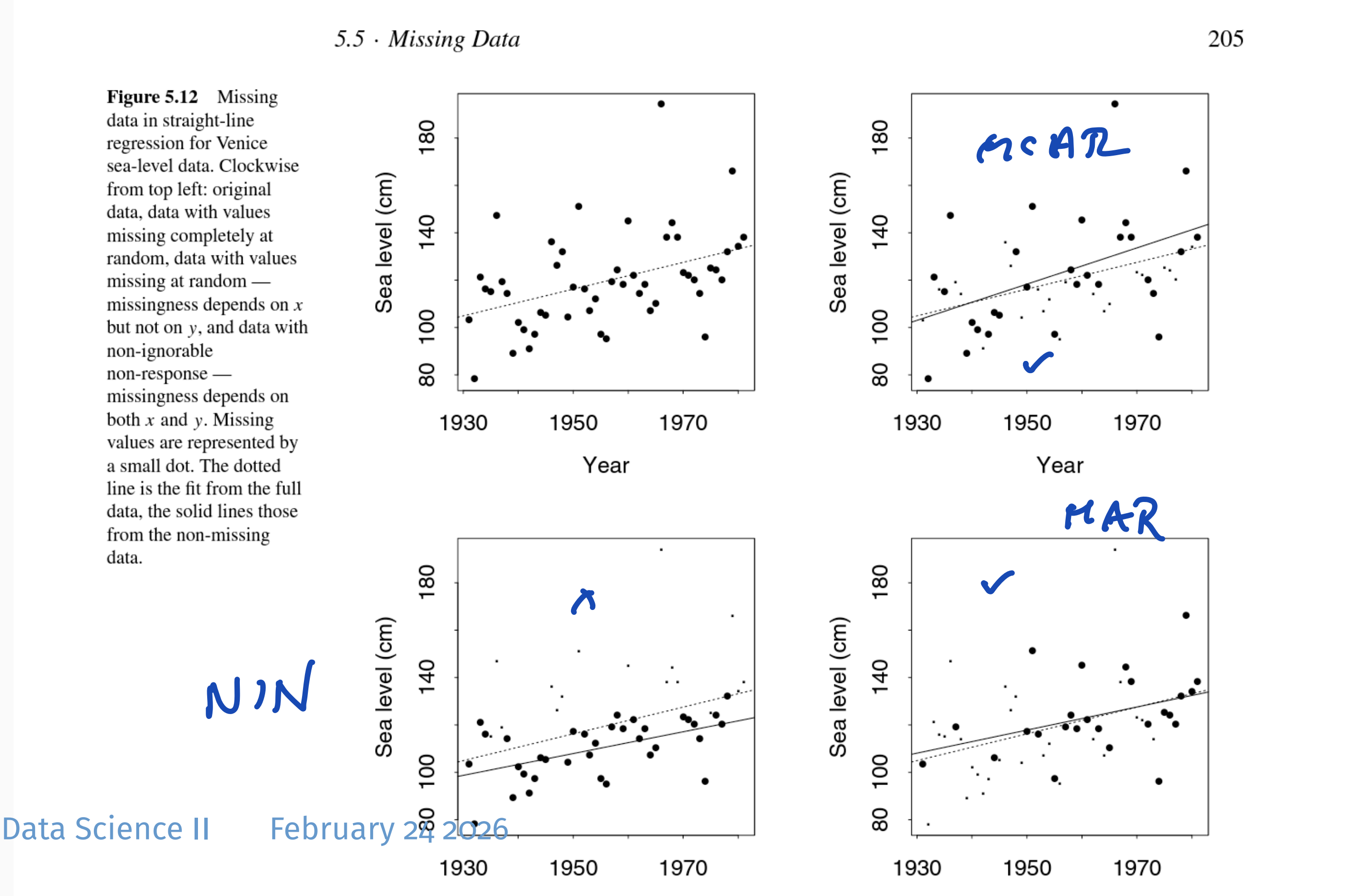
$$L(\theta; \boldsymbol{R}, \boldsymbol{x}, \boldsymbol{y}) \propto \prod_{i=1}^{n} \{f(y_i \mid x_i; \theta)\}^{r_i} f(x_i; \theta)$$

- usually $f(x_i)$ free of $\theta$, so $L(\theta) \propto \prod_{i=1}^{n} \{f(y_i \mid x_i; \theta)\}^{r_i} = \underline{\prod_{complete\ cases} f(y_i \mid x_i; \theta)}$
- expected information $I(\theta) = \underline{\mathrm{E}_\theta\{-\ell''(\theta)\}}$ will depend on $\underline{\mathrm{pr}(R_i = 1)}$
- use observed information $J(\hat{\theta}) = -\ell''(\hat{\theta})$ for estimating standard error of MLE

# Example



## Annual maximum sea-level in Venice, 1931 – 1981




*5.5 · Missing Data* 205

**Figure 5.12** Missing data in straight-line regression for Venice sea-level data. Clockwise from top left: original data, data with values missing completely at random, data with values missing at random — missingness depends on $x$ but not on $y$, and data with non-ignorable non-response — missingness depends on both $x$ and $y$. Missing values are represented by a small dot. The dotted line is the fit from the full data, the solid lines those from the non-missing data.

# Linear model and simulations

```
> faraway::sumary(venice.lm)
                     Estimate Std. Error t value  Pr(>|t|)
(Intercept)         119.60784    2.60729 45.8744 < 2.2e-16
I(year - mean(year))  0.56697    0.17713  3.2009  0.002406

n = 51, p = 2, Residual SE = 18.61977, R-Squared = 0.17
```

simulate 1000 samples from linear model with $\beta_0 = 120$, $\beta_1 = 0.5$, $\sigma = 20$

```
> faraway::sumary(venice.lm)
                     Estimate Std. Error t value  Pr(>|t|)
(Intercept)         119.60784    2.60729 45.8744 < 2.2e-16
I(year - mean(year))  0.56697    0.17713  3.2009  0.002406

n = 51, p = 2, Residual SE = 18.61977, R-Squared = 0.17
```

simulate 1000 samples from linear model with $\beta_0 = 120$, $\beta_1 = 0.5$, $\sigma = 20$

generate missing data indicators as

$$\mathrm{pr}(R_i = 1 \mid x_i, y_i) = \begin{cases} 0.5, \\ \Phi\{0.05(x_i - \bar{x})\}, \\ \Phi[0.05(x_i - \bar{x}) + \{y_i - \beta_0 - \beta_1(x_i - \bar{x})\}/\sigma] \end{cases}$$

MCAR
MAR
NIN

not ignorable

$$\mathrm{pr}(R = 1 \mid x, y) = \begin{cases} 0.5, \\ \Phi\{0.05(x - \bar{x})\}, \\ \Phi[0.05(x - \bar{x}) + \{y - \beta_0 - \beta_1(x - \bar{x})\}/\sigma] \end{cases}$$

| | Average estimate (average standard error) | | | | |
|---|---|---|---|---|---|
| | Truth | Full | MCAR | MAR | NIN |
| $\beta_0$ | 120 | 120 (2.79) | 120 (4.02) | 120 (4.73) | 132 (3.67) |
| $\beta_1$ | 0.50 | 0.49 (0.19) | 0.48 (0.28) | 0.50 (0.32) | 0.20 (0.25) |

**Table 5.8** Average estimates and standard errors for missing value simulation based on Venice data, for full dataset, with data missing completely at random (MCAR), missing at random (MAR) and with non-ignorable non-response (NIN). 1000 samples were taken. Standard errors for the averages for $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are at most 0.16 and 0.01; those for their standard errors are at most 0.05 and 0.002.

