BGSE Development Summer School

Lecture 2:

RCT Power and Design

Professors: Pamela Jakiela and Owen Ozier

Lecture 2, Part 1:

Power in Randomized Trials

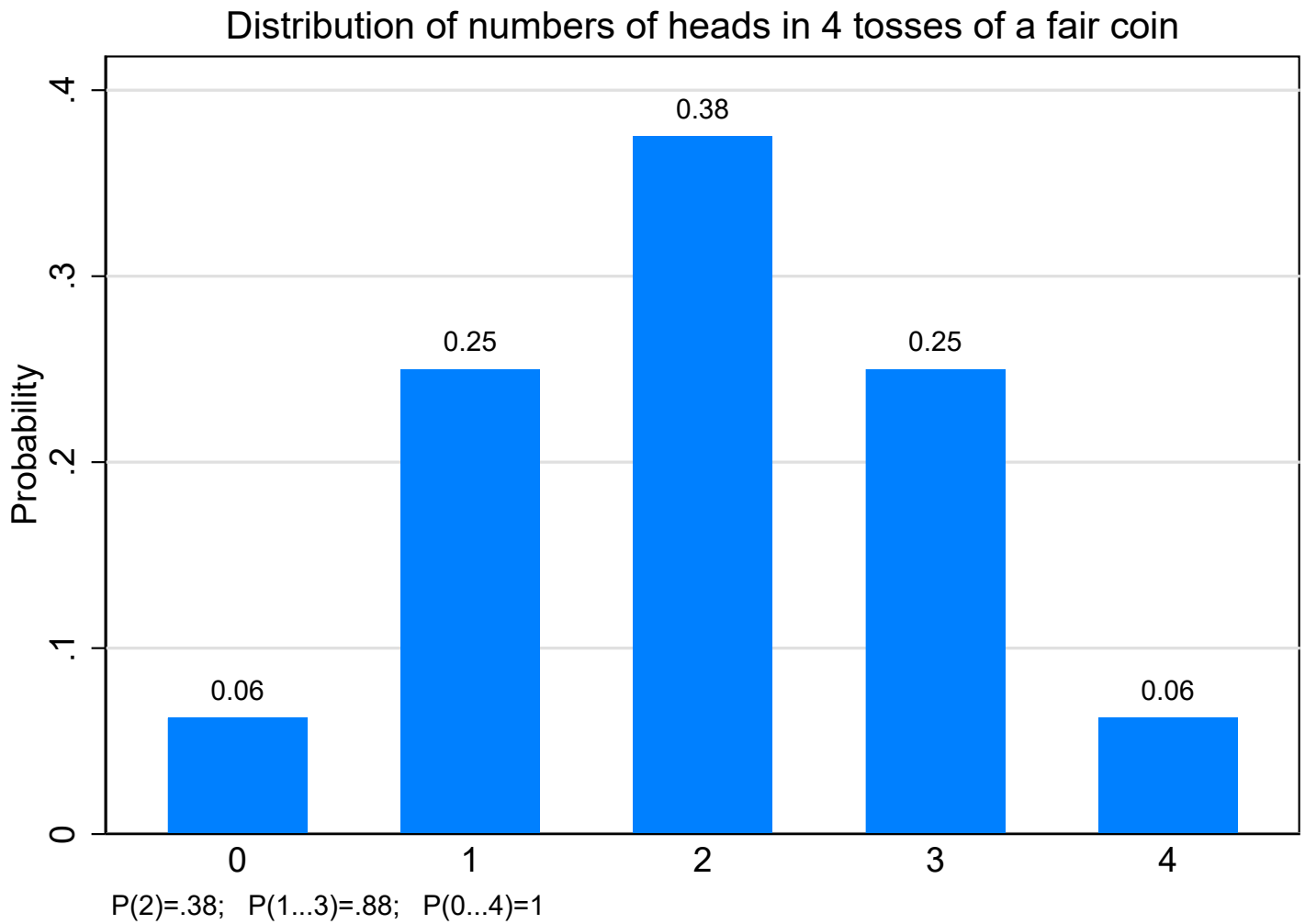
Power

- **Power**:
  probability of rejecting... the null, when... the alternative is true.
- In randomized trials:
  probability of having a statistically significant coefficient on treatment when there is, in fact, an effect of treatment.
- A "power calculation" is... a sample size calculation.
  This means predicting... the standard error.

Coin toss example

- "Null" Hypothesis: the coin is fair
  50% chance of heads, 50% chance of tails.
- Structure of the data:
  Toss the coin a number of times, count heads.
- The test:
  "Fail to reject" null if within some distance of mean under the null;
  "Reject" otherwise.
- If we only had 4 tosses of the coin, what cutoffs could we use?
  Could fail to reject under any of these conditions:
  - (A) never
  - (B) when exactly the mean (2 heads)
  - (C) when within 1 (1, 2, or 3 heads)
  - or (D) always.
- We don't want to reject the null when it is true, though;
  How much accidental rejection would each possible cutoff give us?

Distribution of possible results



Types of error

	Test result	
	"Reject Null," Find an effect!	"Fail to Reject Null," Conclude no effect.
Truth: There is an effect	Great!	"Type II Error" (low power)
Truth: There is NO effect	"Type I Error" (test size)	Great!

The probability of Type I error (given the null) is the "size" of the test. By convention, we are usually interested in tests of "size" 0.05.

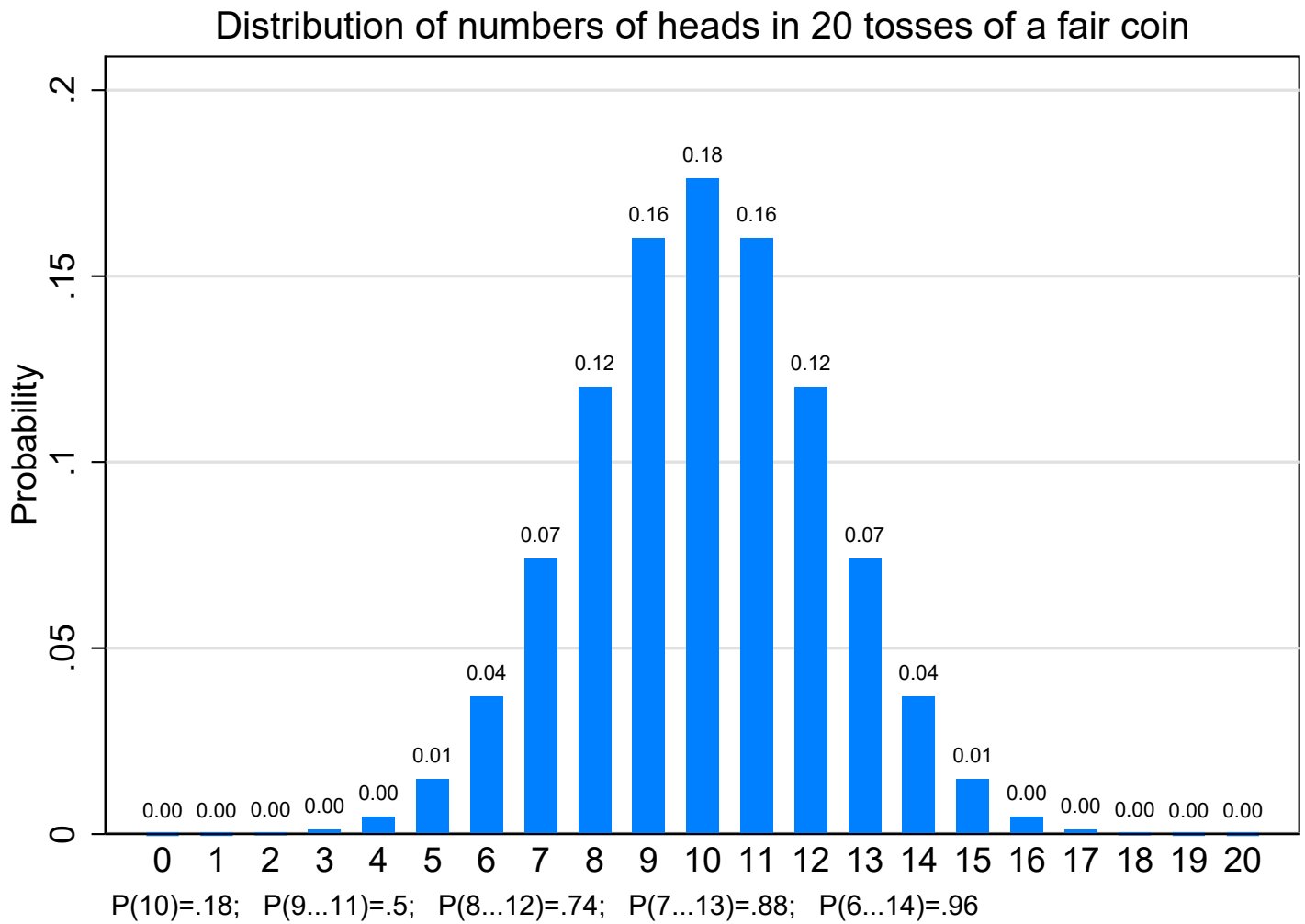
The probability of Type II error is also very important;
If P(failure to detect an effect|there is an effect) $= 1 - \kappa$,
then the power of the test is κ.

Power depends on anticipated effect size; we typically want power $\geq 80\%$.

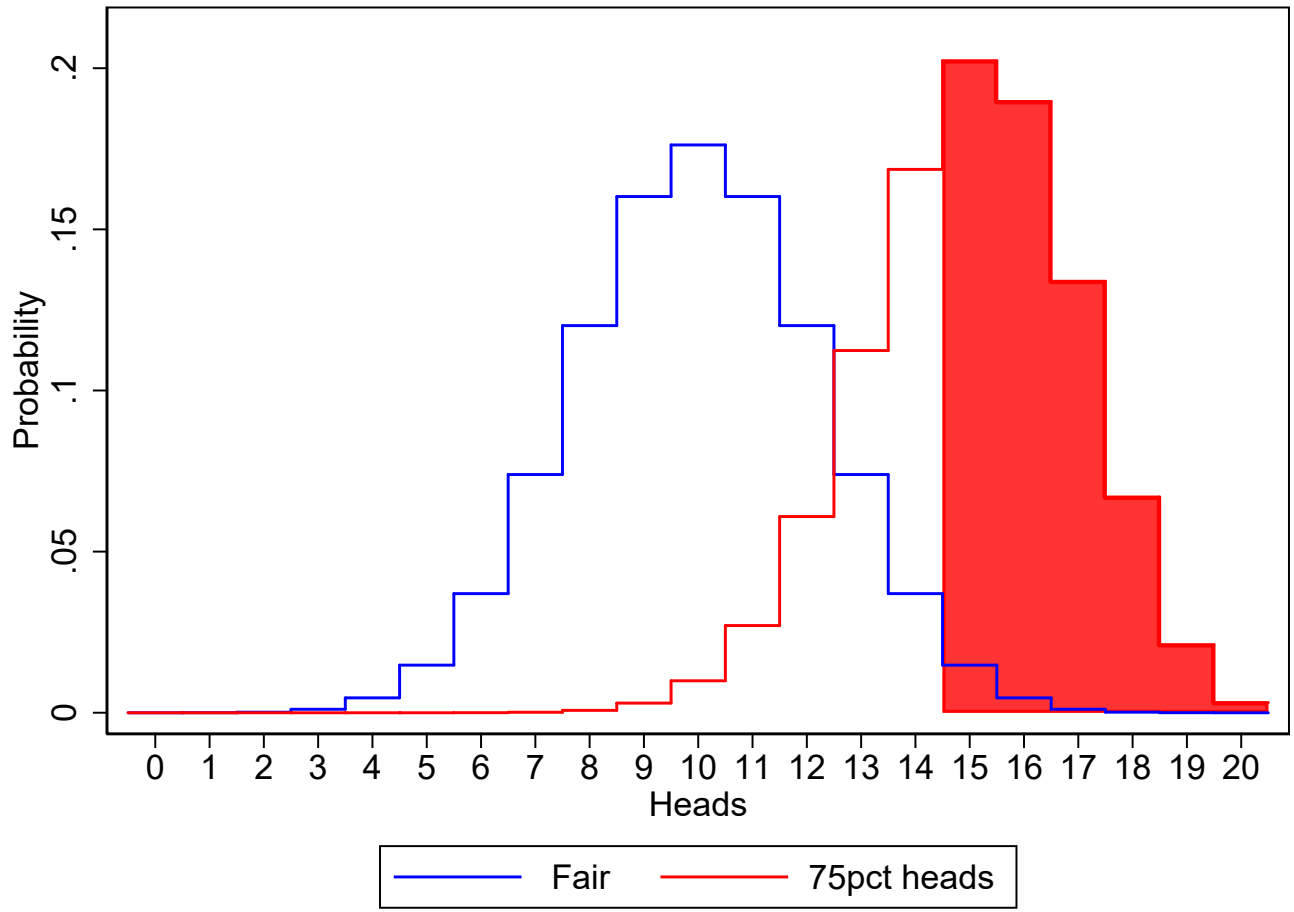
Not enough data even for meaningful test size

- There is no way* to create such a test with four coin tosses so that the chance of accidental rejection under the "null" hypothesis (sometimes written H_0) is less than 5%, a standard in social science.
  ** (Except the "never reject, no matter what" rule. Not very useful.)*
- What about 20 coin tosses?

