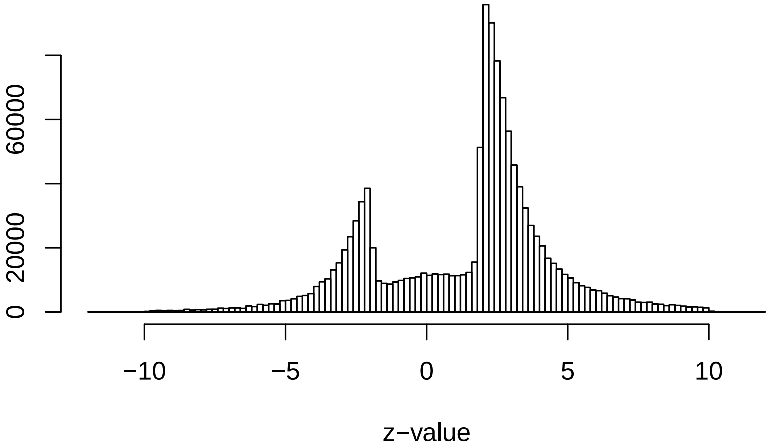
# Lifting the Winner's Curse

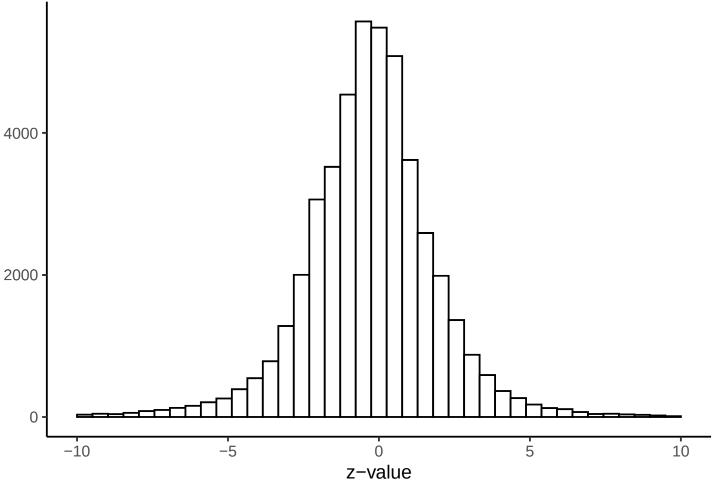
Erik van Zwet, Leiden University Medical Center

BMS Aned
October 7, 2022

# More than one million *z*-values from Medline



# More than 45,000 *z*-values from Cochrane



# 45,955 randomized controlled trials

The "essence" of an RCT is a set of 3 numbers: $(\beta, b, s)$.

- $\beta$ is the primary effect parameter (difference of means, log odds ratio or log hazard ratio)
- $b$ is a normally distributed, unbiased estimator of $\beta$ with standard error $s$.

We have collected 45,955 independent pairs $(b_i, s_i)$ from the Cochrane Database of Systematic Reviews.

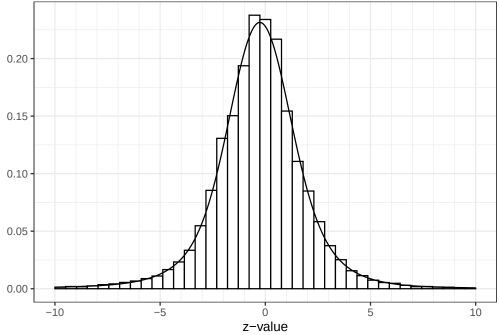
# SNR and $z$-value

We estimate the joint distribution of the $z$-value $z = b/s$ and the signal-to-noise ratio $SNR = \beta/s$ in two steps:

Step 1: We estimate the distribution of $z$ directly from the observed pairs $(b_i, s_i)$.

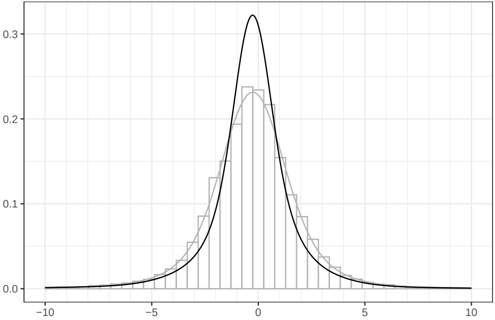
Step 2: The distribution of $z$ is the convolution of the distribution of the $SNR$ with the standard normal. So, the distribution of the $SNR$ can be obtained from the distribution of $z$ by "deconvolution".

# Step 1: Distribution of the *z*-value



The distribution of $z$ is well approximated by a mixture of 4 normal components.

# Step 2: Distribution of the *SNR*



Deconvolution is easy: subtract 1 from the variance of each of the components of the mixture distribution of $z$.