To assess the extent of this bias, we generated 1000 samples from a model with parameters $\beta_0 = 120$, $\beta_1 = 0.5$ and $\sigma = 20$, close to the estimates for the Venice data and with the same covariate $x$. We then computed maximum likelihood estimates for the full data and for those observations that remain after applying the non-response mechanisms

# Example: Publication bias





| Trial | Magnesium $r/m$ | Control $r/m$ | $n$ | $\widehat{\mu}$ | $(v/n)^{1/2}$ |
|---|---|---|---|---|---|
| 1 | 1/25 | 3/23 | 48 | 1.18 | 1.05 |
| 2 | 1/40 | 2/36 | 76 | 0.80 | 0.83 |
| 3 | 2/48 | 2/46 | 94 | 0.04 | 0.75 |
| 4 | 1/50 | 9/53 | 103 | 2.14 | 0.72 |
| 5 | 4/56 | 14/56 | 112 | 1.25 | 0.69 |
| 6 | 3/66 | 6/66 | 132 | 0.69 | 0.63 |
| 7 | 2/92 | 7/93 | 185 | 1.24 | 0.53 |
| 8 | 27/135 | 43/135 | 270 | 0.47 | 0.44 |
| 9 | 10/160 | 8/156 | 316 | −0.20 | 0.41 |
| 10 | 90/1159 | 118/1157 | 2316 | 0.27 | 0.15 |
| Meta-analysis | | | 3652 | 0.41 | 0.11 |
| ISIS-4 | 2216/29011 | 2103/29039 | 58050 | −0.05 | 0.03 |

**Table 5.9** Data from 11 clinical trials to compare magnesium treatment for heart attacks with control, with $n$ patients randomly allocated to treatment and control; there are $r$ deaths out of $m$ patients in each group (Copas, 1999). The estimated log treatment effect $\widehat{\mu}$ will be positive if treatment is effective; $(v/n)^{1/2}$ is its standard error. The huge ISIS-4 trial is not included in the meta-analysis.

# A model for selection bias



- study with $n$ individuals leads to estimate $\hat{\mu} \dot{\sim} N(\mu, \sigma^2/n)$
- study is published ($R = 1$), if $Z > 0$

some measure of randomness in publication

- Suppose $\hat{\mu}$ and $Z$ are related according to the model

$Y \to \hat{\mu}, X \to n$, both pot. missing

$$\hat{\mu} = \mu + \sigma n^{-1/2} U_1, \quad Z = \gamma_0 + \gamma_1 n^{1/2} + U_2, \quad \mathrm{cor}(U_1, U_2) = \rho \geq 0$$

$$U_1 = \frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}}$$

$U_1 =$

# A model for selection bias

- study with $n$ individuals leads to estimate $\hat{\mu} \dot{\sim} N(\mu, \sigma^2/n)$
- study is published ($R = 1$), if $Z > 0$ — some measure of randomness in publication
- Suppose $\hat{\mu}$ and $Z$ are related according to the model — $\underline{Y \to \hat{\mu}}, \underline{X \to n}$, both pot. missing

$$\hat{\mu} = \mu + \sigma n^{-1/2} U_1, \quad Z = \gamma_0 + \gamma_1 n^{1/2} + U_2, \quad \mathrm{cor}(U_1, U_2) = \rho \geq 0$$

$U_1 =$

- $\underline{\mathrm{pr}(R = 1)} = \underline{\mathrm{pr}(Z > 0) = \Phi(\gamma_0 + \gamma_1 n^{1/2})}$
- $\underline{\mathrm{pr}(R = 1 \mid \hat{\mu})} = \mathrm{pr}(Z > 0 \mid \hat{\mu}) = \Phi \underline{\left\{ \frac{\gamma_0 + \gamma_1 n^{1/2} + \rho n^{1/2}(\hat{\mu} - \mu)/\sigma}{(1 - \rho^2)^{1/2}} \right\}}$ — MV Normal
- <u>non-ignorable</u> non-response — depends on $\hat{\mu}$ (y) — unless $\rho = 0$

$$\begin{pmatrix} \underline{Y}_1 \\ \underline{Y}_2 \end{pmatrix} \sim N\left( \begin{pmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right) \qquad Y_1 \mid Y_2 = y_2 \sim N( \quad , \quad )$$

$$Y_1 \mid \underbrace{Y_2 = y_2} \sim N\Big(\mu_1 \overset{?}{+} \Sigma_{12}\Sigma_{22}^{-1}(y_2-\mu_2),$$

$$\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Big)$$

$$\hat{\beta} = \underbrace{(X^TX)^{-1}}_{\Sigma_{22}} \underbrace{X^Ty}_{\Sigma_{21}}$$

$$\begin{pmatrix}\hat{\mu} \\ Z\end{pmatrix} \sim N\left[\begin{pmatrix}\mu \\ 0\end{pmatrix}, \begin{pmatrix}\sigma^2/n & \rho\sigma/\sqrt{n} \\ \frac{\rho\sigma}{\sqrt{n}} & 1\end{pmatrix}\right]$$

$\text{corr} = \rho$

$\text{var}\,\hat{\mu} = \frac{\sigma^2}{n}$ $\quad$ $\text{var}\, Z = 1$

$\text{cor}(\hat{\mu}, Z) =$

# A model for selection bias

- study with $n$ individuals leads to estimate $\hat\mu \dot\sim N(\mu, \sigma^2/n)$
- study is published ($R = 1$), if $Z > 0$ — some measure of randomness in publication
- Suppose $\hat\mu$ and $Z$ are related according to the model — $Y \to \hat\mu, X \to n$, both pot. missing

$$\hat\mu = \mu + \sigma n^{-1/2} U_1, \quad Z = \gamma_0 + \gamma_1 n^{1/2} + U_2, \quad \mathrm{cor}(U_1, U_2) = \rho \geq 0$$

$U_1 =$

- $\mathrm{pr}(R = 1) = \mathrm{pr}(Z > 0) = \Phi(\gamma_0 + \gamma_1 n^{1/2})$
- $\mathrm{pr}(R = 1 \mid \hat\mu) = \mathrm{pr}(Z > 0 \mid \hat\mu) = \Phi\left\{\dfrac{\gamma_0 + \gamma_1 n^{1/2} + \rho n^{1/2}(\hat\mu - \mu)/\sigma}{(1-\rho^2)^{1/2}}\right\}$ — MV Normal
- non-ignorable non-response — unless $\rho = 0$
- estimate of $\mu$ is biased: — small $\gamma_1$

$$E(\hat\mu \mid R = 1) = \mu + \rho\sigma n^{-1/2}\zeta(\gamma_0 + \gamma_1 n^{1/2})$$

$$\doteq \mu + \frac{\rho\sigma}{\sqrt{n}} \, \phi(\gamma_0 + \gamma_1\sqrt{n})$$

$\zeta = \phi/\Phi$

- estimate of $\mu$ is biased: small $\gamma_1$

$$E(\hat{\mu} \mid R = 1) = \mu + \rho\sigma n^{-1/2}\zeta(\gamma_0 + \gamma_1 n^{1/2})$$

$$\doteq \underline{\mu + \rho\sigma\gamma_1\zeta'(\gamma_0)} + \underline{\rho\sigma\zeta(\gamma_0)}n^{-1/2}$$

$$\zeta(\gamma_0 + \sqrt{n}\gamma_1) = \zeta(\gamma_0) + \sqrt{n}\gamma_1\zeta'(\gamma_0) + o(\gamma_1)$$

- estimate of $\mu$ is biased: small $\gamma_1$

$$E(\hat{\mu} \mid R = 1) = \mu + \rho\sigma n^{-1/2}\zeta(\gamma_0 + \gamma_1 n^{1/2})$$
$$\doteq \mu + \rho\sigma\gamma_1\zeta'(\gamma_0) + \rho\sigma\zeta(\gamma_0)n^{-1/2}$$