Distribution of possible results

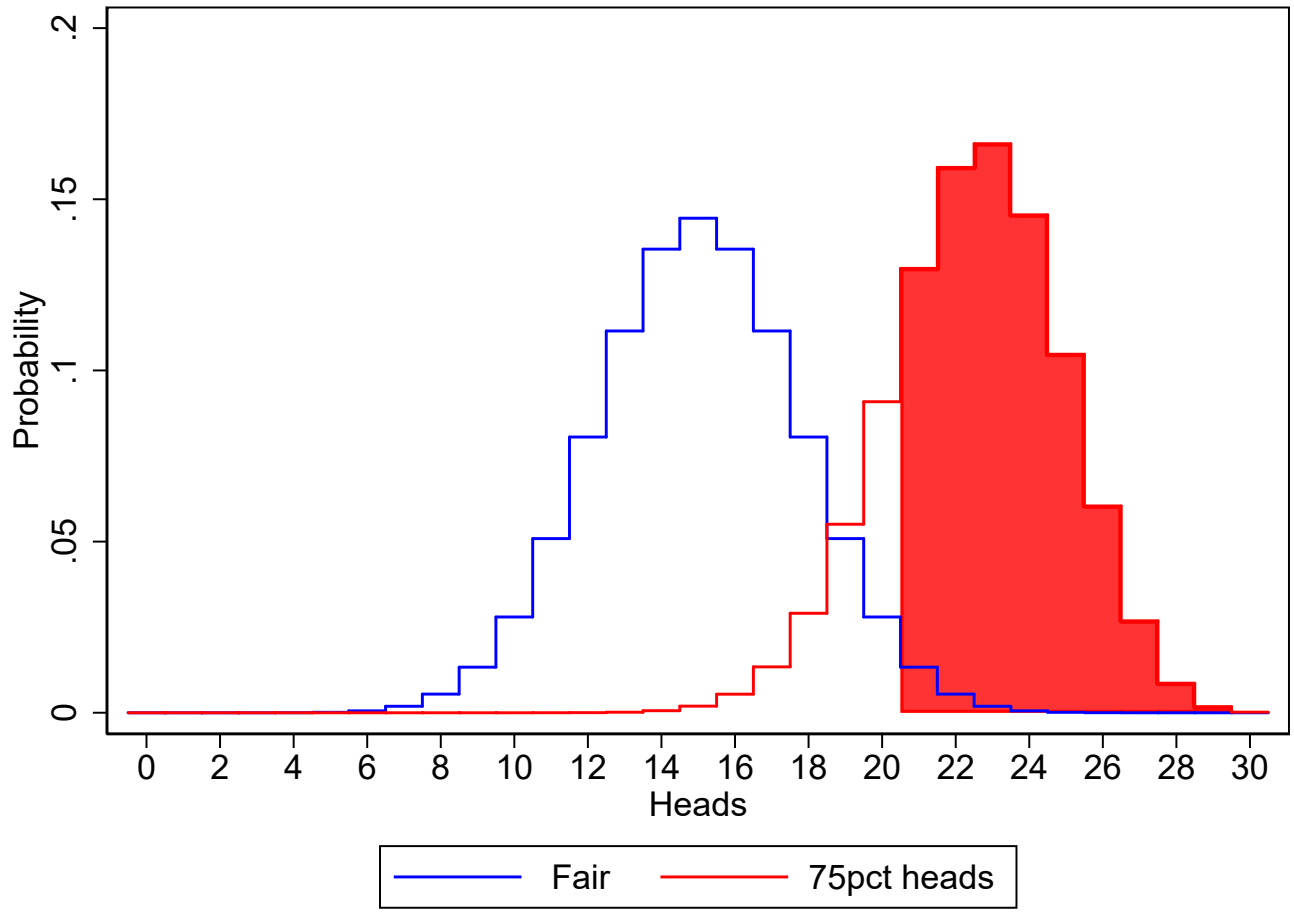


Power with 20 tosses



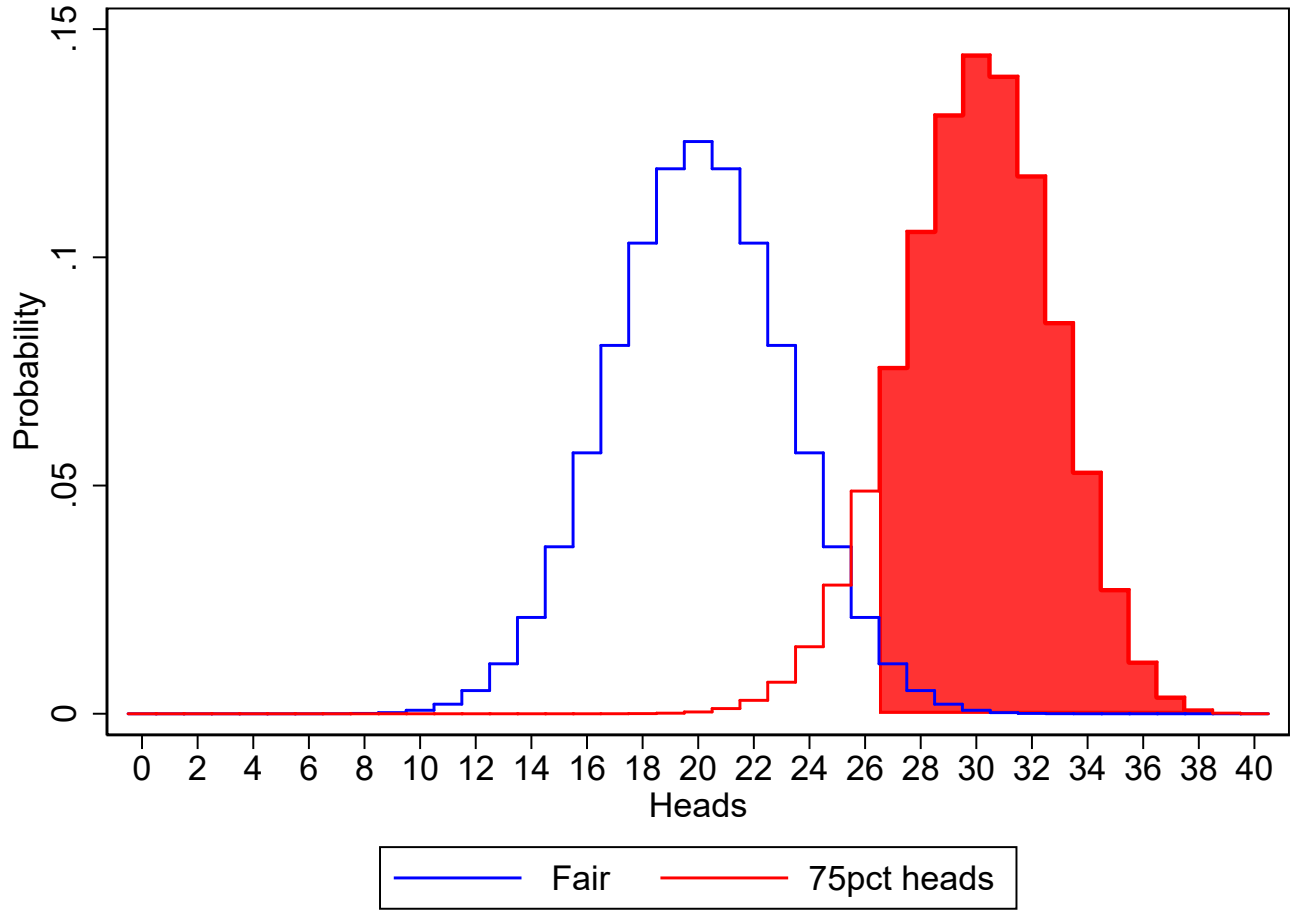
Power: about 0.62

Power with 30 tosses



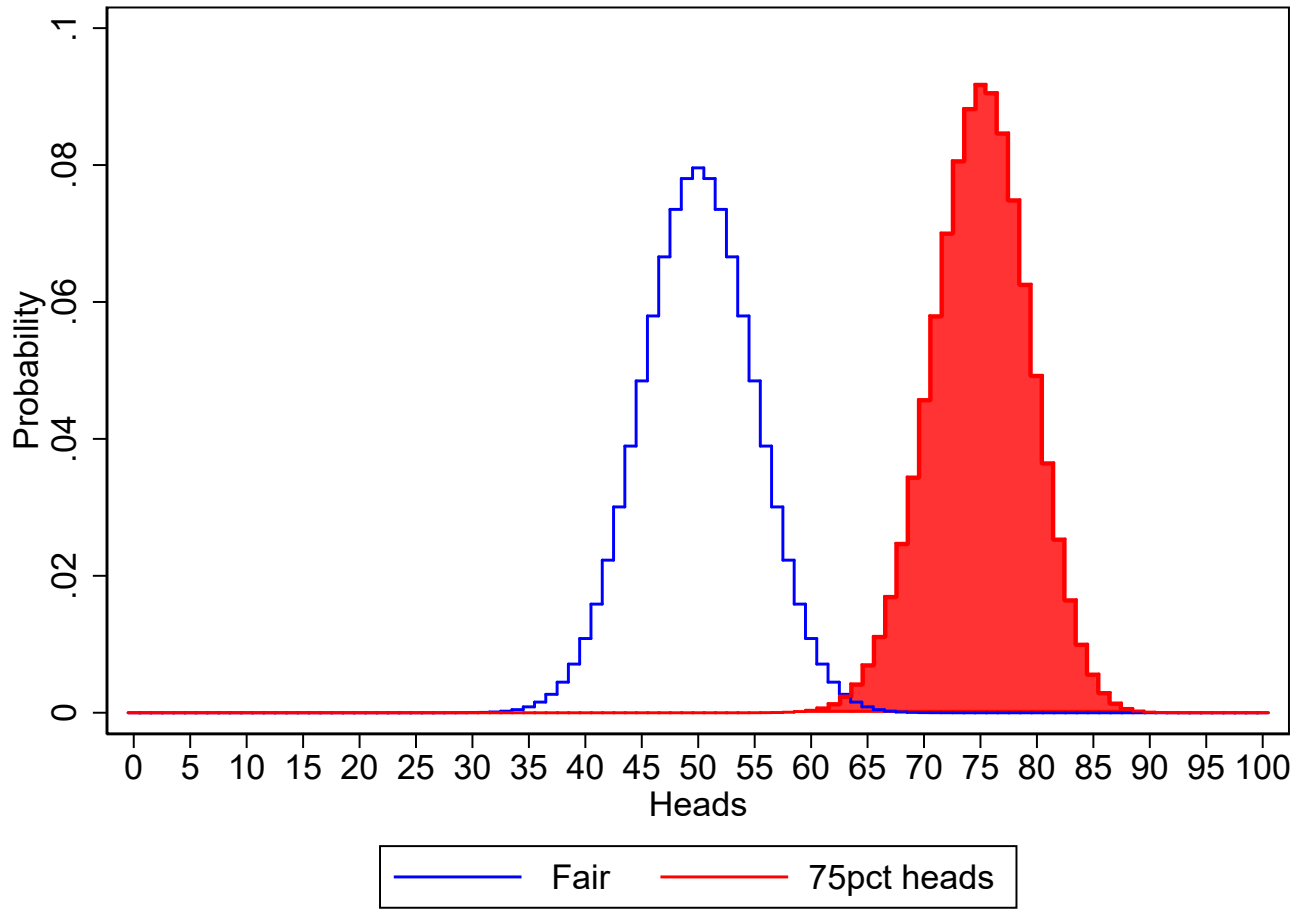
