

Bias-corrected false discovery rates

International Conference on
Mathematical and Computational Medicine

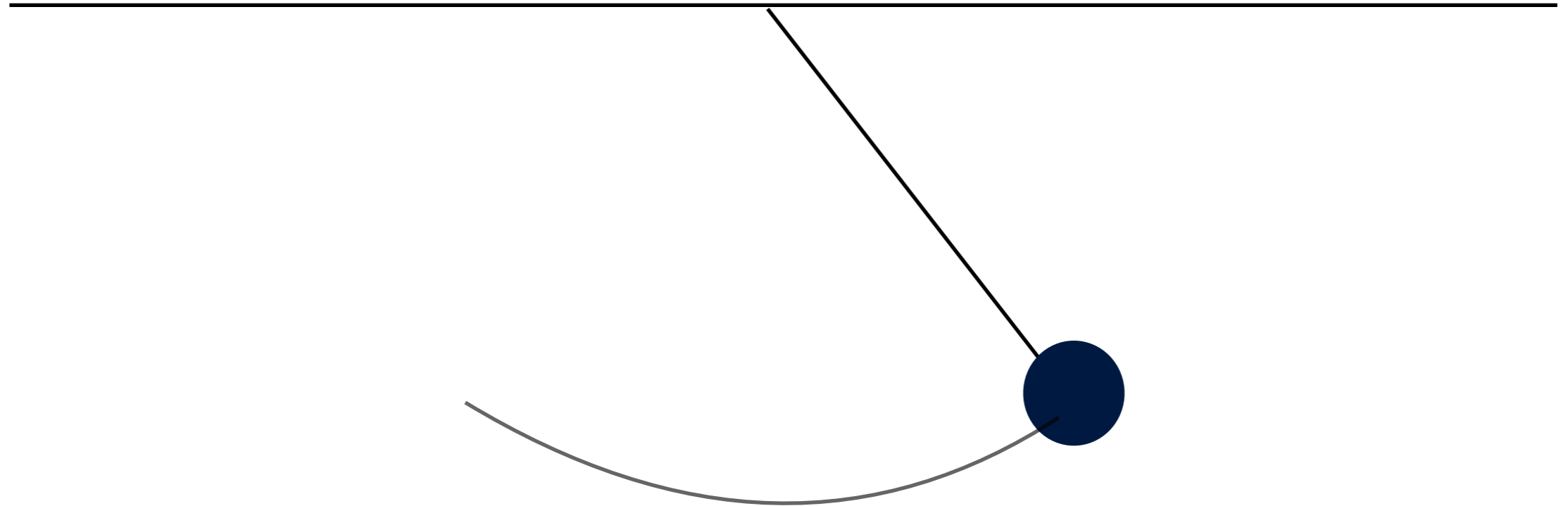
Cancun, Mexico

6 December 2018

David R. Bickel
University of Ottawa



Swung too far?