- Suppose now we have $k$ published studies of the same treatment $\hat{\mu}_1, \dots, \hat{\mu}_k$,
- assume $\hat{\mu}_j \dot{\sim} N(\mu, \sigma^2/n_j)$ same mean, variance depends on study size
-

$$f(\hat{\mu}_j \mid R_j = 1; \theta) = \frac{f(\hat{\mu}_j; \theta)\mathrm{pr}(Z_j > 0 \mid \hat{\mu}_j; \theta)}{\mathrm{pr}(Z_j > 0)}$$

- estimate of $\mu$ is biased: small $\gamma_1$

$$E(\hat\mu \mid R = 1) = \mu + \rho\sigma n^{-1/2}\zeta(\gamma_0 + \gamma_1 n^{1/2})$$
$$\doteq \mu + \rho\sigma\gamma_1\zeta'(\gamma_0) + \rho\sigma\zeta(\gamma_0)n^{-1/2}$$

- Suppose now we have $k$ published studies of the same treatment $\hat\mu_1, \dots, \hat\mu_k$,
- assume $\hat\mu_j \dot\sim N(\mu, \sigma^2/n_j)$ same mean, variance depends on study size
-

$$f(\hat\mu_j \mid R_j = 1; \theta) = \frac{f(\hat\mu_j; \theta)\mathrm{pr}(Z_j > 0 \mid \hat\mu_j; \theta)}{\mathrm{pr}(Z_j > 0)}$$

- log-likelihood function $\theta =$

$\theta = (\mu, \rho, \sigma^2)$

$$\ell(\theta; \hat\mu) = -\sum_{j=1}^{k}\left\{\frac{1}{2}\log\sigma^2 + \frac{n_j}{2\sigma^2}(\hat\mu_j - \mu)^2 + \log\Phi(a_j) - \log\Phi(b_j)\right\}$$

$$a_j = \gamma_0 + \gamma_1 n_j^{1/2}, b_j = \{a_j + \rho n_j^{1/2}(\hat\mu_j - \mu)/\sigma\}(1 - \rho^2)^{-1/2}$$

- log-likelihood function

$$a_j = \gamma_0 + \gamma^1 n_j^{1/2},\; b_j = \{a_j + \rho n_j^{1/2}(\hat\mu_j - \mu)/\sigma\}(1-\rho^2)^{-1/2}$$

$$\ell(\theta;\hat\mu) = -\sum_{j=1}^{k}\left\{\frac{1}{2}\log\sigma^2 + \frac{n_j}{2\sigma^2}(\hat\mu_j - \mu)^2 + \log\Phi(a_j) - \log\Phi(b_j)\right\}$$

- if we set $\rho = 0$, $\quad \hat\mu = \dfrac{\sum n_j\hat\mu_j}{\sum n_j} \;\dot\sim\; N\left(0, \dfrac{\sigma^2}{\Sigma n_j}\right)$ no publication bias

208 5 · *Models*

| Trial | Magnesium $r/m$ | Control $r/m$ | $n$ | $\hat\mu$ | $(v/n)^{1/2}$ |
|---|---|---|---|---|---|
| 1 | 1/25 | 3/23 | 48 | 1.18 | 1.05 |
| 2 | 1/40 | 2/36 | 76 | 0.80 | 0.83 |
| 3 | 2/48 | 2/46 | 94 | 0.04 | 0.75 |
| 4 | 1/50 | 9/53 | 103 | 2.14 | 0.72 |
| 5 | 4/56 | 14/56 | 112 | 1.25 | 0.69 |
| 6 | 3/66 | 6/66 | 132 | 0.69 | 0.63 |
| 7 | 2/92 | 7/93 | 185 | 1.24 | 0.53 |
| 8 | 27/135 | 43/135 | 270 | 0.47 | 0.44 |
| 9 | 10/160 | 8/156 | 316 | −0.20 | 0.41 |
| 10 | 90/1159 | 118/1157 | 2316 | 0.27 | 0.15 |
| Meta-analysis | | | 3652 | 0.41 | 0.11 |
| ISIS-4 | 2216/29011 | 2103/29039 | 58050 | −0.05 | 0.03 |

**Table 5.9** Data from 11 clinical trials to compare magnesium treatment for heart attacks with control, with $n$ patients randomly allocated to treatment and control; there are $r$ deaths out of $m$ patients in each group (Copas, 1999). The estimated log treatment effect $\hat\mu$ will be positive if treatment is effective; $(v/n)^{1/2}$ is its standard error. The huge ISIS-4 trial is not included in the meta-analysis.

$\exp(\hat\mu) = 1.51$,
95% CI $(1.22, 1.86)$

| Trial | Magnesium $r/m$ | Control $r/m$ | $n$ | $\widehat{\mu}$ | $(v/n)^{1/2}$ |
|---|---|---|---|---|---|
| 1 | 1/25 | 3/23 | 48 | 1.18 | 1.05 |
| 2 | 1/40 | 2/36 | 76 | 0.80 | 0.83 |
| 3 | 2/48 | 2/46 | 94 | 0.04 | 0.75 |
| 4 | 1/50 | 9/53 | 103 | 2.14 | 0.72 |
| 5 | 4/56 | 14/56 | 112 | 1.25 | 0.69 |
| 6 | 3/66 | 6/66 | 132 | 0.69 | 0.63 |
| 7 | 2/92 | 7/93 | 185 | 1.24 | 0.53 |
| 8 | 27/135 | 43/135 | 270 | 0.47 | 0.44 |
| 9 | 10/160 | 8/156 | 316 | −0.20 | 0.41 |
| 10 | 90/1159 | 118/1157 | 2316 | 0.27 | 0.15 |
| Meta-analysis | | | 3652 | 0.41 | 0.11 |
| ISIS-4 | 2216/29011 | 2103/29039 | 58050 | −0.05 | 0.03 |

**Table 5.9** Data from 11 clinical trials to compare magnesium treatment for heart attacks with control, with $n$ patients randomly allocated to treatment and control; there are $r$ deaths out of $m$ patients in each group (Copas, 1999). The estimated log treatment effect $\widehat{\mu}$ will be positive if treatment is effective; $(v/n)^{1/2}$ is its standard error. The huge ISIS-4 trial is not included in the meta-analysis.

$\hat{\mu} = 0.41$ $e^{\hat{\mu}} = 1.$

$\hat\lambda$ overall death rate

$m_1 \doteq m_2$, $r_j$ small

$$\hat{\mu}_j = \underline{\log(r_{2j}/m_{2j})} - \underline{\log(r_{1j}/m_{1j})}, \quad \sigma^2 \doteq 41.24$$