Power: about 0.80

Power with 40 tosses



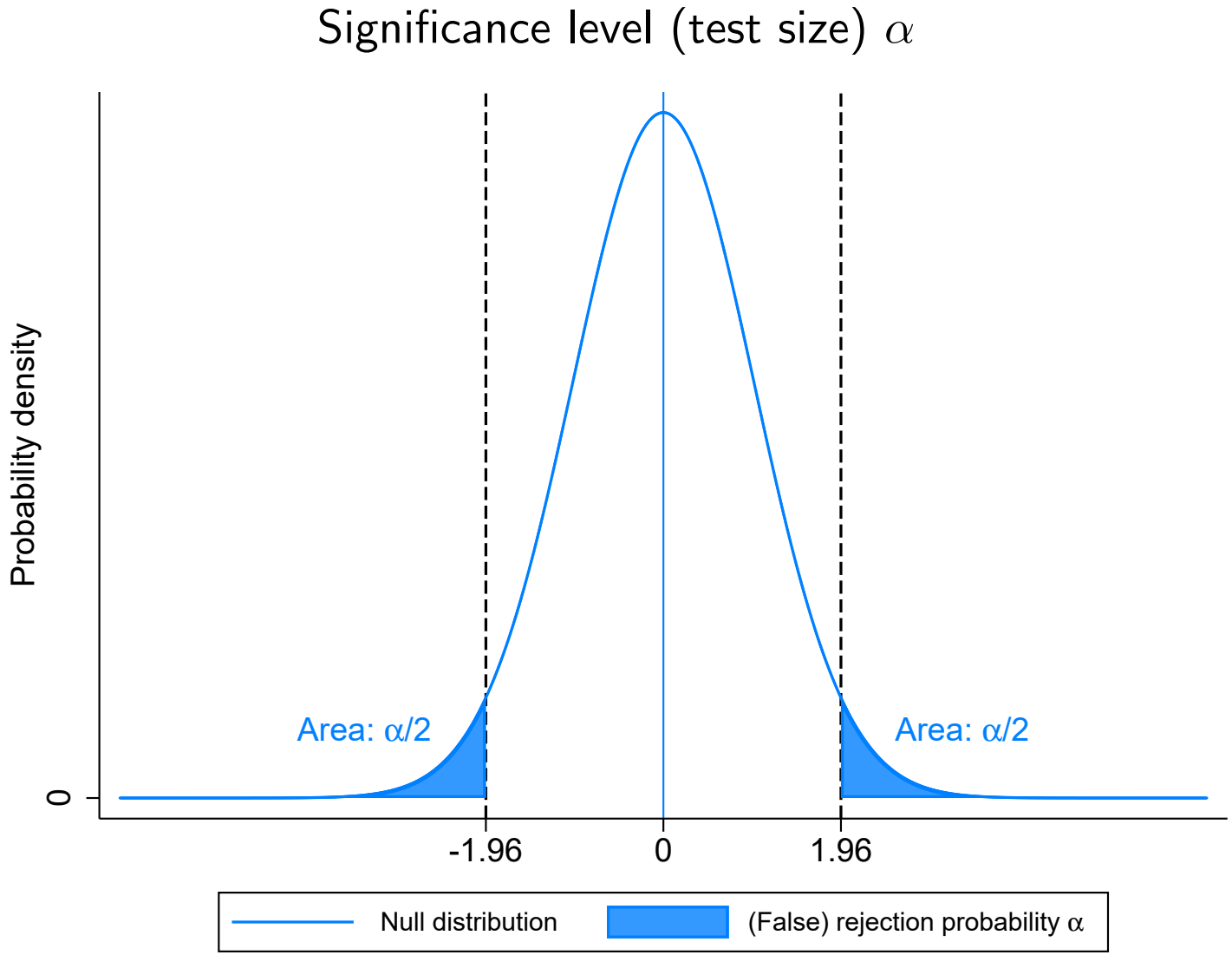
Power: about 0.90

Power with 100 tosses

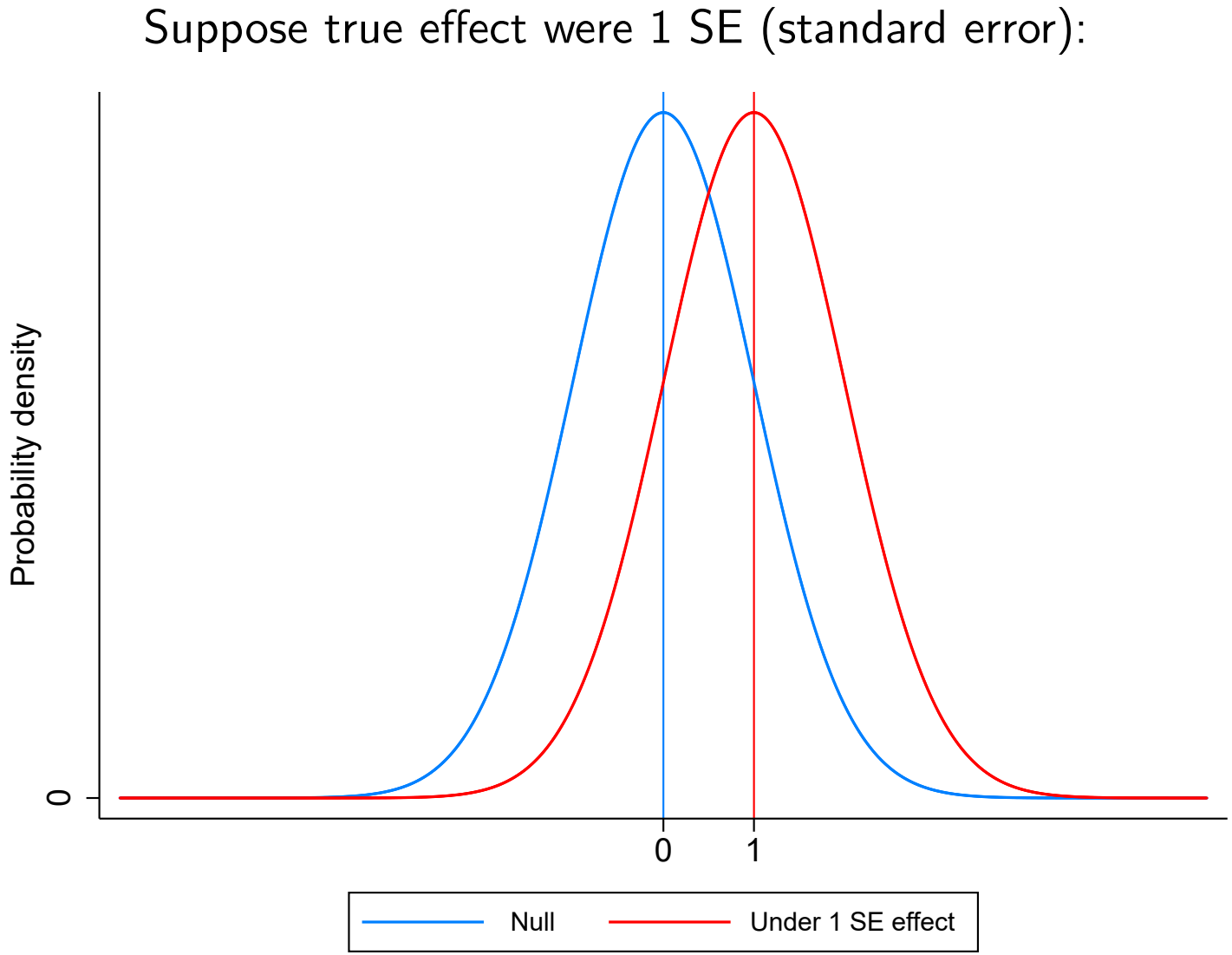


Power: about 0.9997

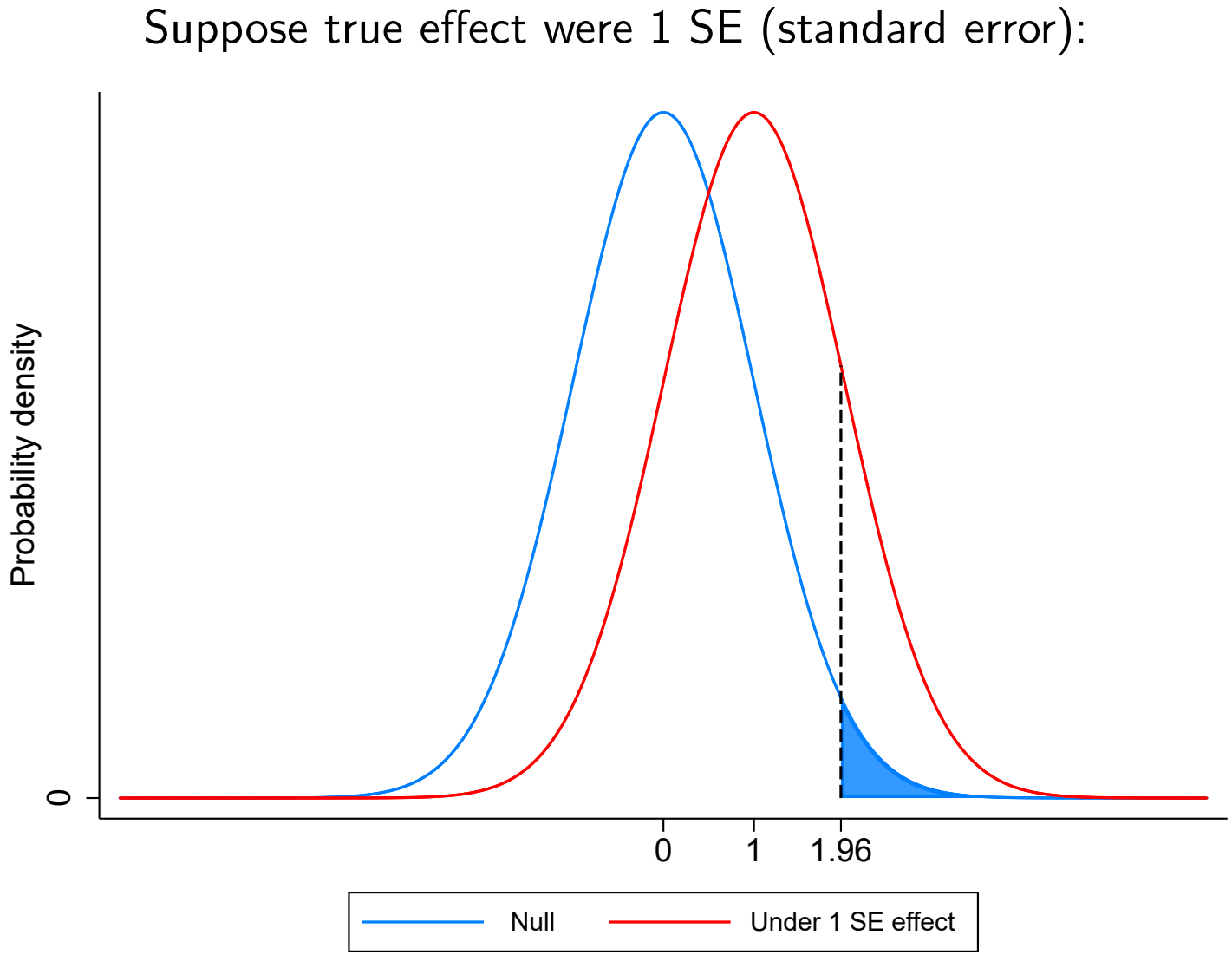
