Journal of Behavioral Medicine (2026) 49:298–306
https://doi.org/10.1007/s10865-025-00595-6

Determining the power of a 1-sided *z*-test given only the power of the corresponding 2-sided test

Amy Liang[1] · Kristopher J. Preacher[1] · Nathaniel J. Williams[2,3] · Paul D. Allison[4] · Steven C. Marcus[5,6] · Sonya K. Sterba[1]

Received: 15 November 2024 / Accepted: 17 July 2025
© The Author(s) 2025, modified publication 2026

Abstract
Estimating statistical power is essential for designing behavioral medicine studies efficiently and conserving finite resources. Sometimes behavioral medicine researchers are interested in calculating power for 1-sided *z*-tests of individual parameters (e.g., slopes) in complex models such as multilevel structural equation models or multilevel mixture regression models. For such models, calculating power for 1-sided *z*-tests is cumbersome because: (a) online *z*-test power calculator tools are inapplicable, (b) commonly-used power analysis software provides power only for 2-sided *z*-tests and does not allow changing alpha, and (c) published power tables typically provide power results only for 2-sided *z*-tests. Hence, here we introduce straightforward and resource-efficient conversion formulas to estimate the power of 1-sided *z*-tests of individual parameters in any model by using direct power conversions from the corresponding 2-sided tests. We then implement these conversion formulas in accessible R and Excel software. This brief report thus provides behavioral medicine researchers with a convenient and practical solution for power calculation that minimizes the time, financial, and computational resources typically needed for power estimation.

Keywords Statistical power · Directional and non-directional hypotheses · Hypothesis testing · Power conversion

The *power* of a statistical test is the probability of rejecting a false null hypothesis (H_0) given that some alternative hypothesis (H_A) is true. Power estimation has been widely recognized as crucial in behavioral medicine research, as it enables researchers to conserve finite research resources, avoid wasting participants' time, and minimize the chance of failing to detect a true effect (e.g., Davidson et al., 2003). In conducting power analyses, researchers have the option to adopt either non-directional (2-sided) or directional (1-sided) hypotheses.

Directional hypotheses

Across scientific disciplines, traditional 2-sided tests of non-directional hypotheses are commonly employed, especially in exploratory settings where researchers do not want to hinder discovery of effects in the unexpected direction. An example *non-directional hypothesis* posited by a behavioral medicine researcher is: child cerebral palsy patients treated with myofascial-release techniques plus functional training have *no difference* in gross motor skills compared

✉ Kristopher J. Preacher
kris.preacher@vanderbilt.edu

1 Department of Psychology and Human Development, Vanderbilt University, 230 Appleton Place, Nashville, TN 37203-5721, USA

2 School of Social Work, Boise State University, Boise, ID, USA

3 Institute for the Study of Behavioral Health and Addiction, Boise State University, 1910 University Drive, Boise, ID 83725-1940, USA

4 Statistical Horizons LLC, P.O. Box 282, Ardmore, PA 19003, USA

5 Penn Center for Mental Health, University of Pennsylvania School of Medicine, 3535 Market Street, Philadelphia, PA 19104, USA

6 School of Social Policy and Practice, University of Pennsylvania, 3701 Locust Walk, Philadelphia, PA 19104-6214, USA

Published online: 09 September 2025

with those receiving only functional training. However, in reviews of over 490 published applications of structural equation modeling, regression, and ANOVA containing over 2700 stated hypotheses, over 90% of hypotheses stated were actually directional (Cho & Abe, 2012). An example *directional hypothesis* posited by a behavioral medicine researcher is: child cerebral palsy patients treated with myofascial-release techniques plus functional training will have *more improvement* in gross motor skills than those receiving only functional training (Rafat & Srivastav, 2024). When theory is strong enough to suggest an expected direction for an effect, it makes little sense to conduct a 2-sided significance test in which half of the rejection region lies in the opposite, unexpected direction; such a test would be inconsistent with the hypothesis.

Thus, 1-sided tests are a logically consistent choice for evaluating directional hypotheses. 1-sided tests can be useful when there is a reasonable expectation of an outcome occurring in a given direction based on theory or prior knowledge and when a result in the direction opposite of expectation would have little implications for theory or practice (e.g., Hick, 1952; Levitt, 1994; Rice & Gaines, 1994; Ruxton & Neuhäuser, 2010; Wike, 1971). For instance, 1-sided tests are used in placebo-controlled clinical trials to test directional hypotheses that a new drug (e.g., for insomnia) is superior to a placebo—where the new drug would not receive approval if it were less effective than the placebo (e.g., Murphy, 1998). As another example, 1-sided tests are used to test directional hypotheses that product exposure (e.g., BPA [bisphenol-A] in can linings) harms vulnerable groups such as children or pregnant women—necessitating placing a warning label on the product (e.g., Ruxton & Neuhäuser, 2010). No warning label would be affixed if the product had either a neutral or net beneficial effect.