$\mathrm{var}\,\hat{\mu}_j \approx \dfrac{4}{\hat{\lambda}(m_{1j}+m_{2j})}$

*5.5 · Missing Data*

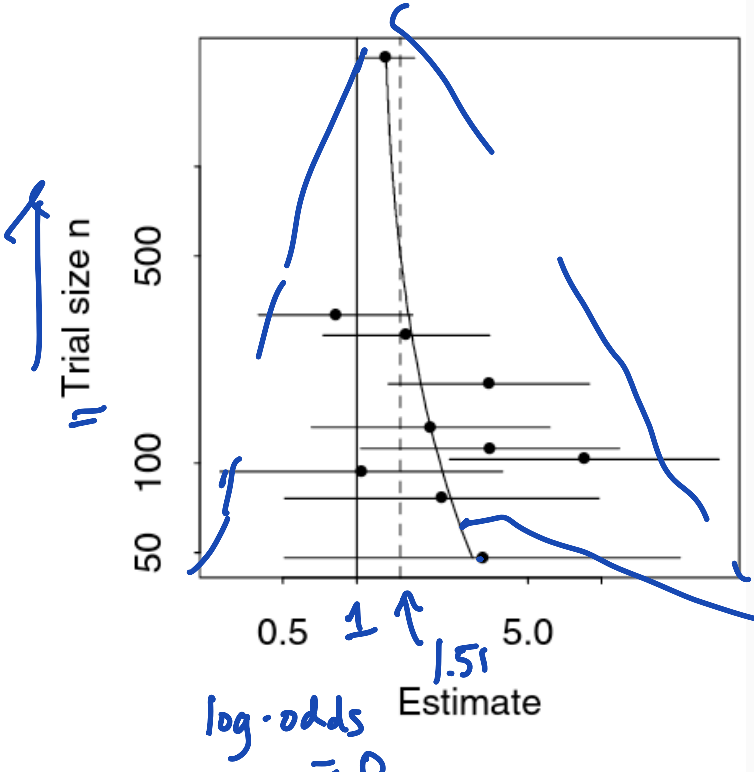
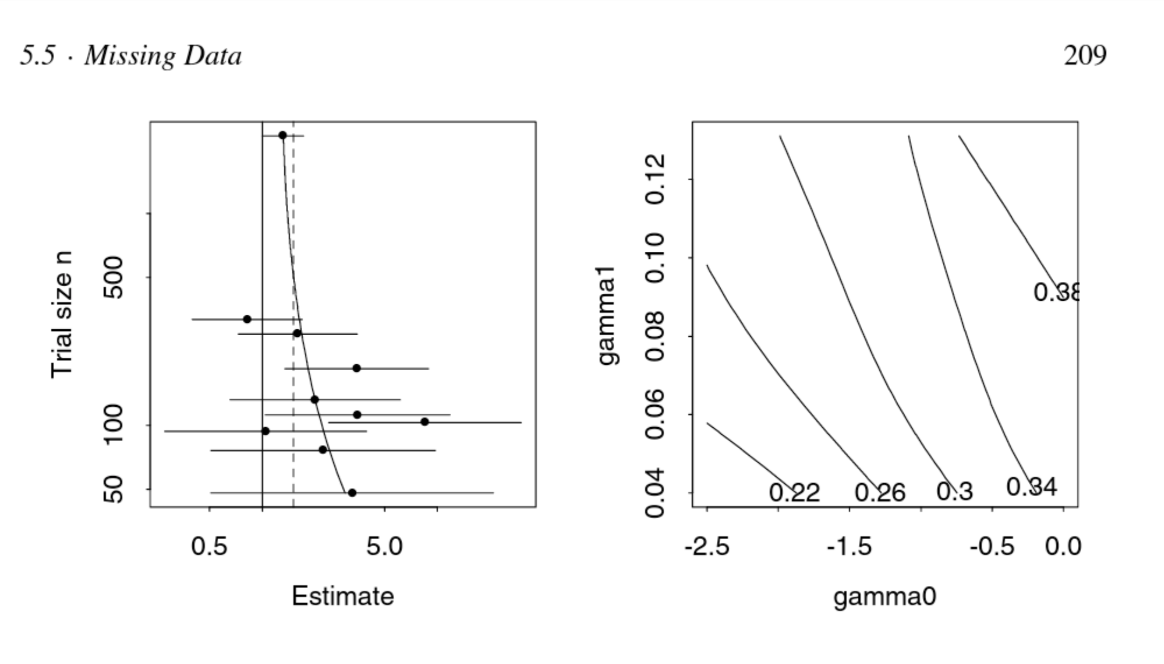



**Figure 5.13** Likelihood analysis of magnesium data. Left: funnel plot showing variation of $\widehat{\mu}$ with trial size $n$, with 95% confidence interval for $\mu$ based on each trial. The vertical dotted line is the combined estimate of $\mu$ from the ten small trials, ignoring the possibility of publication bias; the vertical solid line shows no treatment effect. The solid line is the estimated conditional mean (5.33). Right: contours of $\widehat{\mu}$ as a function of $\gamma_0$ and $\gamma_1$.

- smaller studies have wider confidence intervals
- seem to be missing small, negative, studies
- simple weighted average is positive (dashed line)
- estimate of average conditonal on publication favours smaller studies

$E(\hat{\mu} \mid R = 1)$




back-of the envelope calculation suggests $\hat{\rho} = 0.5$ and $\hat{\mu} = 0.27 \pm 0.12$

$$\exp(\hat{\mu}) = 1.31, \quad 95\% \text{ CI } (1.03, 1.66)$$

Large RCT (ISIS-4) found no benefit

preview

BMJ

*BMJ* 2011;342:d4002 doi: 10.1136/bmj.d4002 Page 1 of 8

RESEARCH METHODS & REPORTING

## Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials

Funnel plots, and tests for funnel plot asymmetry, have been widely used to examine bias in the results of meta-analyses. Funnel plot asymmetry should not be equated with publication bias, because it has a number of other possible causes. This article describes how to interpret funnel plot asymmetry, recommends appropriate tests, and explains the implications for choice of meta-analysis model

Jonathan A C Sterne *professor*[1], Alex J Sutton *professor*[2], John P A Ioannidis *professor and director*[3], Norma Terrin *associate professor*[4], David R Jones *professor*[2], Joseph Lau *professor*[4], James Carpenter *reader*[5], Gerta Rücker *research assistant*[6], Roger M Harbord *research associate*[1], Christopher H Schmid *professor*[4], Jennifer Tetzlaff *research coordinator*[7], Jonathan J Deeks *professor*[8], Jaime Peters *research fellow*[9], Petra Macaskill *associate professor*[10], Guido Schwarzer *research assistant*[6], Sue Duval *assistant professor*[11], Douglas G Altman *professor*[12], David Moher *senior scientist*[7], Julian P T Higgins *senior statistician*[13]

[1]School of Social and Community Medicine, University of Bristol, Bristol BS8 2PS, UK; [2]Department of Health Sciences, University of Leicester, Leicester, UK; [3]Stanford Prevention Research Center, Stanford University School of Medicine, Stanford, CA, USA; [4]Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA, USA; [5]Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London, UK; [6]Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Germany; [7]Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada; [8]School of Health and Population Sciences, University of Birmingham, Birmingham, UK; [9]Peninsula Medical School, University of Exeter, Exeter, UK; [10]School of Public Health, University of Sydney, NSW, Australia; [11]University of Minnesota School of Public Health, Minneapolis, MN, USA; [12]Centre for Statistics in Medicine, University of Oxford, Oxford, UK; [13]MRC Biostatistics