Rejecting H_0 in critical region



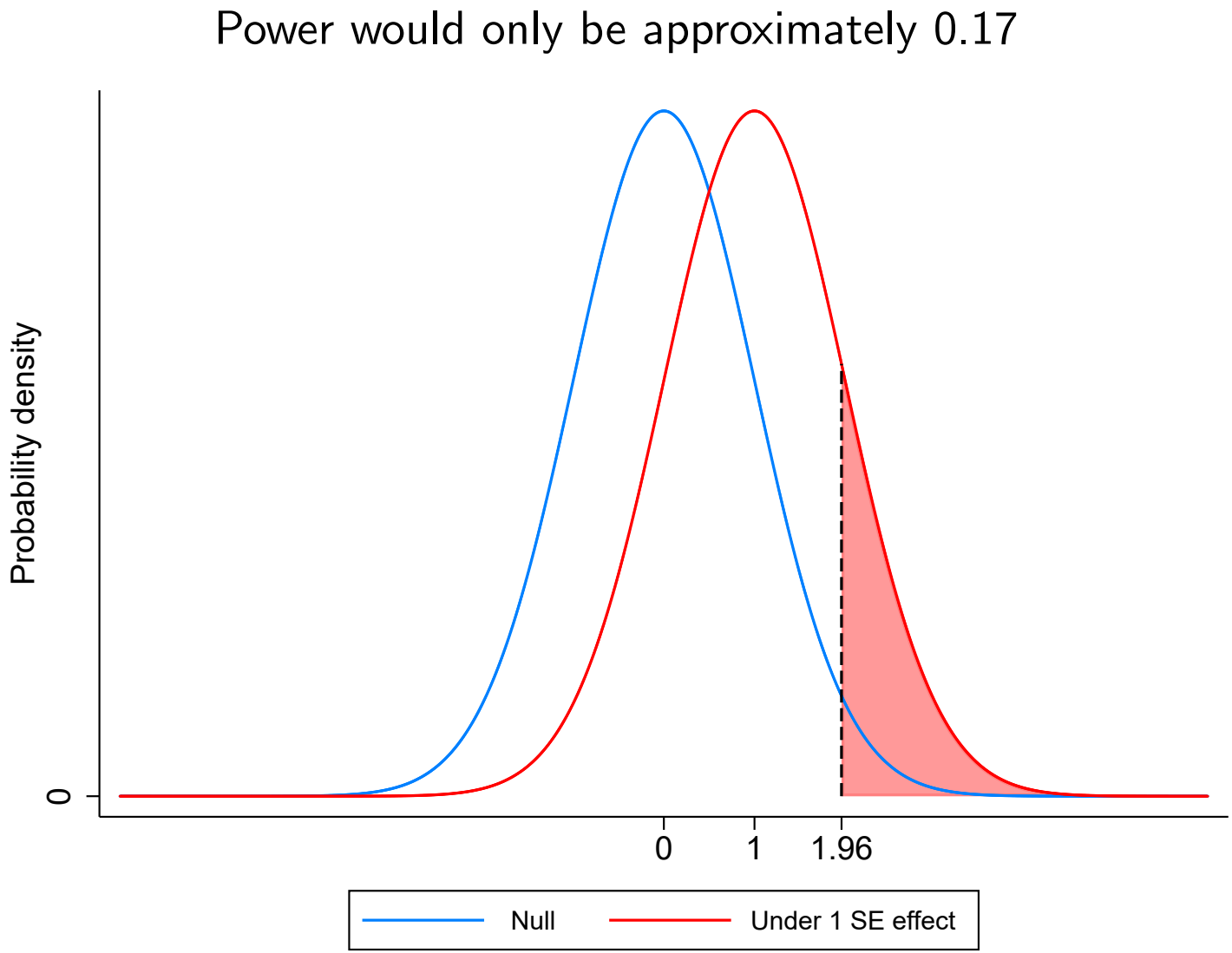
Under an alternative:



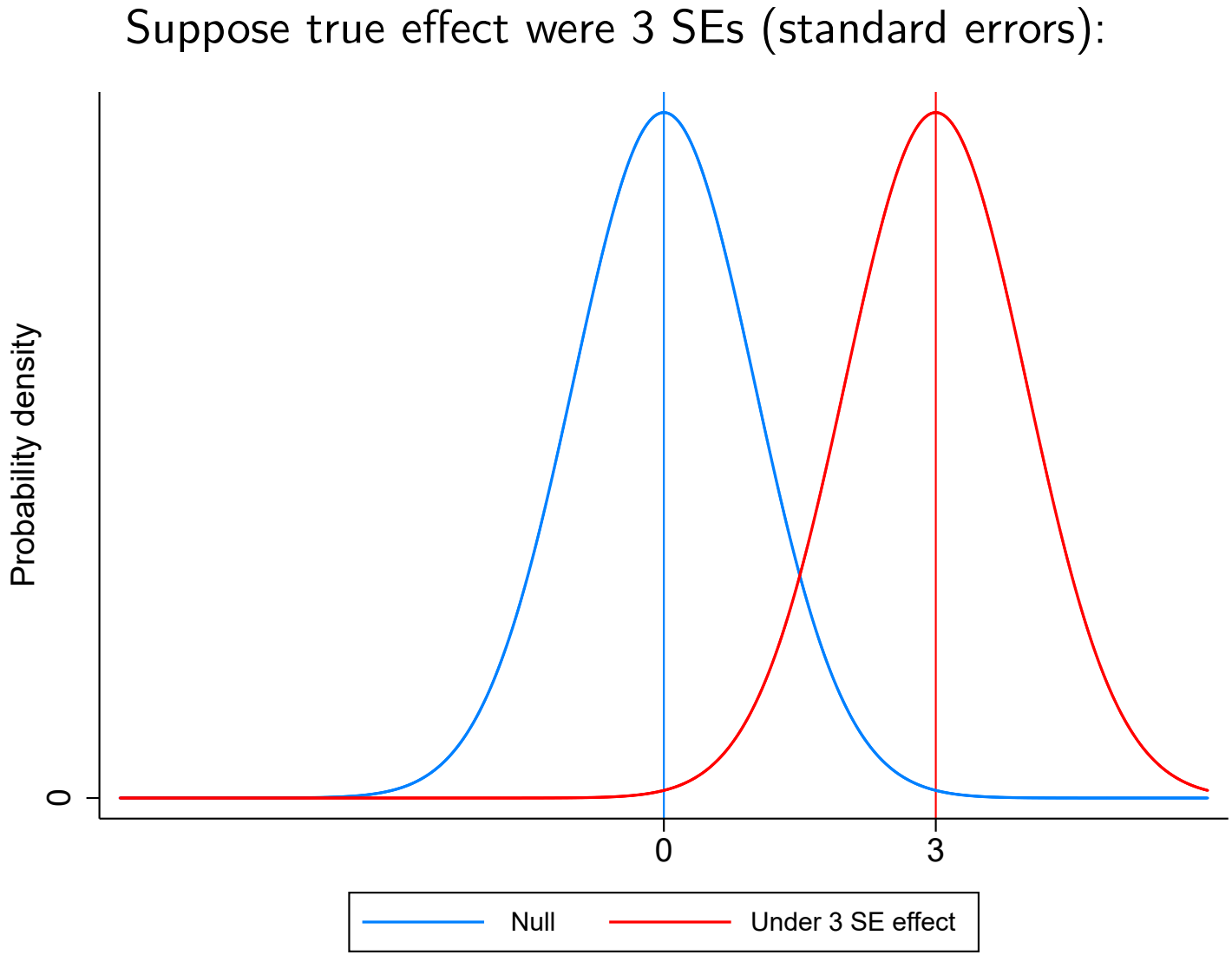
Under an alternative:



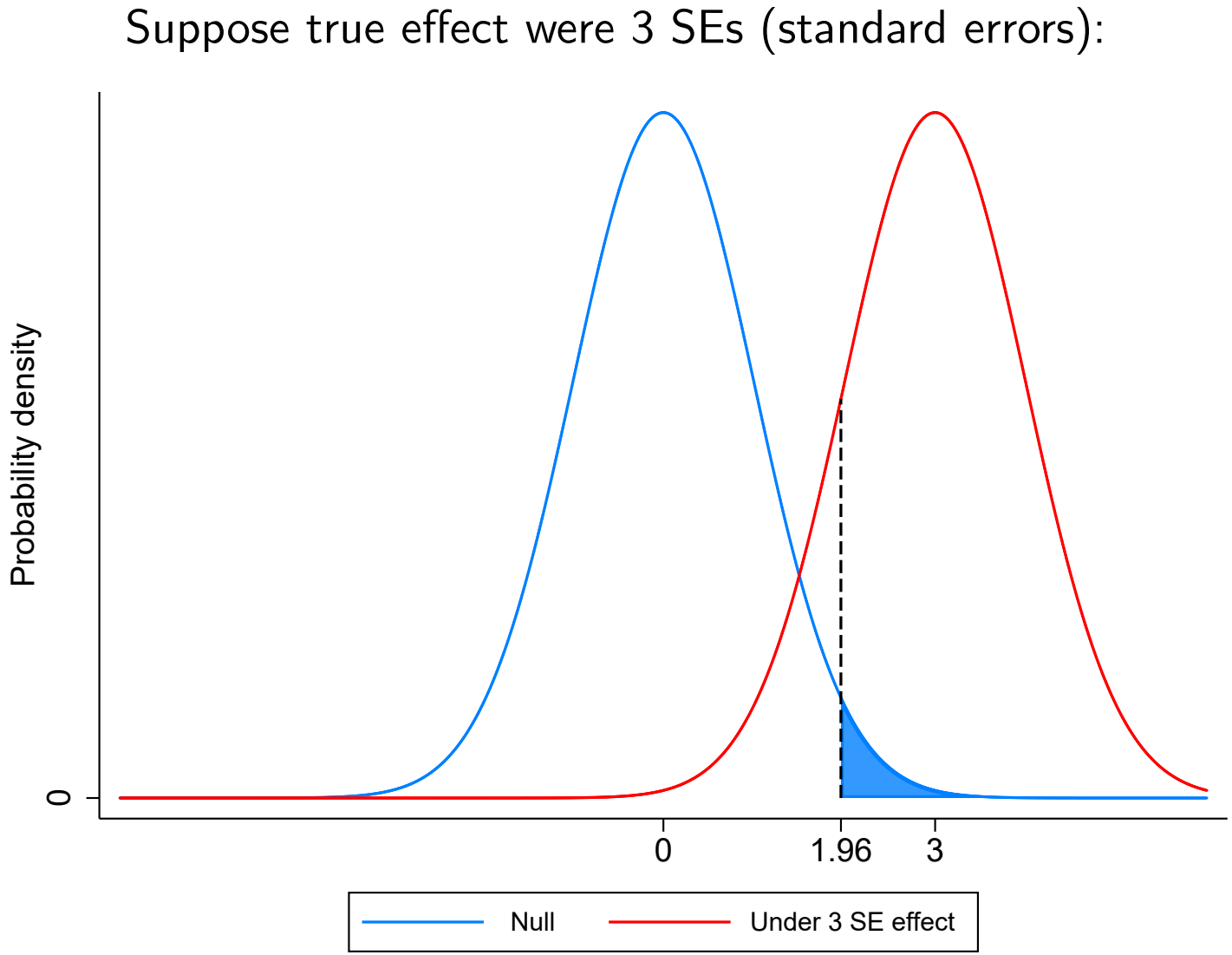
Under an alternative:



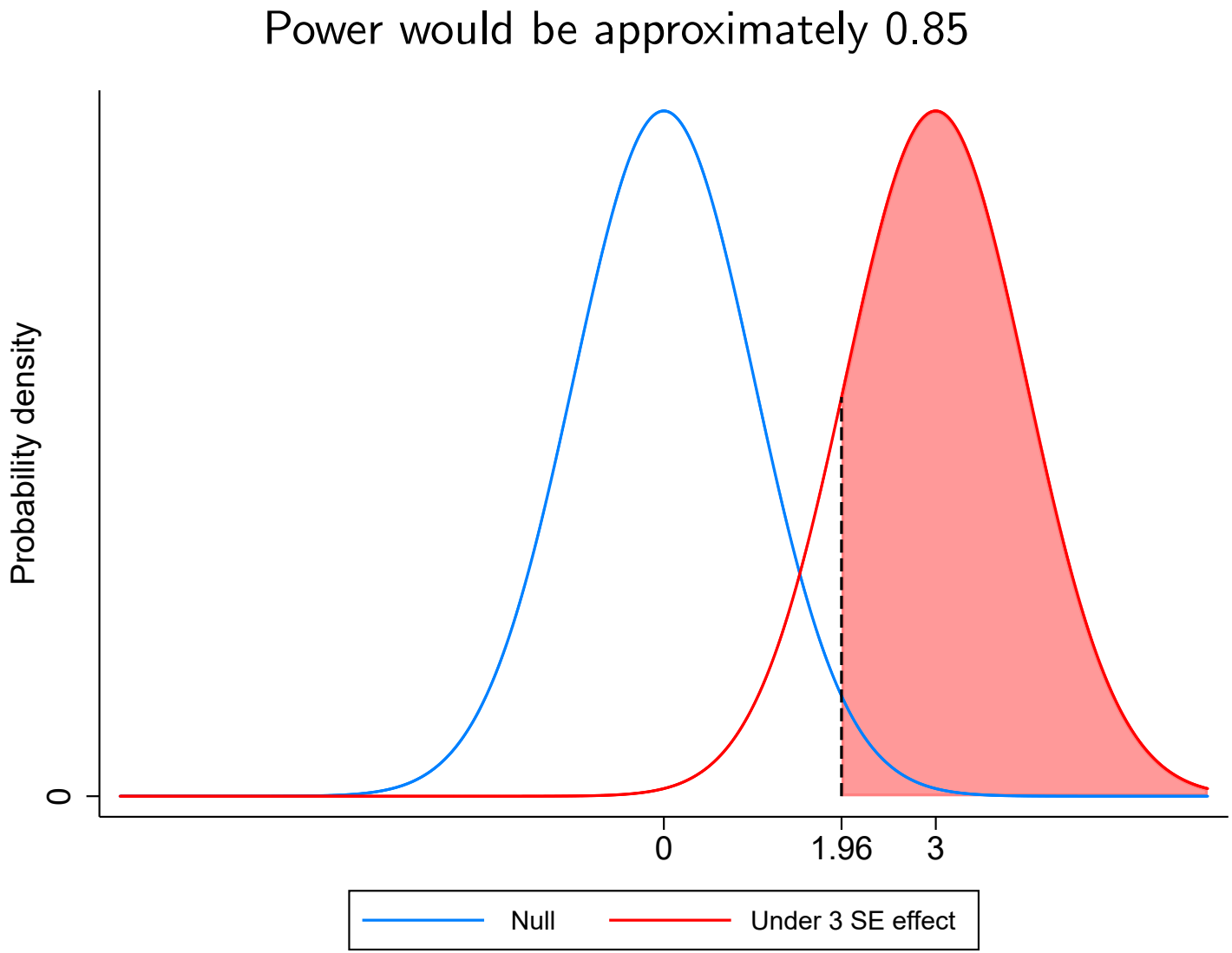
Under an alternative:



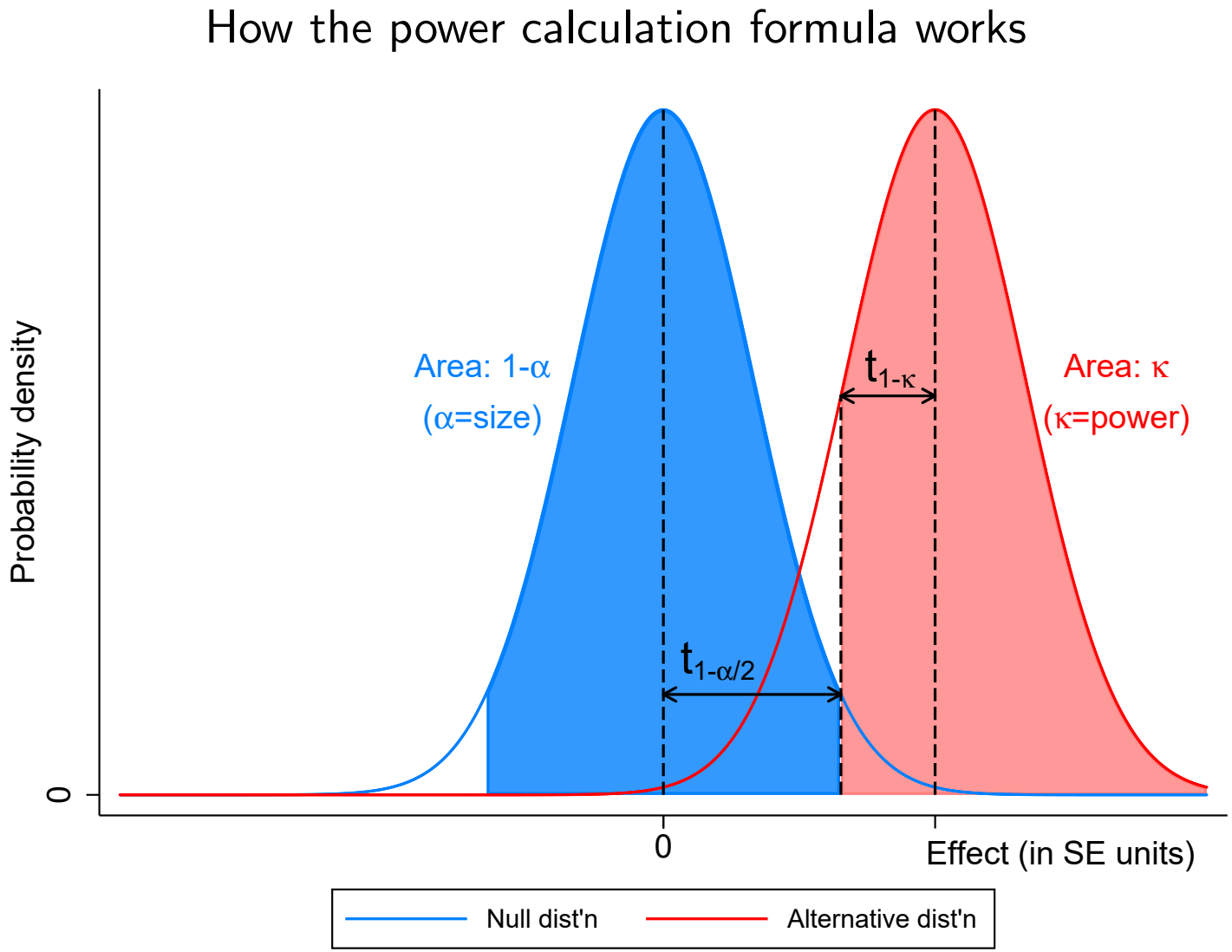
Under an alternative:



Under an alternative:



Power calculation, visually



Note: see the related figure in the *Toolkit* paper.

The formula: for power κ and size α,

$Effect > (t_{1-\kappa}+t_{\alpha/2})SE(\hat{\beta})$ Notation: $t_{1-p} = p^{th}$ percentile of the t dist'n.

Note that the formula above works no matter the design.
Usually: $\alpha = 0.05$, $\kappa = 0.80$, N is large, so:

$$\text{Minimum Detectable Effect} \approx (0.84 + 1.96)SE(\hat{\beta}) \approx 2.8SE(\hat{\beta})$$

We focus on sample size. But how would imperfect compliance or baseline data affect this? Below, I continue for the standard RCT case.

$$MDE = (t_{1-\kappa}+t_{\alpha/2})\sqrt{\frac{1}{P(1-P)}}\sqrt{\frac{\sigma^2}{N}} \quad \approx \quad (z_{1-\kappa}+z_{\alpha/2})\sqrt{\frac{1}{P(1-P)}}\sqrt{\frac{\sigma^2}{N}}$$

In practice (Stata): **sampsi**
Note: Stata uses z rather than t distribution (skirting D.O.F. issue).
We could also flip this equation around:

$$\iff N = (z_{1-\kappa} + z_{\alpha/2})^2 \cdot \left(\frac{1}{P(1-P)}\right) \cdot \left(\frac{\sigma^2}{MDE^2}\right)$$

The formula: for power κ and size α,

Where do these numbers come from, σ^2 and the effect size?
Two basic options:

- Consider standardized effect sizes in terms of standard deviations
- Draw on existing data: What is available that could inform your project?

What if treatment is assigned by groups?

We have been thinking here of randomizing at the individual level.
But in practice, we often randomize larger units.
Examples:

- **Entire schools** are assigned to treatment or comparison;
  we observe outcomes at the level of the individual pupil
- **Classes within a school** are assigned to treatment or comparison;
  we observe outcomes at the level of the individual pupil
- **Households** are assigned to treatment or comparison;
  we observe outcomes at the level of the individual family member
- **Sub-district locations** are assigned to treatment or comparison;
  we observe outcomes at the level of the individual road
- **Bank branch offices** are assigned to treatment or comparison;
  we observe outcomes at the level of the individual borrower

What does this do?
It depends on how much variation is explained by the group each individual is in.

What happens to the variance of the estimator?

Suppose $y_i = \beta t_i + \epsilon_i$. We compare the means of those with $t_i = 1$ to those with $t_i = 0$. Departure point: iid ϵ_i having variance σ^2_ϵ, and equal numbers of observations in treatment and control ($N/2$ in each):

$$\hat{\beta} = \frac{1}{N/2}\sum_T y_i - \frac{1}{N/2}\sum_C y_i$$

$$\hat{\beta} = \beta + \frac{1}{N/2}\sum_T \epsilon_i - \frac{1}{N/2}\sum_C \epsilon_i$$

$$Var(\hat{\beta}) = \frac{1}{N/2}\sigma^2_\epsilon + \frac{1}{N/2}\sigma^2_\epsilon = \frac{4}{N}\sigma^2_\epsilon$$

$$SE(\hat{\beta}) = \sqrt{4}\sqrt{\frac{\sigma^2_\epsilon}{N}}$$

This is the formula from before, with $P = 1/2$:

$$\sqrt{\frac{1}{P(1-P)}}\sqrt{\frac{\sigma^2}{N}}$$

What happens to the variance of the estimator?

Now suppose $y_i = \beta t_i + \epsilon_i$, but $\epsilon_i = \nu_g + \eta_{ig}$ for groups g of fixed size n_g. We still compare the means of those with $t_i = 1$ to those with $t_i = 0$. Departure point: within a group, treatment is either 1 or 0; iid ν_g having variance σ^2_ν, iid η_{ig} having variance σ^2_η, so that $\sigma^2_\epsilon = \sigma^2_\nu + \sigma^2_\eta$, and equal numbers of observations in treatment and control (still $N/2$ in each). Define the "the intra-cluster correlation," ρ:

$$\rho_\epsilon = \frac{\sigma^2_\nu}{\sigma^2_\nu + \sigma^2_\eta} = \frac{\sigma^2_\nu}{\sigma^2_\epsilon}$$

Two other ways of writing this will be convenient:

$$\sigma^2_\nu = \rho_\epsilon \sigma^2_\epsilon$$

$$\sigma^2_\eta = (1 - \rho_\epsilon)\sigma^2_\epsilon$$

What happens to the variance of the estimator?