Family-wise error rates

Too many false negatives

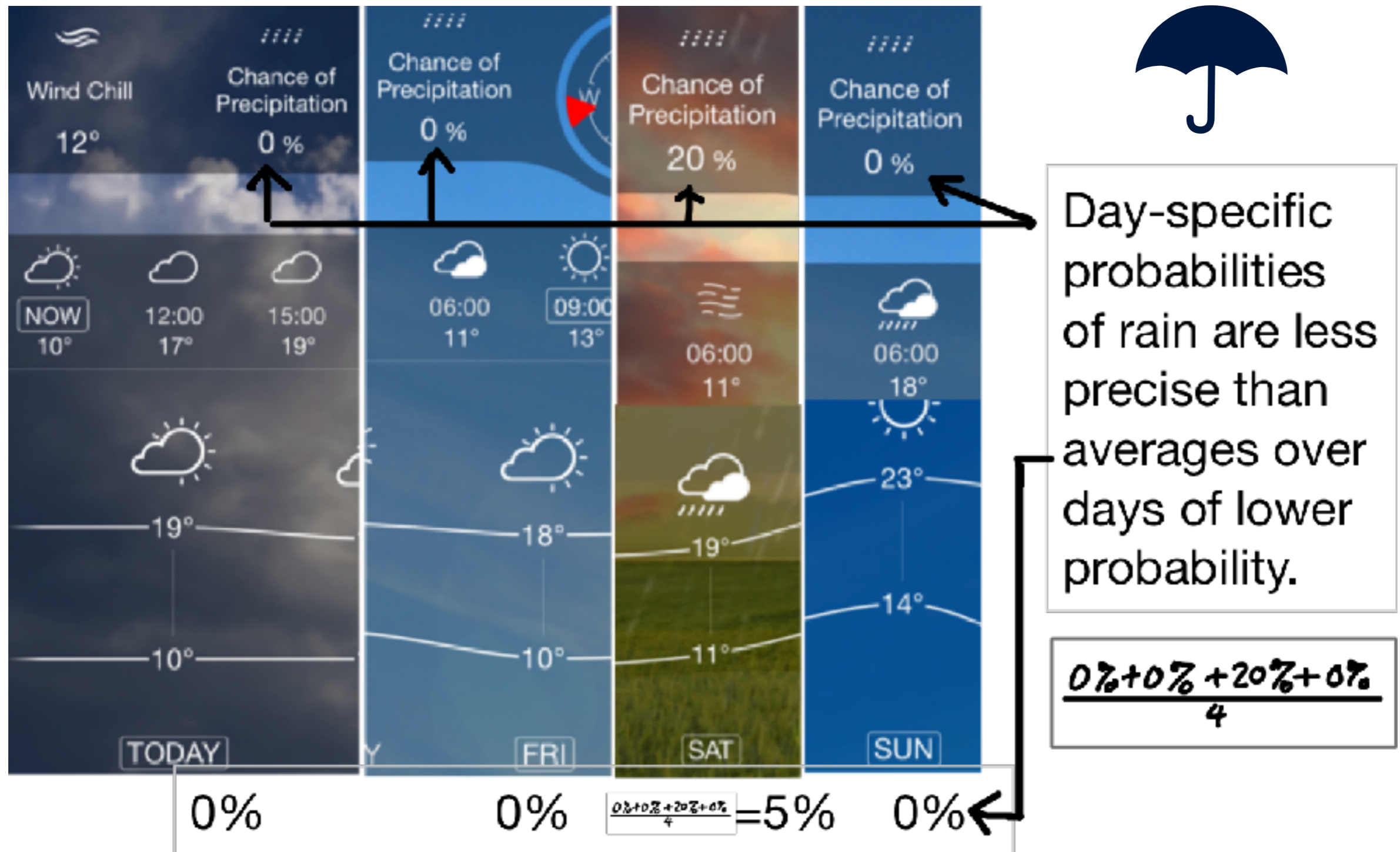
False discovery rates

Too many false positives

Local false discovery rates

Large variance

Probabilities of rain



FDRs & local FDRs

$$\begin{aligned}
 \text{FDR}(0.01) &= \frac{\text{average number of false discoveries at 0.01 significance}}{\text{average number of discoveries at 0.01 significance}} \\
 &= \frac{\text{average number of p-values} < 0.01 \text{ for equivalently expressed genes}}{\text{average number of p-values} < 0.01}
 \end{aligned}$$

$$\begin{aligned}
 \widehat{\text{FDR}}(\alpha) &= \frac{\text{estimated average number of false discoveries}}{\text{estimated average number of discoveries}} \\
 &= \frac{\text{estimated average number of false discoveries}}{\text{number of discoveries}} \\
 &= \begin{cases} \frac{\alpha d}{\#(p(x_i) \leq \alpha)} & \text{if } \frac{\alpha d}{\#(p(x_i) \leq \alpha)} < 1 \\ 1 & \text{if } \frac{\alpha d}{\#(p(x_i) \leq \alpha)} > 1. \end{cases}
 \end{aligned}$$

$$\text{FDR}(\alpha) \approx \frac{\text{LFDR}(p_1) + \text{LFDR}(p_2) + \dots + \text{LFDR}(p_{\#(p(x_i) \leq \alpha)})}{\#(p(x_i) \leq \alpha)}$$