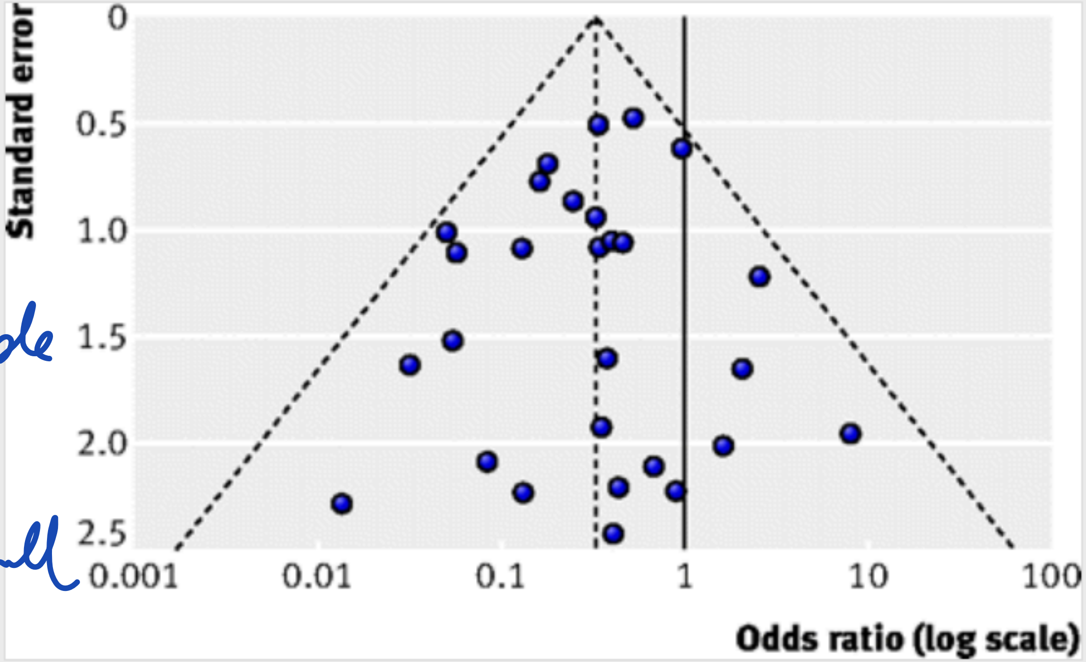
# More on funnel plots

Sterne et al. 2011

large

sample size

small

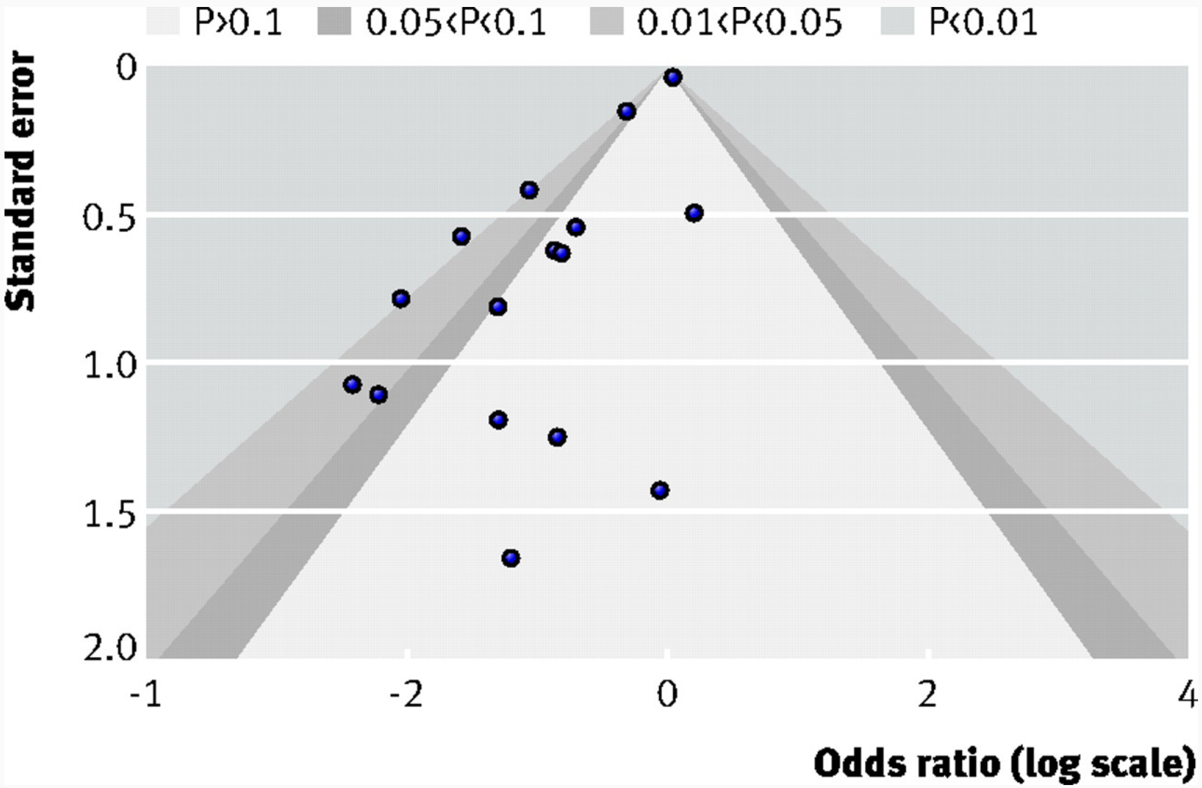



Centered at aggregate effect estimate, not null value.

**Fig 1** Example of symmetrical funnel plot. The outer dashed lines indicate the triangular region within which 95% of studies are expected to lie in the absence of both biases and heterogeneity (fixed effect summary log odds ratio±1.96×standard error of summary log odds ratio). The solid vertical line corresponds to no intervention effect

# … More on funnel plots

Sterne et al. 2011

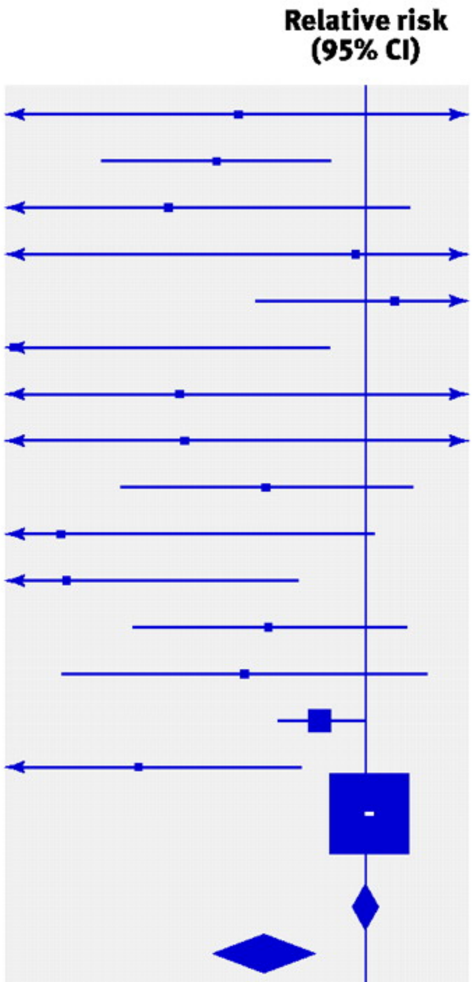



"Beneficial effects on mortality, found in a meta-analysis of small studies, were subsequently contradicted when the very large ISIS-4 study found no evidence of benefit.33 A contour enhanced funnel plot (fig 4) gives a clear visual impression of asymmetry"