Let's think through two pieces of ϵ_i and the variance of their sums; we will need this in just a moment. Within a single study arm (so consider $N/2$ observations). First, the **simple** case, η_{ig}:

$$
\begin{aligned}
Var\left(\frac{1}{N/2}\sum_{arm}\eta_{ig}\right) &= \frac{1}{(N/2)^2}Var\left(\sum_{arm}\eta_{ig}\right) \\
&= \frac{1}{(N/2)^2}Var\left(\sum_{1}^{(N/2)}\eta_{ig}\right) \\
&= \frac{1}{(N/2)^2}\left(\frac{N}{2}\right)\sigma_\eta^2 \\
&= \frac{1}{N/2}\sigma_\eta^2
\end{aligned}
$$

What happens to the variance of the estimator?

Let's think through two pieces of ϵ_i and the variance of their sums; we will need this in just a moment. Within a single study arm (so consider $N/2$ observations). Now, the **slightly more complicated** case, ν_g:

$$
\begin{aligned}
Var\left(\frac{1}{N/2}\sum_{arm}\nu_g\right) &= \frac{1}{(N/2)^2}Var\left(\sum_{arm}\nu_g\right) \\
&= \frac{1}{(N/2)^2}Var\left(\sum_{1}^{(N/2n_g)} n_g\nu_g\right) \\
&= \frac{1}{(N/2)^2}n_g^2 Var\left(\sum_{1}^{(N/2n_g)} \nu_g\right) \\
&= \frac{1}{(N/2)^2}n_g^2\left(\frac{N}{2n_g}\right)\sigma_\nu^2 \\
&= \frac{n_g}{N/2}\sigma_\nu^2
\end{aligned}
$$

What happens to the variance of the estimator?

As before,

$$\hat{\beta} = \beta + \frac{1}{N/2}\sum_{T}\epsilon_i - \frac{1}{N/2}\sum_{C}\epsilon_i$$

$$\hat{\beta} = \beta + \frac{1}{N/2}\sum_{T}\nu_g + \frac{1}{N/2}\sum_{T}\eta_{ig} - \frac{1}{N/2}\sum_{C}\nu_g - \frac{1}{N/2}\sum_{C}\eta_{ig}$$

$$\begin{aligned}
Var(\hat{\beta}) =& \frac{n_g}{N/2}\sigma_\nu^2 + \frac{1}{N/2}\sigma_\eta^2 + \frac{n_g}{N/2}\sigma_\nu^2 + \frac{1}{N/2}\sigma_\eta^2 \\
=& \frac{4n_g}{N}\sigma_\nu^2 + \frac{4}{N}\sigma_\eta^2 \\
=& \frac{4}{N}\left(n_g\sigma_\nu^2 + \sigma_\eta^2\right) \\
=& \frac{4}{N}\left(n_g\rho_\epsilon\sigma_\epsilon^2 + (1-\rho_\epsilon)\sigma_\epsilon^2\right) = \frac{4}{N}\sigma_\epsilon^2\left((n_g-1)\rho_\epsilon + 1\right)
\end{aligned}$$

$$SE(\hat{\beta}) = \sqrt{4}\sqrt{\frac{\sigma_\epsilon^2}{N}}\sqrt{(n_g-1)\rho_\epsilon + 1} = \sqrt{\frac{1}{P(1-P)}}\sqrt{\frac{\sigma^2}{N}}\sqrt{(n_g-1)\rho_\epsilon + 1}$$

The formula

Scale the effective standard error by:

$$\textit{Design Effect (“Moulton factor”)} = \sqrt{1 + (n_{groupsize} - 1)\rho}$$

ρ (“rho”) is the intra-class correlation.
In practice (Stata): **loneway** and **sampclus**

Recall earlier formula:

$$MDE = (t_{1-\kappa} + t_{\alpha/2})\sqrt{\frac{1}{P(1-P)}}\sqrt{\frac{\sigma^2}{N}}\sqrt{1 + (n_{groupsize} - 1)\rho}$$

We could also flip this equation around (swapping z for t):

$$\iff N = (z_{1-\kappa} + z_{\alpha/2})^2 \cdot \left(\frac{1}{P(1-P)}\right) \cdot \left(\frac{\sigma^2}{MDE^2}\right) \cdot (1 + (n_{groupsize} - 1)\rho)$$

Estimation example: clustered standard errors

Stata:

$$V_{cluster} = (X'X)^{-1} \sum_{j=1}^{n_c} u_j' u_j (X'X)^{-1}$$

where

$$u_j = \sum_{j_{cluster}} e_i x_i$$

Angrist and Pischke 8.2.6:

$$\hat{\Omega}_{cl} = (X'X)^{-1} \left(\sum_g X_g' \hat{\Psi}_g X_g \right) (X'X)^{-1}$$

where

$$\hat{\Psi}_g = a \hat{e}_g \hat{e}_g' = a \begin{bmatrix} \hat{e}_{1g}^2 & \hat{e}_{1g}\hat{e}_{2g} & \dots & \hat{e}_{1g}\hat{e}_{n_g g} \\ \hat{e}_{2g}\hat{e}_{1g} & \hat{e}_{2g}^2 & \dots & \hat{e}_{2g}\hat{e}_{n_g g} \\ \dots & \dots & \dots & \dots \\ \hat{e}_{n_g g}\hat{e}_{1g} & \hat{e}_{n_g g}\hat{e}_{2g} & \dots & \hat{e}_{n_g g}^2 \end{bmatrix}$$

Estimation example: clustered standard errors

But remember, in the simplest case, X_g' is either:

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \end{bmatrix} \text{ or } \begin{bmatrix} 0 & 0 & \dots & 0 \\ 1 & 1 & \dots & 1 \end{bmatrix}$$

So

$$X_g' \begin{bmatrix} \hat{e}_{1g}^2 & \hat{e}_{1g}\hat{e}_{2g} & \dots & \hat{e}_{1g}\hat{e}_{n_g g} \\ \hat{e}_{2g}\hat{e}_{1g} & \hat{e}_{2g}^2 & \dots & \hat{e}_{2g}\hat{e}_{n_g g} \\ \dots & \dots & \dots & \dots \\ \hat{e}_{n_g g}\hat{e}_{1g} & \hat{e}_{n_g g}\hat{e}_{2g} & \dots & \hat{e}_{n_g g}^2 \end{bmatrix} X_g$$

Count the terms. diagonal: n_g; off-diagonal: $n_g(n_g - 1)$.
Diagonal terms have expectation σ_ϵ^2,
while off-diagonal terms have expectation $\sigma_\nu^2 = \rho\sigma_\epsilon^2$.
The matrix product then has expectation:

$$\sigma_\epsilon^2 n_g(1 + (n_g - 1)\rho) \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \text{ or } \sigma_\epsilon^2 n_g(1 + (n_g - 1)\rho) \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Estimation example: clustered standard errors

So:

$$E\left[\left(\sum_g X_g' \hat{\Psi}_g X_g\right)\right] = \sigma_\epsilon^2(1+(n_g-1)\rho)\begin{bmatrix} \frac{N}{2} & \frac{N}{2} \\ \frac{N}{2} & N \end{bmatrix}$$

This is a familiar matrix - it is $X'X$!

and thus

$$E\left[\hat{\Omega}_{cl}\right] = E\left[(X'X)^{-1}\left(\sum_g X_g' \hat{\Psi}_g X_g\right)(X'X)^{-1}\right]$$

$$= (1+(n_g-1)\rho)\ (X'X)^{-1}\ \sigma_\epsilon^2$$

Intra-cluster correlation ρ (greek letter "rho")

But where does this ρ number come from before you have endline data?
Two basic options:

- Consider what might be reasonable assumptions
- Draw on existing data (again):
  What is available that could inform your project?

Intra-class correlations we have known

Data source	ICC (ρ)
Madagascar Math + Language	0.5
Busia, Kenya Math + Language	0.22
Udaipur, India Math + Language	0.23
Mumbai, India Math + Language	0.29
Vadodara, India Math + Language	0.28
Busia, Kenya Math	0.62
Busia, Kenya Language	0.43
Busia, Kenya Science	0.35

Duflo, Glennerster, and Kremer (2006) Using Randomization in Development Economics Research: A Toolkit

Data source	ICC (ρ)
US Elementary Math, unconditional	0.22
US Elementary Math, rural only, unconditional	0.15
US Elementary Math, rural only, conditional on previous scores	0.12

Hedges & Hedberg (2007), Intraclass correlations for planning group randomized experiments in rural education.

More variations

For discussion or further reading:

- Imperfect compliance with treatment;
- Alternative tests
- Small numbers of groups
- *"A first comment is that, despite all the precision of these formulas, power calculations involve substantial guess work in practice."*

May be discussed in later lectures:

- Multiple treatments, multiple testing, attrition

Next:

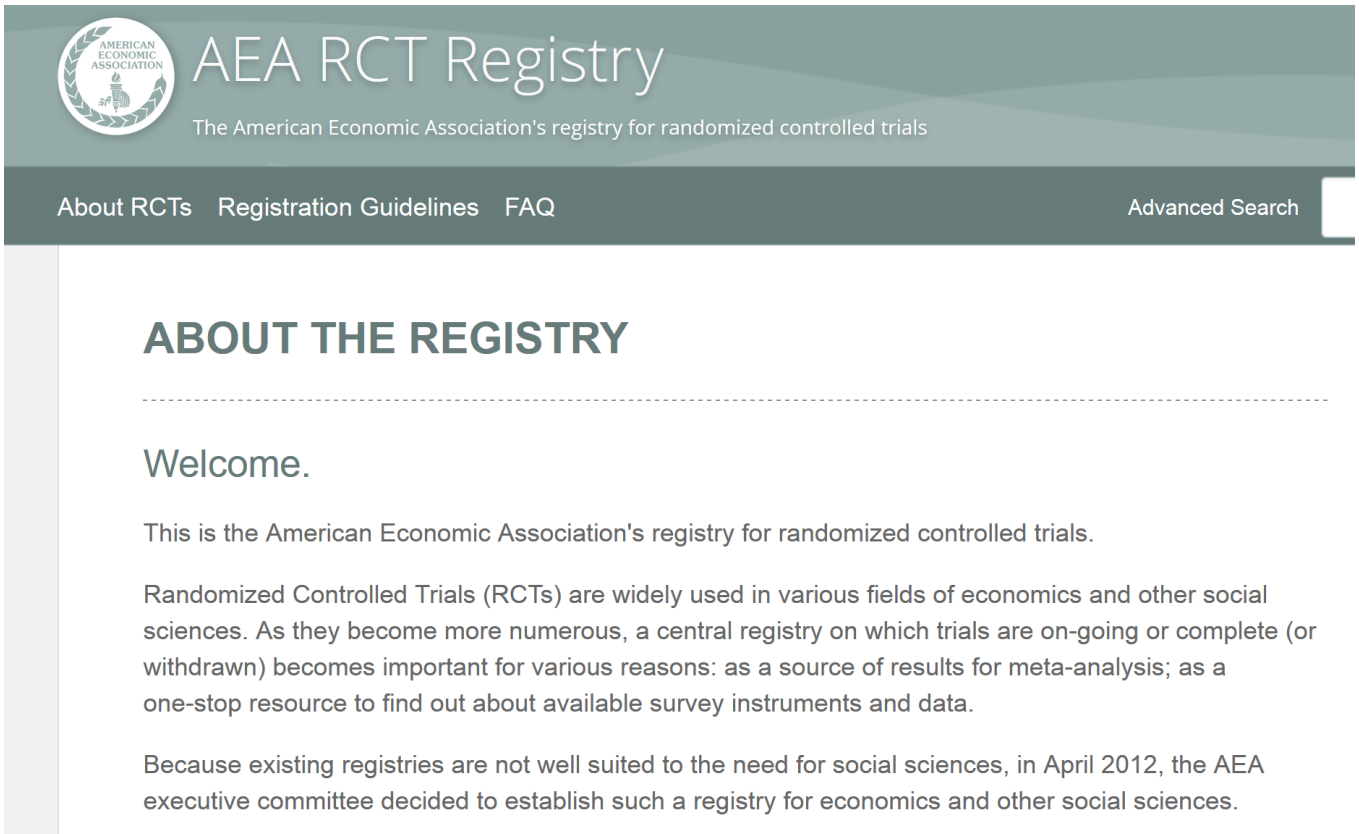
- Actual mechanics of randomization; covariates; stratification

Lecture 2, Part 2:

Design and Balance in Randomized Trials

Besides statistics, registration

Economics: since 2012



Besides statistics, registration

Other disciplines: this example since 2000

BioMed Central
The Open Access Publisher

ISRCTN registry

What is the ISRCTN registry?