Interpreting the false discovery rate

- If a discovery of differential expression is made whenever the p-value is less than 0.05, then the false discovery rate is the average of all LFDRs corresponding to discoveries

$$\text{FDR}(0.05) = \text{mean}(\text{LFDR}(p(x_i)) | p(x_i) < 0.05)$$

- false discovery rate = probability that randomly selected discovery is false

$$\text{FDR}(0.05) = P(A_i = 0 | p(x_i) < 0.05)$$

Local false discovery rates

- local false discovery rate (LFDR) = posterior probability of equivalent expression: $\text{LFDR}(0.00832) = P(A_i = 0 | p(X_i) = 0.00832)$
- evidence of differential expression = likelihood ratio:

$$\frac{L_i(1)}{L_i(0)} \approx \frac{P(p(X_i) \approx 0.00832 | A_i = 1)}{P(p(X_i) \approx 0.00832 | A_i = 0)}$$

- posterior odds that gene i of p-value 0.00832 is differentially expressed:

$$\frac{1 - \text{LFDR}(0.00832)}{\text{LFDR}(0.00832)} = \frac{P(A_i = 1 | p(X_i) = 0.00832)}{P(A_i = 0 | p(X_i) = 0.00832)} = \frac{P(A_i = 1)}{P(A_i = 0)} \times \frac{L_i(1)}{L_i(0)}$$

Achieved FDR

$$\begin{aligned}
 \text{FDR}(0.01) &= \frac{\text{average number of false discoveries at 0.01 significance}}{\text{average number of discoveries at 0.01 significance}} \\
 &= \frac{\text{average number of p-values} < 0.01 \text{ for equivalently expressed genes}}{\text{average number of p-values} < 0.01}
 \end{aligned}$$

$$\widehat{\text{FDR}}(\alpha) = \begin{cases} \frac{\alpha d}{\#(p(x_i) \leq \alpha)} & \text{if } \frac{\alpha d}{\#(p(x_i) \leq \alpha)} < 1 \\ 1 & \text{if } \frac{\alpha d}{\#(p(x_i) \leq \alpha)} > 1 \end{cases}$$

$$\widehat{\text{FDR}}(p(x_j)) = \begin{cases} \frac{p(x_j) d}{\#(p(x_i) \leq p(x_j))} & \text{if } \frac{p(x_j) d}{\#(p(x_i) \leq p(x_j))} < 1 \\ 1 & \text{if } \frac{p(x_j) d}{\#(p(x_i) \leq p(x_j))} > 1 \end{cases}$$

Bias in false discovery rates

$$\begin{aligned}\text{FDR}(p(x_j)) &\approx \frac{\text{LFDR}(p_1) + \text{LFDR}(p_2) + \cdots + \text{LFDR}(p_{\#(p(x_i) \leq p(x_j))})}{\#(p(x_i) \leq p(x_j))} \\ &= \frac{\text{LFDR}(p_1) + \text{LFDR}(p_2) + \cdots + \text{LFDR}(p_j)}{\#(p(x_i) \leq p(x_j))}\end{aligned}$$

$$p(x_1) < p(x_2) < \cdots < p(x_d)$$

$$\text{LFDR}(p(x_1)) < \text{LFDR}(p(x_2)) < \cdots < \text{LFDR}(p(x_d))$$

$$\text{LFDR}(p_j) > \frac{\text{LFDR}(p_1) + \text{LFDR}(p_2) + \cdots + \text{LFDR}(p_j)}{\#(p(x_i) \leq p(x_j))} = \text{FDR}(p(x_j))$$

$$\text{FDR}(p(x_j)) < \text{LFDR}(p_j)$$

Correcting the bias

$$p(x_1) < p(x_2) < \dots < p(x_d) \quad \text{LFDR}(p(x_1)) < \text{LFDR}(p(x_2)) < \dots < \text{LFDR}(p(x_d))$$

$$\widehat{\text{FDR}}(p(x_j)) = \begin{cases} \frac{p(x_j)d}{\#(p(x_i) \leq p(x_j))} & \text{if } \frac{p(x_j)d}{\#(p(x_i) \leq p(x_j))} < 1 \\ 1 & \text{if } \frac{p(x_j)d}{\#(p(x_i) \leq p(x_j))} > 1 \end{cases}$$

$$\text{CFDR}(x_j) = \begin{cases} \left(\frac{1}{1-1+1}\right) \widehat{\text{FDR}}(p(x_j)) & \text{if } j = 1 \\ \left(\frac{1}{2-1+1} + \frac{1}{2-2+1}\right) \widehat{\text{FDR}}(p(x_j)) & \text{if } j = 2 \\ \left(\frac{1}{3-1+1} + \frac{1}{3-2+1} + \frac{1}{3-3+1}\right) \widehat{\text{FDR}}(p(x_j)) & \text{if } j = 3 \\ \left(\frac{1}{j-1+1} + \frac{1}{j-2+1} + \dots + \frac{1}{j-j+1}\right) \widehat{\text{FDR}}(p(x_j)) & \text{if } j \geq 4 \end{cases}$$