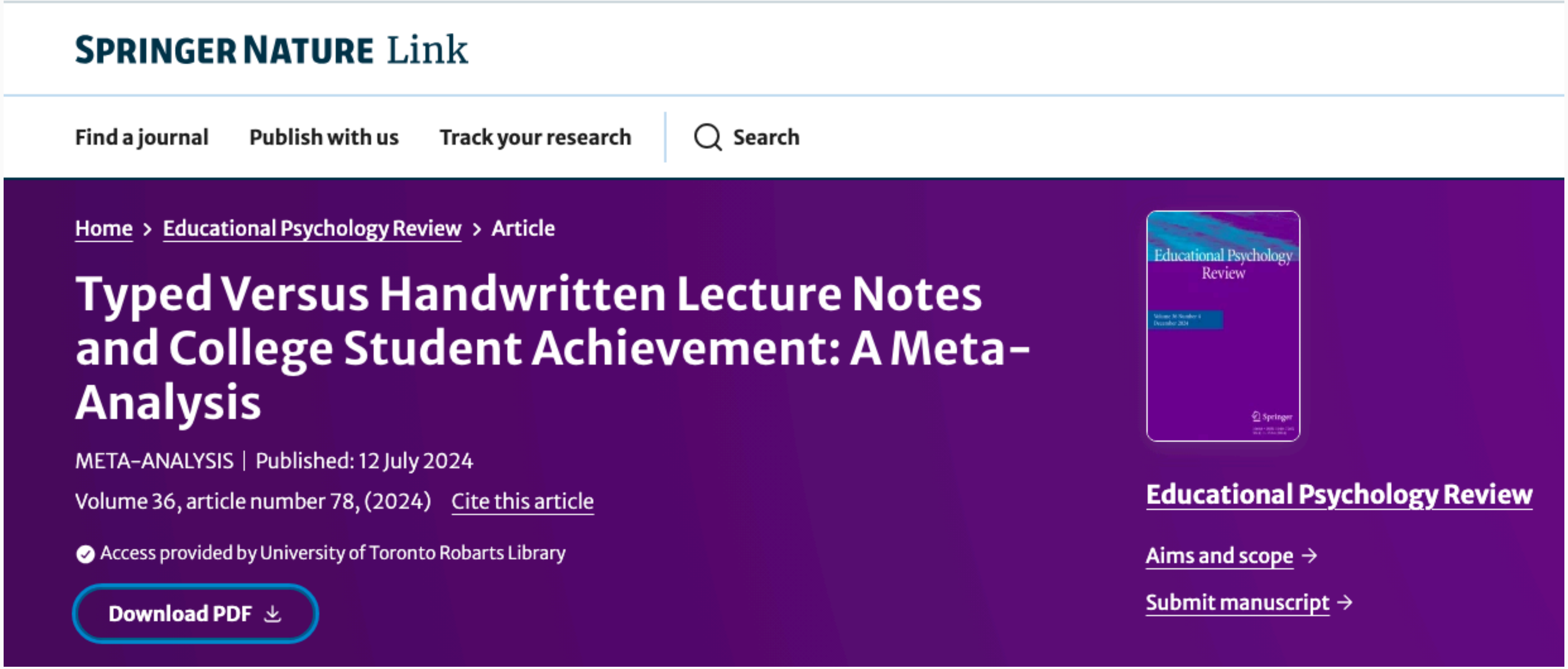
This is a funnel plot for the magnesium/heart meta-analysis, it has slightly different numbers than Davison's book, and a couple of extra studies, but essentially the same conclusions

# Forest plots






| Trial | Events: Treatment | Events: Control | Weight (%) | Relative risk (95% CI) |
|---|---|---|---|---|
| Morton 1984 | 1/40 | 2/36 | 0.09 | 0.45 (0.04 to 4.76) |
| Rasmussen 1986 | 9/135 | 23/135 | 0.98 | 0.39 (0.19 to 0.81) |
| Smith 1986 | 2/200 | 7/200 | 0.30 | 0.29 (0.06 to 1.36) |
| Abraham 1987 | 1/48 | 1/46 | 0.04 | 0.96 (0.06 to 14.87) |
| Feldstedt 1988 | 10/150 | 8/148 | 0.34 | 1.23 (0.50 to 3.04) |
| Shechter 1989 | 1/59 | 9/56 | 0.39 | 0.11 (0.01 to 0.81) |
| Ceremuzynski 1989 | 1/25 | 3/23 | 0.13 | 0.31 (0.03 to 2.74) |
| Bertschat 1989 | 0/22 | 1/21 | 0.07 | 0.32 (0.01 to 7.42) |
| Singh 1990 | 6/76 | 11/75 | 0.47 | 0.54 (0.21 to 1.38) |
| Pereira 1990 | 1/27 | 7/27 | 0.30 | 0.14 (0.02 to 1.08) |
| Shechter 1 1991 | 2/89 | 12/80 | 0.54 | 0.15 (0.03 to 0.65) |
| Golf 1991 | 5/23 | 13/33 | 0.46 | 0.55 (0.23 to 1.33) |
| Thogersen 1991 | 4/130 | 8/122 | 0.35 | 0.47 (0.14 to 1.52) |
| LIMIT-2 1992 | 90/1159 | 118/1157 | 5.04 | 0.76 (0.59 to 0.99) |
| Shechter 2 1995 | 4/107 | 17/108 | 0.72 | 0.24 (0.08 to 0.68) |
| ISIS-4 1995 | 2216/29 011 | 2103/29 039 | 89.76 | 1.05 (1.00 to 1.12) |
| Fixed-effect (M-H) estimate: $I^2$=67%, P=0.000 | 2353/31 301 | 2343/31 306 | 100.0 | 1.01 (0.95 to 1.06) |
| Random-effects (D+L) estimate | | | | 0.53 (0.38 to 0.75) |