# Some statistical quantities

Important quantities depend on $(\beta, b, s)$ only through $(z, SNR)$:

- coverage:
  $b - 1.96s < \beta < b + 1.96s \Leftrightarrow z - 1.96 < SNR < z + 1.96$
- significance: $|z| > 1.96$
- correct sign: $\beta \cdot b > 0 \Leftrightarrow SNR \cdot z > 0$
- exaggeration: $|b/\beta| = |z/SNR|$

# Coverage

We have the coverage statement

$$P(b - 1.96\,s < \beta < b + 1.96\,s \mid \beta, s) = 0.95$$

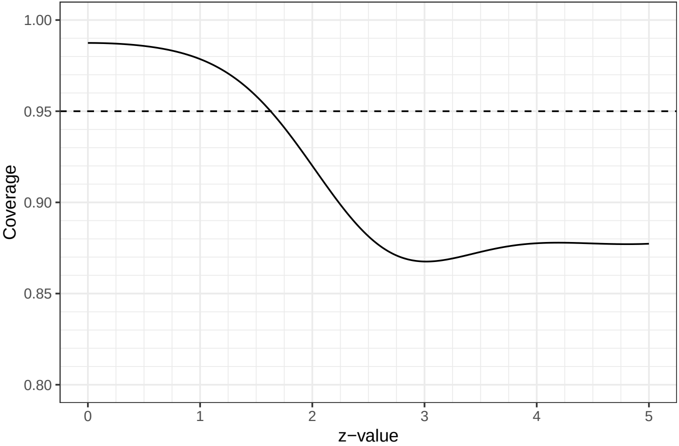
which must *not* be confused with

$$P(b - 1.96\,s < \beta < b + 1.96\,s \mid b, s) = 0.95$$

With the joint distribution of the *SNR* and the *z*-value we can compute the conditional probability of coverage, given the *z*-value:

$$P(b{-}1.96\,s < \beta < b{+}1.96\,s \mid z) = P(z{-}1.96 < SNR < z{+}1.96 \mid z).$$

# Coverage



# Power

RCTs are designed to have 80% or 90% power for testing $H_0 : \beta = 0$ against an alternative that is considered to be of clinical interest, or plausible, or both.

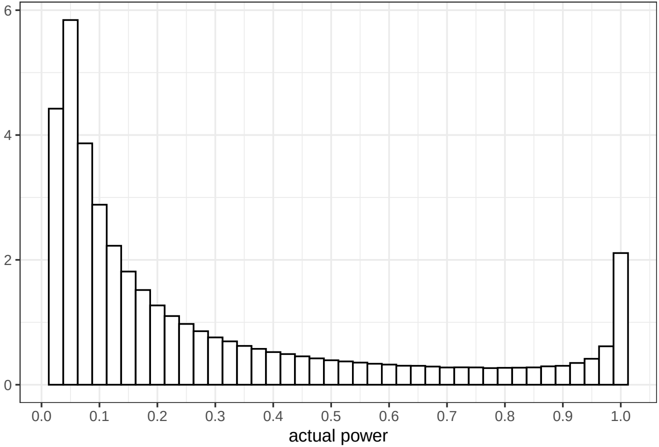
The power against the *true* effect is given by

$$P(|z| > 1.96 \mid \beta, s) = \Phi(-1.96 - SNR) + 1 - \Phi(1.96 - SNR).$$

The probability of a significant result in the right direction is $\Phi(-1.96 + |SNR|)$.

- We can transform our estimate of the distribution of the *SNR* into an estimate of the distribution of the power against the true effect.

# Distribution the power (median=14%, mean=29%)



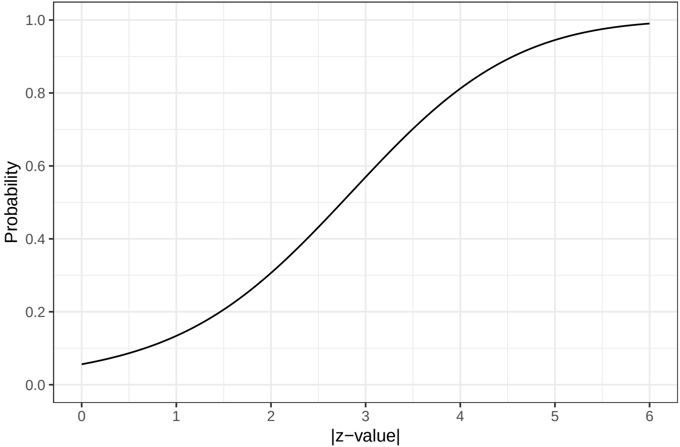
# Predictive power

The power is just a function of the *SNR*. Since we have the joint distribution of the *SNR* and the $z$-value, we also have the conditional distribution of the power given the $z$-value.

The conditional *expectation* of the power given the $z$-value is sometimes called the predictive power.

Predictive power = the probability of a significant result when a study with a particular $z$-value would be repeated exactly.