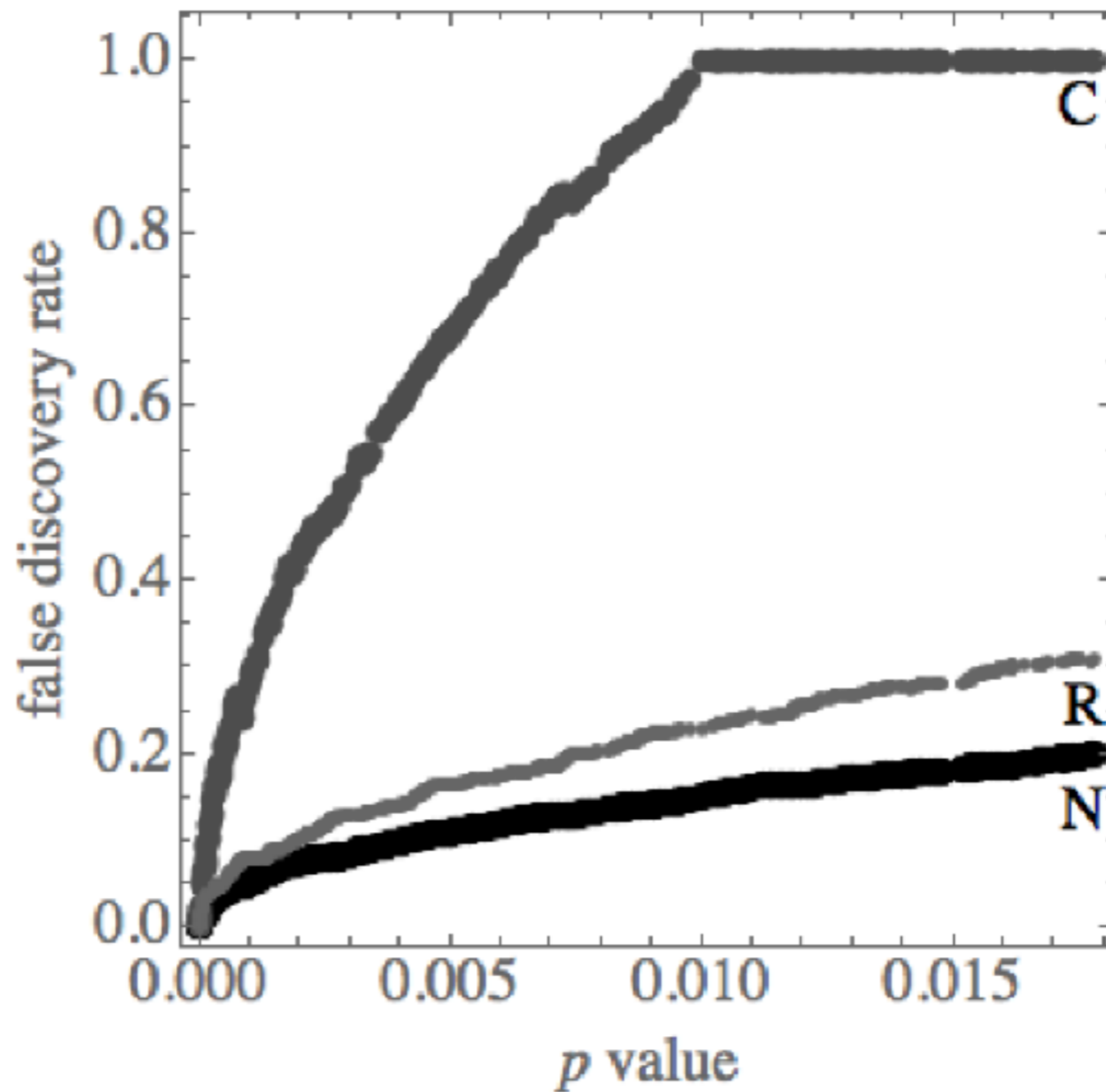
Re-ranked false discovery rate

$$p(x_1) < p(x_2) < \cdots < p(x_d) \quad \text{LFDR}(p(x_1)) < \text{LFDR}(p(x_2)) < \cdots < \text{LFDR}(p(x_d))$$