SPRINGER NATURE Link

Find a journal Publish with us Track your research Search

Home > Educational Psychology Review > Article

# Typed Versus Handwritten Lecture Notes and College Student Achievement: A Meta-Analysis

META-ANALYSIS | Published: 12 July 2024

Volume 36, article number 78, (2024) Cite this article

Access provided by University of Toronto Robarts Library

Download PDF

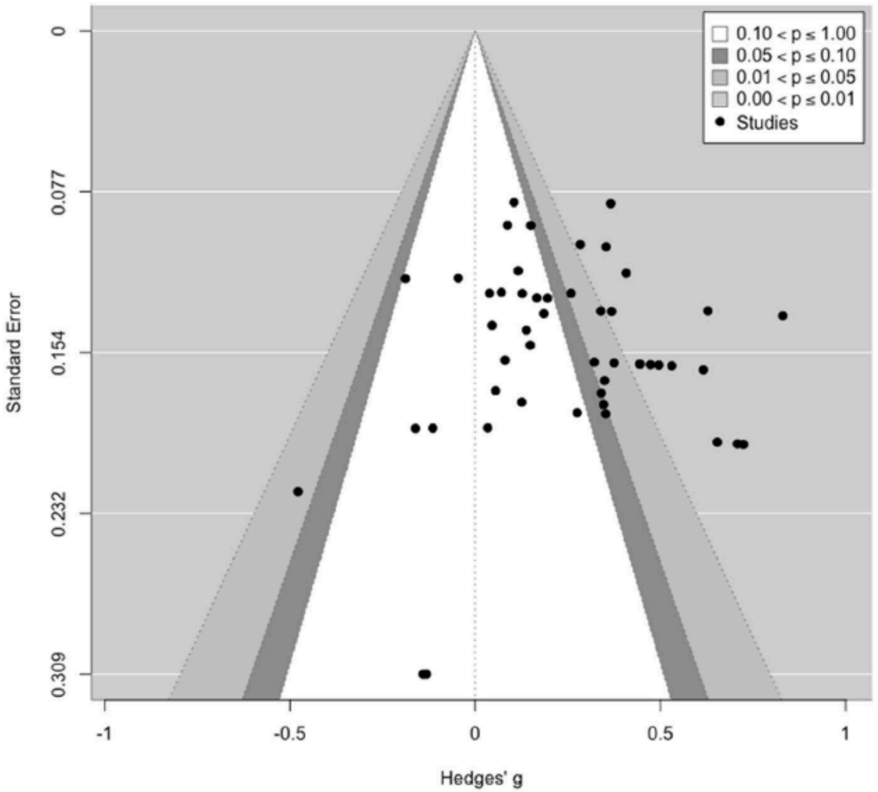
Educational Psychology Review

Aims and scope →

Submit manuscript →

# Meta-analysis

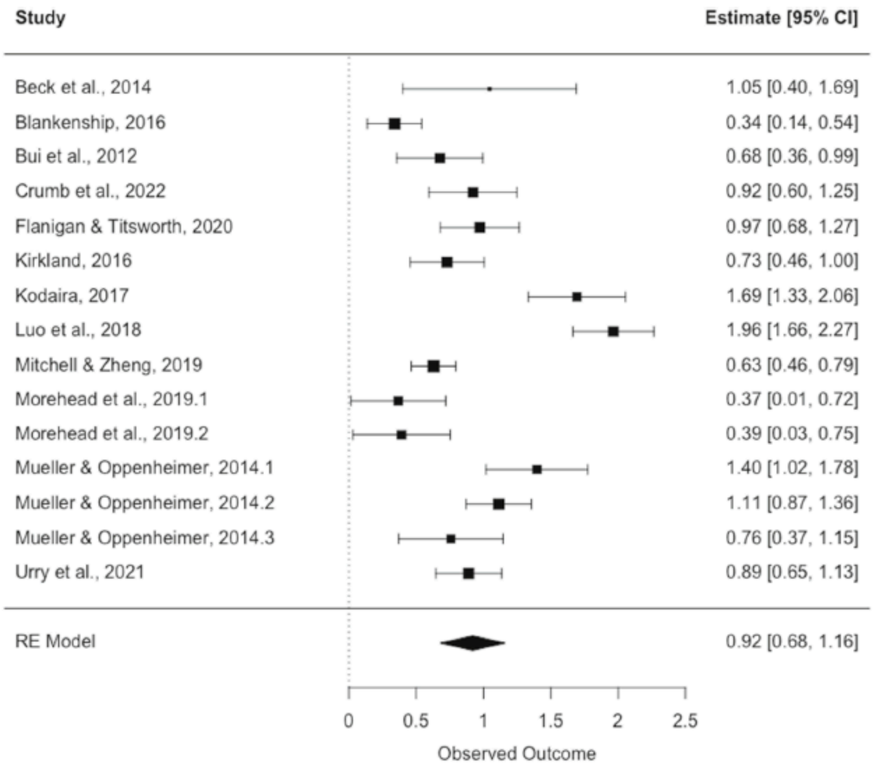





***Note.*** A positive standardized effect size indicates that students who handwrote their notes had higher achievement than those who typed their notes.

**Fig. 1** Funnel plot depicting Hedges' *g* and precision in assessing achievement: handwritten vs. typed






***Note.*** A positive standardized effect size indicates that students who typed their notes recorded a larger number of words compared to those who handwrote their notes.

# Inference with missing data

- if MAR or MCAR, can use usual likelihood-based inference with observed information to estimate ~~variance~~ parameters
- if not, but the missing-ness pattern can be modelled, may be able to adjust estimates accordingly

pub bias

- adjustments will depend on the missing-ness model being approx'ly correct
- there is a large literature on re-weighting standard estimators to accommodate missing-ness

# … Inference with missing data

- what about missing values of covariates?
- use only complete cases – may result in substantial reduction in sample size
- imputation of missing values is a popular choice
- based on prediction of missing covariate value, given observed values of other units

MICE Example

| ID | Age_Original | Income_Original | Age_Imp1 | Income_Imp1 | Age_Imp2 | Income_Imp2 |
|---|---|---|---|---|---|---|
| 1 | 25 | 50000 | 25 | 50000 | 25 | 50000 |
| 2 | | 55000 | 25 | 55000 | 50 | 55000 |
| 3 | 35 | | 35 | 65000 | 35 | 55000 |
| 4 | 40 | 70000 | 40 | 70000 | 40 | 70000 |
| 5 | | 65000 | 25 | 65000 | 50 | 65000 |
| 6 | 50 | | 50 | 75000 | 50 | 65000 |
| 7 | 45 | 80000 | 45 | 80000 | 45 | 80000 |
| 8 | | 90000 | 35 | 90000 | 29 | 90000 |
| 9 | 38 | | 38 | 75000 | 38 | 70000 |
| 10 | 29 | 75000 | 29 | 75000 | 29 | 75000 |

# … Inference with missing data

Research

JAMA | Original Investigation

## A Digital Health Behavior Intervention to Prevent Childhood Obesity
### The Greenlight Plus Randomized Clinical Trial

William J. Heerman, MD, MPH; Russell L. Rothman, MD, MPP; Lee M. Sanders, MD, MPH;
Jonathan S. Schildcrout, PhD; Kori B. Flower, MD, MS, MPH; Alan M. Delamater, PhD;
Melissa C. Kay, PhD, MPH, MS, RD, CLC; Charles T. Wood, MD, MPH; Rachel S. Gross, MD, MS; Aihua Bian, MPH;
Laura E. Adams, RD, MBA; Evan C. Sommer, BS, BA; H. Shonna Yin, MD, MSc; Eliana M. Perrin, MD, MPH;
and the Greenlight Investigators

**IMPORTANCE** Infant growth predicts long-term obesity and cardiovascular disease. Previous interventions designed to prevent obesity in the first 2 years of life have been largely unsuccessful. Obesity prevalence is high among traditional racial and ethnic minority groups.

**OBJECTIVE** To compare the effectiveness of adding a digital childhood obesity prevention intervention to health behavior counseling delivered by pediatric primary care clinicians.

**DESIGN, SETTING, AND PARTICIPANTS** Individually randomized, parallel-group trial conducted at 6 US medical centers and enrolling patients shortly after birth. To be eligible, parents spoke English or Spanish, and children were born after 34 weeks' gestational age. Study enrollment occurred between October 2019 and January 2022, with follow-up through January 2024.

**INTERVENTIONS** In the clinic-based health behavior counseling (clinic-only) group, pediatric clinicians used health literacy–informed booklets at well-child visits to promote healthy behaviors (n = 451). In the clinic + digital intervention group, families also received health literacy–informed, individually tailored, responsive text messages to support health behavior goals and a web-based dashboard (n = 449).

**MAIN OUTCOMES AND MEASURES** The primary outcome was child weight-for-length trajectory over 24 months. Secondary outcomes included weight-for-length *z* score, body mass index (BMI) *z* score, and the percentage of children with overweight or obesity.

**RESULTS** Of 900 randomized children, 86.3% had primary outcome data at the 24-month follow-up time point; 143 (15.9%) were Black, non-Hispanic; 405 (45.0%) were Hispanic; 185

Visual Abstract

Multimedia

Supplemental content

"Missing baseline variables were imputed 1000 times with chained equations" (p.4)

# A curious example

- data $(X_1, R_1, Y_1), \dots, (X_n, R_n, Y_n)$ i.i.d.
  1. $X_i \sim$ Uniform from $\{1, \dots, B\}$
  2. $R_i \sim \text{Bernoulli}(\xi_{X_i})$
  3. If $R_i = 1$, $Y_i \sim \text{Bernoulli}(\theta_{X_i})$
- $\theta = (\theta_1, \dots, \theta_B)$ unknown, $0 \le \theta_j \le 1$
- $\xi = (\xi_1, \dots, \xi_B)$ known, $0 < \delta \le \xi_j \le 1 - \delta < 1$

# A curious example

AoS Ex. 11.9

- data $(X_1, R_1, Y_1), \ldots, (X_n, R_n, Y_n)$ i.i.d.
  1. $X_i \sim$ Uniform from $\{1, \ldots, B\}$
  2. $R_i \sim$ Bernoulli$(\xi_{X_i})$
  3. If $R_i = 1$, $Y_i \sim$ Bernoulli$(\theta_{X_i})$
- $\theta = (\theta_1, \ldots, \theta_B)$ unknown, $0 \le \theta_j \le 1$
- $\xi = (\xi_1, \ldots, \xi_B)$ known, $0 < \delta \le \xi_j \le 1 - \delta < 1$
- parameter of interest $\psi = \underbrace{\mathrm{pr}(Y_i = 1)}$ $= \underbrace{\sum_{j=1}^{B} \mathrm{pr}(Y_i = 1 \mid X_i = j)\mathrm{pr}(X_i = j)}$ $= \frac{1}{B}\sum_j \theta_j$
- An unbiased estimator of $\psi$:

$$\hat{\psi} = \frac{1}{n}\sum_{i=1}^{n} \frac{R_i Y_i}{\xi_{X_i}}$$

- observed values are averaged, but weighted by probability of being observed
- Horvitz-Thompson estimator