ISRCTN is a registry and curated database containing the basic set of data items deemed essential to describe a study at inception, as per the requirements set out by the World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP) and the International Committee of Medical Journal Editors (ICMJE) guidelines. All study records in the database are freely accessible and searchable and have been assigned an ISRCTN ID.

The registry was launched in 2000, in response to the growing body of opinion in favour of prospective registration of randomised controlled trials (RCTs). Originally ISRCTN stood for 'International Standard Randomised Controlled Trial Number'; however, over the years the scope of the registry has widened beyond randomized controlled trials to include any study designed to assess the efficacy of health interventions in a human population. This includes both observational and interventional trials.

Other registries include non-RCTs, focus on specific fields, etc.

Besides statistics, documentation: CONSORT

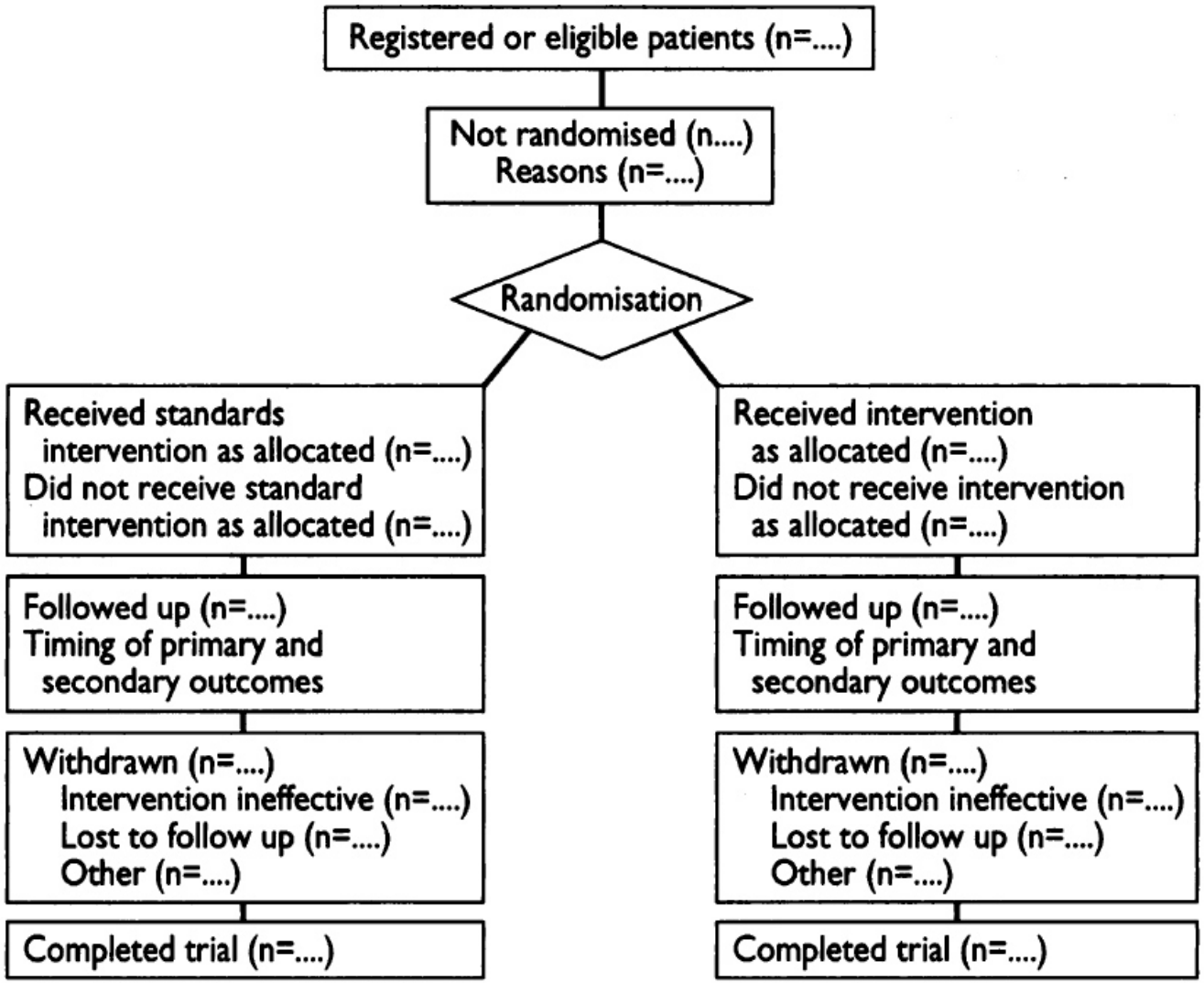
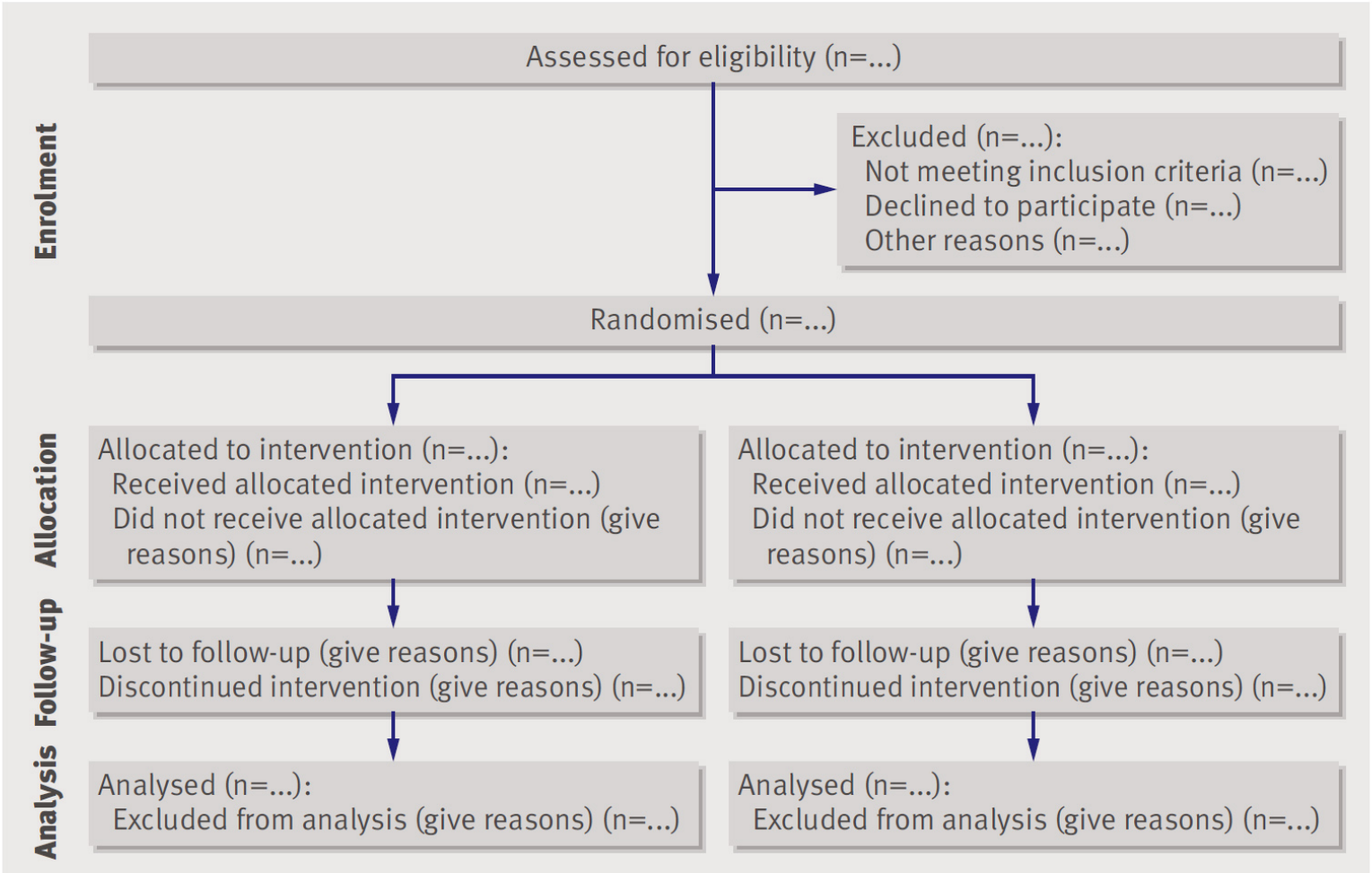


Fig 1—*Flow chart describing progress of patients through randomised trial (reproduced from JAMA)*[9]

Besides statistics, documentation: CONSORT



Flow diagram of the progress through the phases of a parallel randomised trial of two groups (that is, enrolment, intervention allocation, follow-up, and data analysis)

CONSORT-style example from QJE

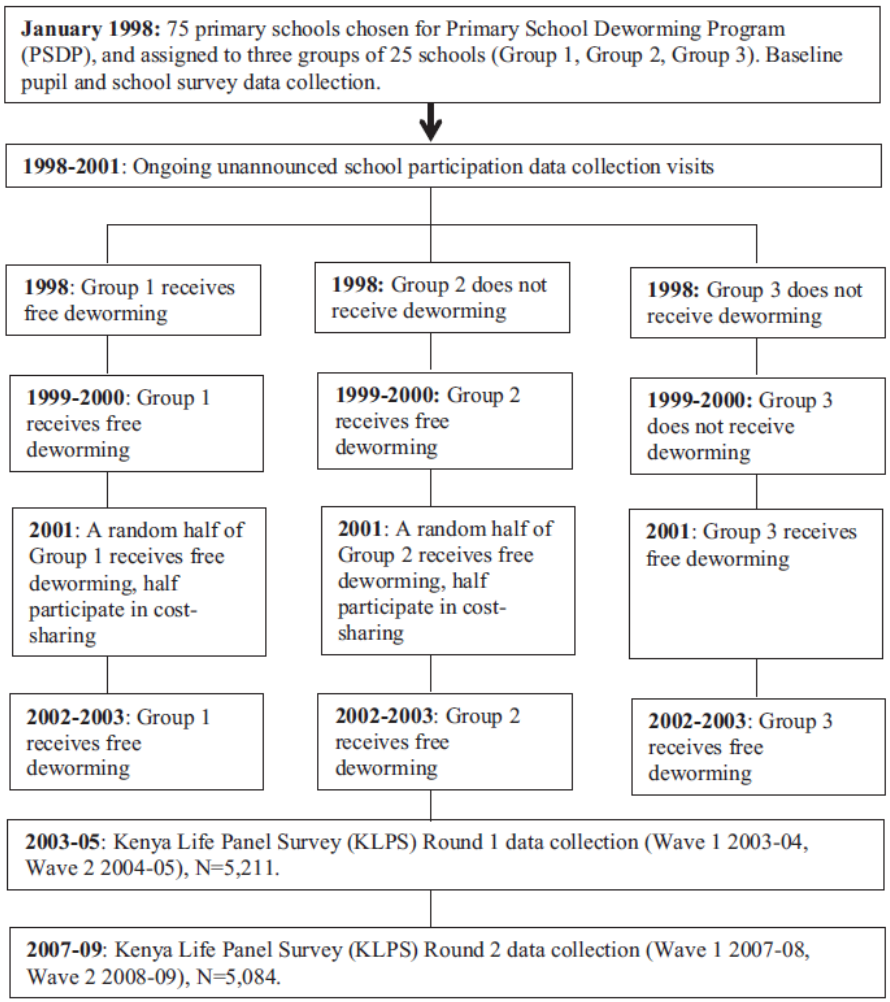


FIGURE I

Project Timeline of the Primary School Deworming Program (PSDP) and the Kenya Life Panel Survey (KLPS)