$$\widehat{\text{FDR}}(p(x_j)) = \begin{cases} \frac{p(x_j)d}{\#(p(x_i) \leq p(x_j))} & \text{if } \frac{p(x_j)d}{\#(p(x_i) \leq p(x_j))} < 1 \\ 1 & \text{if } \frac{p(x_j)d}{\#(p(x_i) \leq p(x_j))} > 1 \end{cases}$$

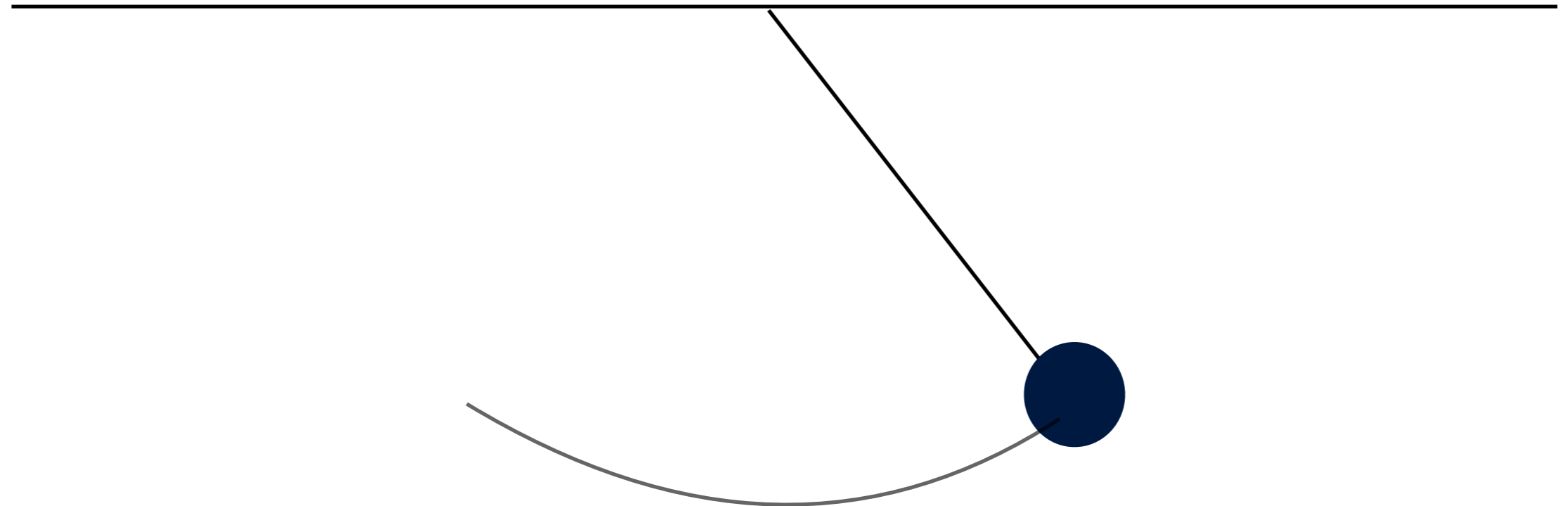
$$\text{RFDR}(x_j) = \begin{cases} \widehat{\text{FDR}}(p(x_{[1.6j]})) & \text{if } [1.6j] \leq d \\ 1 & \text{if } [1.6j] > d \end{cases}$$

Gene expression data



D. R. Bickel, deposited in uO Research at <https://goo.gl/GcUjJe>

The right balance



Family-wise error rates

Too many false negatives

CFDR

RFDR

Other local FDR estimators?

False discovery rates

Too many false positives

Slides and preprint: www.davidbickel.com

Acknowledgements

- Natural Sciences and Engineering Research Council of Canada
- Agriculture and Agri-Food Canada
- Faculty of Medicine of the University of Ottawa