IPW estimator

# … A curious example

- data $(X_1, R_1, Y_1), \dots, (X_n, R_n, Y_n)$ i.i.d.
  1. $X_i \sim$ Uniform from $\{1, \dots, B\}$
  2. $R_i \sim$ Bernoulli$(\xi_{X_i})$
  3. If $R_i = 1$, $Y_i \sim$ Bernoulli$(\theta_{X_i})$
- one term in likelihood function:

$$\underline{f(X_i)f(R_i \mid X_i)f(Y_i \mid X_i)^{R_i}} = \frac{1}{B}\underbrace{\xi_{X_i}^{R_i}(1-\xi_{X_i})^{1-R_i}}\underbrace{\theta_{X_i}^{Y_iR_i}(1-\theta_{X_i})^{(1-Y_i)R_i}}$$

# … A curious example

- data $(X_1, R_1, Y_1), \dots, (X_n, R_n, Y_n)$ i.i.d.
  1. $X_i \sim$ Uniform from $\{1, \dots, B\}$
  2. $R_i \sim$ Bernoulli$(\xi_{X_i})$
  3. If $R_i = 1$, $Y_i \sim$ Bernoulli$(\theta_{X_i})$

$\psi = \frac{1}{B}\sum \theta_j$

- one term in likelihood function:

$$f(X_i)f(R_i \mid X_i)f(Y_i \mid X_i)^{R_i} = \frac{1}{B}\xi_{X_i}^{R_i}(1-\xi_{X_i})^{1-R_i}\theta_{X_i}^{Y_iR_i}(1-\theta_{X_i})^{(1-Y_i)R_i}$$

- likelihood function: $L(\theta) \propto \prod_{i=1}^n \theta_{X_i}^{Y_iR_i}(1-\theta_{X_i})^{(1-Y_i)R_i} = \prod_{j=1}^B \theta_j^{n_j}(1-\theta_j)^{m_j}$
- $n_j = \#\{i : Y_i = 1, R_i = 1, X_i = j\}$, $\quad m_j = \#\{i : Y_i = 1, R_i = 0, X_i = j\}$ ← ntbc
  - ? 0 1
- most $n_j, m_j = 0$ ($B$ very large) $\implies$ mle of $\theta_j$ doesn't exist for many $j$
  $\implies \pi(\theta \mid \text{data}) \propto \pi(\theta)$

$$f(y;\vartheta) = \sum_{r=1}^{p} f(y_r;\vartheta)\pi_r, \quad 0 \le \pi_r \le 1, \Sigma\pi_r = 1$$

- missing data: $U_1,\ldots,U_p; U_r \sim Bernoulli(\pi_r)$ indexes sub-model
- complete-data log-likelihood function:

$$\log f(y,u;\theta) = \sum_{r=1}^{p} 1(U=r)\{\log\pi_r + \log f_r(y;\theta)\} \qquad = \log f(y;\cancel{\theta}) + \log f(u \mid y;\theta)$$

$\vartheta$ ?

ntbc

$\theta = (\vartheta, \pi)$

$E_{\theta'} \log f(u \mid y;\theta)$

$$f(y;\theta) = \sum_{r=1}^{p} f(y_r;\vartheta)\pi_r, \quad 0 \le \pi_r \le 1, \Sigma\pi_r = 1$$

- missing data: $U_1, \dots, U_p$; $U_r \sim Bernoulli(\pi_r)$ indexes sub-model
- complete-data log-likelihood function:

$$\log f(y,u;\theta) = \sum_{r=1}^{p} 1(U=r)\{\log\pi_r + \log f_r(y;\theta)\} \qquad = \log f(y;\theta) + \log f(u \mid y;\theta)$$

- maximize $Q(\theta,\theta') = \log f(y;\theta) + \mathrm{E}_{\theta'} \log f(u \mid y;\theta)$ $\leftarrow \theta^{(j)} \rightarrow Q(\theta,\theta^{(j)})$
- conditional distribution

$\rightarrow \mathrm{E}_{\hat\theta^{(j)}}\{\log f(u \mid y;\theta)\}$

$$\mathrm{pr}(U=r \mid Y=y;\theta') = \frac{\pi'_r f_r(y;\theta')}{\sum_{s=1}^{p}\pi_s f_s(y;\theta')} = w_r(y;\theta')$$

M

est'g

## The EM Algorithm

(0) Pick a starting value $\theta^0$. Now for $j = 1, 2, \ldots,$ repeat steps 1 and 2 below:

(1) (The E-step): Calculate

$$J(\theta|\theta^j) = \mathbb{E}_{\theta^j}\left(\log \frac{f(Y^n, Z^n; \theta)}{f(Y^n, Z^n; \theta^j)} \,\Bigg|\, \underline{Y^n = y^n}\right). \qquad E$$

$Q(\theta, \theta^j) =$

The expectation is over the missing data $Z^n$ treating $\theta^j$ and the observed data $\underline{Y^n}$ as fixed.

(2) Find $\theta^{j+1}$ to maximize $J(\theta|\theta^j)$. $\leftarrow$ M

$$\prod_{i=1}$$

Maximizing this function over the five parameters is hard. Imaging that we were given extra information telling us which of the two normals every observation came from. These "complete" data are of the form $(Y_1, Z_1), \ldots, (Y_n, Z_n)$, where $Z_i = 0$ represents the first normal and $Z_i = 1$ represents the second. Note that $\mathbb{P}(Z_i = 1) = p$. We shall soon see that the likelihood for the complete data $(Y_1, Z_1), \ldots, (Y_n, Z_n)$ is much simpler than the likelihood for the observed data $Y_1, \ldots, Y_n$. ■

Now we describe the EM algorithm.

The EM Algorithm

(0) Pick a starting value $\theta^0$. Now for $j = 1, 2, \ldots,$ repeat steps 1 and 2 below:

(1) (The E-step): Calculate

$$J(\theta|\theta^j) = \mathbb{E}_{\theta^j}\left(\log \frac{f(Y^n, Z^n; \theta)}{f(Y^n, Z^n; \theta^j)} \,\middle|\, Y^n = y^n\right).$$

The expectation is over the missing data $Z^n$ treating $\theta^i$ and the observed data $Y^n$ as fixed.

(2) Find $\theta^{j+1}$ to maximize $J(\theta|\theta^j)$.

We now show that the EM algorithm always increases the likelihood, that is, $\mathcal{L}(\theta^{j+1}) \geq \mathcal{L}(\theta^j)$. Note that

$$\begin{aligned} J(\theta^{j+1}|\theta^j) &= \mathbb{E}_{\theta^j}\left(\log \frac{f(Y^n, Z^n; \theta^{j+1})}{f(Y^n, Z^n; \theta^j)} \,\middle|\, Y^n = y^n\right) \\ &= \log \frac{f(y^n; \theta^{j+1})}{f(y^n; \theta^j)} + \mathbb{E}_{\theta^j}\left(\log \frac{f(Z^n|Y^n; \theta^{j+1})}{f(Z^n|Y^n; \theta^j)} \,\middle|\, Y^n = y^n\right) \end{aligned}$$