# Predictive power

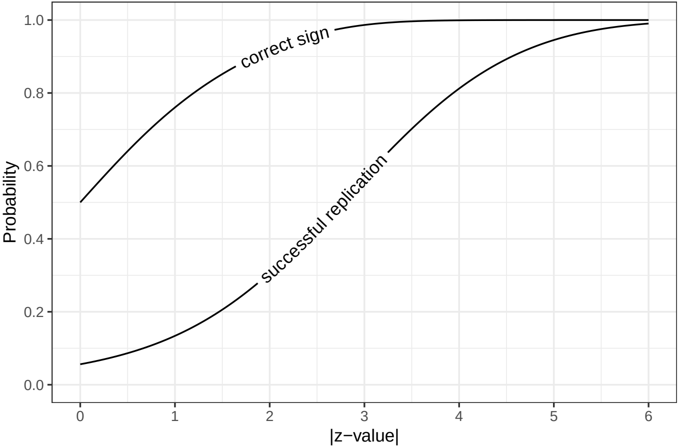


# Sign agreement

With the joint distribution of the $SNR$ and the $z$-value we can compute the conditional probability that $b$ has the same sign as $\beta$, given the observed $z$-value:

$$P(b \cdot \beta > 0 \mid z) = P(z \cdot SNR > 0 \mid z).$$

# The gap



correct sign
successful replication
Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
|z−value|

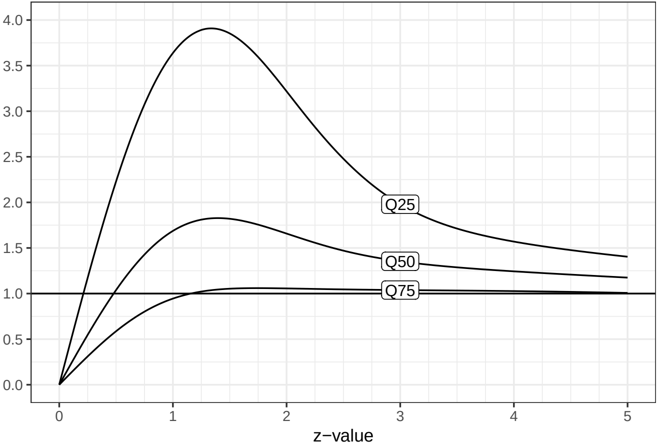
# Exaggeration

Define the exaggeration

$$\frac{|b|}{|\beta|} = \frac{|b|/s}{|\beta|/s} = \frac{|z|}{|SNR|}.$$

From the joint distribution of the $SNR$ and $z$, we can obtain the conditional distribution of the exaggeration given $z$.

# The conditional quartiles of $|b/\beta|$ given $z$



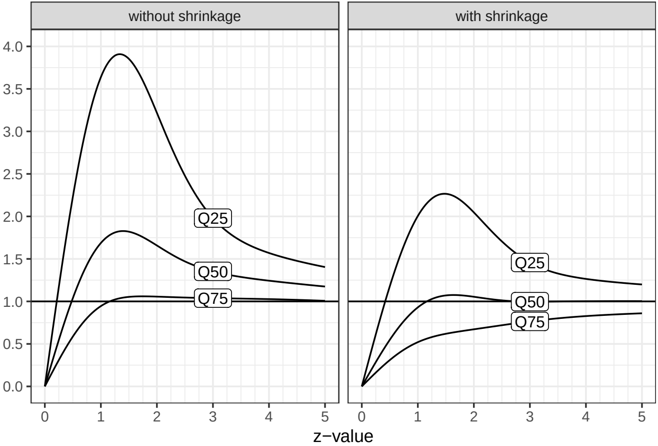
# Shrinkage

The exaggeration can be addressed with shrinkage. Recalling that $SNR = \beta/s$, we propose

$$\hat{\beta} = s\,\mathbb{E}(SNR \mid z).$$

as an alternative to the unbiased estimator $b$.

# The conditional quartiles of $|b/\beta|$ given $z$



# Discussion

- The distribution of the signal-to-noise ratio across the Cochrane database (CDSR) says something about "how medical research is done".
- We can use this information to improve our inferences on average across the CDSR.
- To use this information for the interpretation a particular trial, we must view this trial as *exchangeable* with the trials in the CDSR. This means ignoring all distinguishing features.

# References

1. with Andrew Gelman: A proposal for informative default priors scaled by the standard error of estimates (2022) in *The American Statistician*
2. with Simon Schwab and Stephen Senn: The statistical properties of RCTs and a proposal for shrinkage (2021) in *Statistics in Medicine*
3. with Simon Schwab and Sander Greenland: Addressing exaggeration of effects from single RCTs (2022) in *Significance*
4. with Steven Goodman: How large should the next study be? Predictive power and sample size requirements for replication studies (2022) in *Statistics in Medicine*