CONSORT-style example from ECRQ

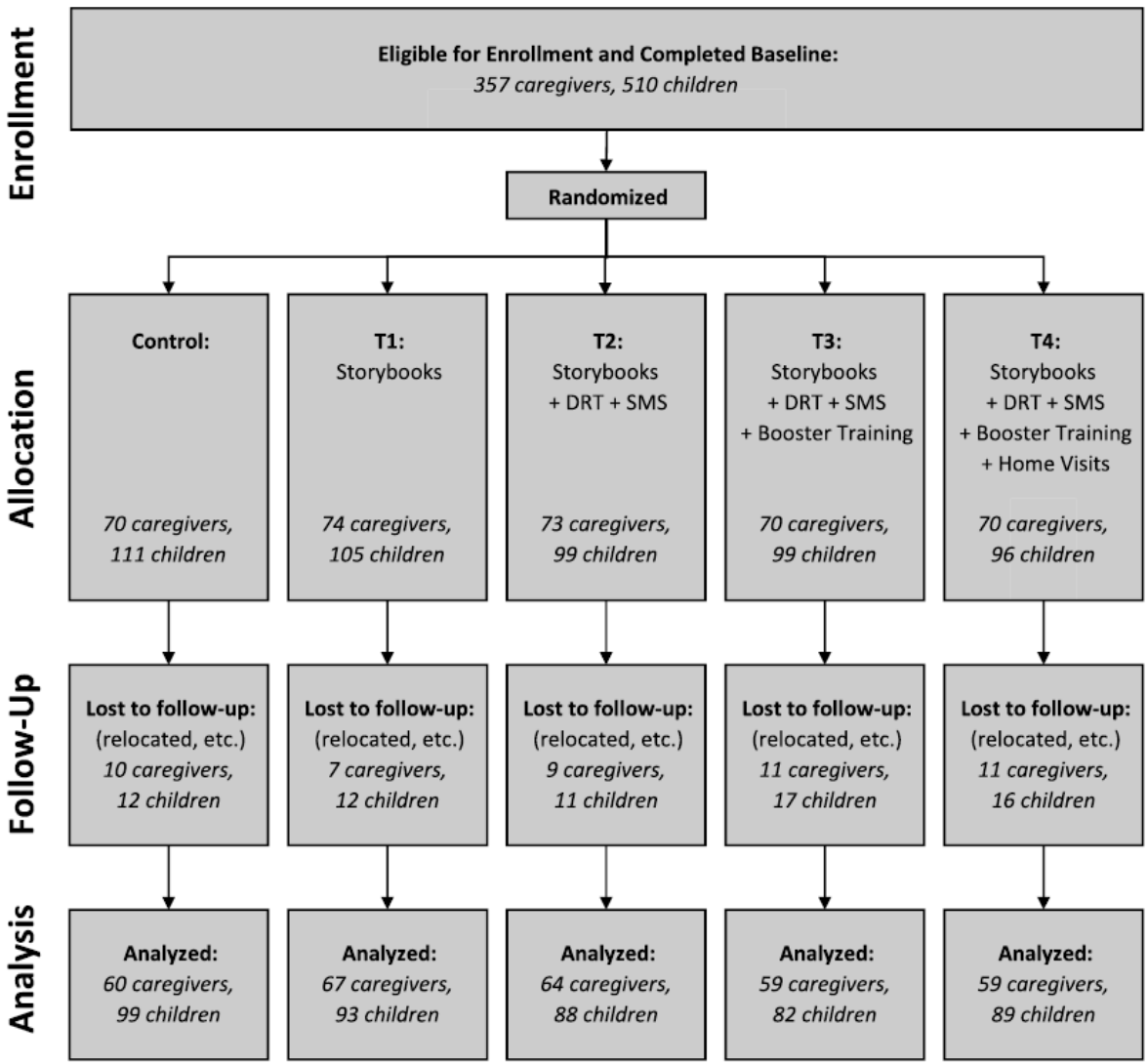


Fig. 1. CONSORT flow diagram.

Besides statistics, documentation: CONSORT

Table 1—*Items that should be included in reports of randomised trials (reproduced from* JAMA*)*[9]

Heading	Subheading	Descriptor
Title		Identify the study as a randomised trial
Abstract		Use a structured format
Introduction		State prospectively defined hypothesis, clinical objectives, and planned subgroup or covariate analyses
Methods	Protocol	Describe Planned study population, together with inclusion or exclusion criteria Planned interventions and their timing Primary and secondary outcome measure(s) and the minimum important difference(s), and indicate how the target sample size was projected Rationale and methods for statistical analyses, detailing main comparative analyses and whether they were completed on an intention to treat basis Prospectively defined stopping rules (if warranted)
	Assignment	Describe Unit of randomisation (for example, individual, cluster, geographic) Method used to generate the allocation schedule Method of allocation concealment and timing of assignment Method to separate the generator from the executor of assignment
	Masking (blinding)	Describe Mechanism (for example, capsules, tablets) Similarity of treatment characteristics (for example, appearance, taste) Allocation schedule control (location of code during trial and when broken) Evidence for successful blinding among participants, person doing intervention, outcome assessors, and data analysts
Results	Participant flow and follow up	Provide a trial profile (fig 1) summarising participant flow, numbers and timing of randomisation assignment, interventions, and measurements for each randomised group
	Analysis	State estimated effect of intervention on primary and secondary outcome measures, including a point estimate and measure of precision (confidence interval) State results in absolute numbers when feasible (for example, 10/20, not 50%) Present summary data and appropriate descriptive and interferential statistics in sufficient detail to permit alternative analyses and replication Describe prognostic variables by treatment group and any attempt to adjust for them Describe protocol deviations from the study as planned, together with the reasons
Discussion		State specific interpretation of study findings, including sources of bias and imprecision (internal validity) and discussion of external validity, including appropriate quantitative measures when possible State general interpretation of the data in light of the totality of the available evidence

Besides statistics, documentation: CONSORT

Table 1—*Items that should be included in reports of randomised trials (reproduced from JAMA)*[9]

Heading	Descriptor
Title	Identify the study as a randomised trial
Abstract	Use a structured format
Introduction	State prospectively defined hypothesis, clinical objectives, and planned subgroup or covariate analyses

Some of these are clearly more applicable to economics than others.

Besides statistics, documentation: CONSORT

Methods	Protocol	Describe Planned study population, together with inclusion or exclusion criteria Planned interventions and their timing Primary and secondary outcome measure(s) and the minimum important difference(s), and indicate how the target sample size was projected Rationale and methods for statistical analyses, detailing main comparative analyses and whether they were completed on an intention to treat basis Prospectively defined stopping rules (if warranted)
	Assignment	Describe Unit of randomisation (for example, individual, cluster, geographic) Method used to generate the allocation schedule Method of allocation concealment and timing of assignment Method to separate the generator from the executor of assignment
	Masking (blinding)	Describe Mechanism (for example, capsules, tablets) Similarity of treatment characteristics (for example, appearance, taste) Allocation schedule control (location of code during trial and when broken) Evidence for successful blinding among participants, person doing intervention, outcome assessors, and data analysts

Besides statistics, documentation: CONSORT

Results	Provide a trial profile (fig 1) summarising participant flow, numbers and timing of randomisation assignment, interventions, and measurements for each randomised group
	State estimated effect of intervention on primary and secondary outcome measures, including a point estimate and measure of precision (confidence interval)
	State results in absolute numbers when feasible (for example, 10/20, not 50%)
	Present summary data and appropriate descriptive and interferential statistics in sufficient detail to permit alternative analyses and replication
	Describe prognostic variables by treatment group and any attempt to adjust for them
	Describe protocol deviations from the study as planned, together with the reasons
Discussion	State specific interpretation of study findings, including sources of bias and imprecision (internal validity) and discussion of external validity, including appropriate quantitative measures when possible
	State general interpretation of the data in light of the totality of the available evidence

Bruhn and McKenzie - Approach

A survey of practitioners, then **six datasets**:

- Microenterprise profits in Sri Lanka
- Employment survey in Mexico
- Indonesia Family Life Survey: children in school
- Indonesia Family Life Survey: household expenditure
- Learning & Educational Achievement Project (Pakistan): math test
- Learning & Educational Achievement Project: height z-score

Bruhn and McKenzie - Approach

Then, five **randomization methods**:

- Randomization (single random draw)
- Stratification
- Pair-wise matching
- Rerandomization: redraw if anything is significant
- Rerandomization: minimum maximum t statistic

Bruhn and McKenzie - Approach

Really important: **choosing the variables**.

> "The set of outcomes we have chosen spans a range of the ability of the baseline variables to predict future outcomes. At one end is microenterprise profits in Sri Lanka, where baseline profits and 6 baseline individual and firm characteristics explain only 12.2 percent of the variation in profits 6 months later. ... The math test scores and height z-scores in the LEAPS data have the most variation explained by baseline characteristics, with 43.6 percent of the variation in follow-up test scores explained by the baseline test score and 6 baseline characteristics."

Bruhn and McKenzie - Recommendation 1

"Better reporting of the **method of random assignment** is needed. This should include a description of:
a. Which randomization method was used and why.
b. Which variables were used for balancing?
c. For stratification, how many strata were used?
d. For rerandomization, which cutoff rules were used?
This is particularly important for experiments with small samples, where the randomization method makes more difference."

(Obvious in retrospect?)

Bruhn and McKenzie - Recommendation 2

"Clearly describe **how the randomization was carried out** in practice.
a. Who performed the randomization?
b. How was the randomization done (coin toss, random number generator, etc.)?
c. Was the randomization carried out in public or private?"

Bruhn and McKenzie - Recommendation 3

"**Re-think the common use of rerandomization.**
Our simulations find pair-wise matching to generally perform as well, or better, than rerandomization in terms of balance and power, and like rerandomization, matching allows balance to be sought on more variables than possible under stratification. Adjusting for the method of randomization is statistically cleaner with matching or stratification than with rerandomization."

Bruhn and McKenzie - Recommendation 4

"When deciding which variables to balance on, strongly consider the **baseline outcome variable and geographic region dummies, in addition to variables desired for subgroup analysis.**"

Bruhn and McKenzie - Recommendation 5

"**Be aware that over-stratification can lead to a loss of power** in extreme cases. This is because using a large number of strata involves a downside in terms of loss in degrees of freedom when estimating standard errors, possibly more cases of missing observations, and odd numbers within strata when stratification is used."

Bruhn and McKenzie - Recommendation 6

"**As ye randomize, so shall ye analyze.**" (Include dummies for strata in analysis.) "Similarly, pair dummies should be included for matched randomization, or linear variables used for rerandomizations."

Bruhn and McKenzie - Recommendation 7

"In the ex post analysis, **do not automatically control for baseline variables that show a statistically significant difference in means.** The previous literature, and our simulations, suggest that it is a better rule to control for variables that are thought to influence follow-up outcomes, independent of whether their difference in means is statistically significant or not. ... One should still be cautious in the use of ex post controls, given the potential for finite-sample bias if treatment heterogeneity is correlated with the square of these covariates."

McKenzie (2012)

"The vast majority of randomized experiments in economics rely on a single baseline and single follow-up survey. While such a design is suitable for study of highly autocorrelated and relatively precisely measured outcomes in the health and education domains, it is unlikely to be optimal for measuring noisy and relatively less autocorrelated outcomes such as business profits, and household incomes and expenditures. Taking multiple measurements of such outcomes at relatively short intervals allows one to average out noise, increasing power. When the outcomes have low autocorrelation and budget is limited, it can make sense to do no baseline at all. Moreover, I show how for such outcomes, more power can be achieved with multiple follow-ups than allocating the same total sample size over a single follow-up and baseline. I also highlight the large gains in power from ANCOVA analysis rather than difference-in-differences analysis when autocorrelations are low."