Another reason behavioral medicine researchers might choose directional hypothesis tests is that they are more powerful than their non-directional counterparts and therefore require smaller sample sizes (e.g., Hales, 2024; Hernández et al., 2018; Lawler & Zimmermann, 2021). This is an especially important consideration when the costs of sampling are high, such as in highly pragmatic trials (Ford & Norrie, 2016; Loudon et al., 2015). Consider the simple situation in which a behavioral medicine researcher wishes to conduct a significance test for the effect (slope) of goal-striving (x) on chronic pain (y) in the context of simple linear regression. Slope estimators quickly approach normality with increasing n, justifying the use of an asymptotic z-test. Given a slope estimate for the effect of goal-striving on chronic pain of $\hat{b} = .16$ ($SE = .09$), the test statistic ($z = \hat{b}/SE = 1.78$) would be significant at $\alpha = .05$ for the 1-sided test (p=.0377) but nonsignificant for the 2-sided test (p=.0754). Of course, ethically the decision of whether to use a 1-sided test must be made in advance, not after inspecting 2-sided test results (Burke, 1953; Panter & Sterba, 2011). This can be accomplished during pre-registration of a study or trial (e.g., Bosnjak et al., 2022) wherein the researcher would specify in advance whether hypotheses are non-directional or directional and whether corresponding tests are one-sided or two-sided before analyzing the data (see Lawler & Zimmermann, 2021). This would avoid what Lawler and Zimmerman (2021) call "hypothesis hacking"—formulating or changing a hypothesis after statistical testing to correspond with the testing results.

Power for 1-sided tests to evaluate directional hypotheses

Thus, directional hypothesis tests may be more consistent with theory and always will be more powerful than non-directional tests. It follows that the power for directional tests can be of interest. Although some online calculators[1] often can compute power for 1 or 2-sided z-tests, such online calculators are designed for simple situations where one knows n and wants to compute power (e.g., a z-test for testing a single mean against a constant or testing the difference in two means). In those cases, n is explicitly included in the z-test formula. These online calculators *do not apply* to more complex cases where z-tests are used but n is not explicitly included in the formula– e.g., tests for slopes in models using maximum likelihood estimation, such as path coefficients in mixture regression models, multilevel structural equation models, multilevel mixture models, etc. that are increasingly of interest in behavioral medicine research (e.g., Baumann et al., 2020; Deng et al., 2022; de Vries McClintock et al., 2014; Kessels et al., 2021; Mun et al., 2016; Romano et al., 2020; St Fleur et al. 2025; Watanabe & Yamauchi, 2016; Wi et al., 2024). For example, in Mun et al. (2016) the slope of chronic pain on goal striving is one parameter embedded in a multilevel structural equation model; thus, power for a one-sided z-test of this slope cannot be obtained using such online calculators. In these more complex situations, researchers often (1) turn to power tables from published simulations or (2) conduct Monte Carlo simulations themselves using, for instance, M*plus* (Muthén & Muthén, 1998–2025), simsem (Pornprasertmanit et al., 2021), or WebPower (Zhang & Yuan, 2018)—see Liang and Sterba (in press) for a review of such programs.

Although many behavioral medicine researchers still prefer to use tabled power values instead of running Monte Carlo simulation software programs for power analyses,

[1] Examples of such online calculators include: https://www.statskingdom.com/32test_power_t_z.html and https://mathcracker.com/power-calculator-testing-one-mean.

unfortunately the majority of tabled *z*-test power values provided in published simulation studies (see, e.g., Egbewale et al., 2014; Geminiani et al., 2021; Kaplan & George, 1995; MacKinnon et al., 2002; Preacher et al., 2011; Van Horn et al., 2008) are only for non-directional (2-sided) tests, rather than for directional (1-sided) tests. For researchers who instead opt to use Monte Carlo simulation software programs to compute power for *z*-tests from complex models, unfortunately some frequently-used software programs for this purpose (e.g., M*plus;* Muthén & Muthén, 1998–2025) generate statistical power only for 2-sided, rather than for 1-sided, *z*-tests and do not permit doubling alpha in order to obtain power for 1-sided *z*-tests. Behavioral medicine researchers are thus in need of a method to quickly convert a 2-sided test's power to a 1-sided test's power.

To address this limitation, in this brief report we provide behavioral medicine researchers with convenient formulas that facilitate the conversion from power for non-directional (2-sided) *z*-tests to power for directional (1-sided) *z*-tests, when only non-directional (2-sided) power values are available. Our conversion formulas are applicable under the same outcome-normality assumptions invoked when determining power for *z*-tests using published power tables mentioned earlier or using typical implementations of Monte Carlo power analyses (Liang & Sterba, in press), that don't explicitly involve generating nonnormal data. We also implement these conversion formulas using widely-used R and Excel software. Readers wishing to bypass the derivation can skip to the software implementation section.

Table 1 Symbols and their meanings

Symbols	Meaning
Φ	Standard normal cdf,[a] which converts *z*-scores into *p*-values
Φ^{-1}	Inverse standard normal cdf, which converts *p*-values into *z*-scores
α	Nominal Type I error rate
n	Total sample size
θ	Proposed/theorized effect[b]
$\hat{\theta}$	Estimated/predicted effect
θ_0	Hypothesized effect under H_0
θ_{c1}	Critical value under H_0 for the 1-sided test
θ_{c2+}	Right-side critical value under H_0 for the 2-sided test
θ_{c2-}	Left-side critical value under H_0 for the 2-sided test
σ	Standard error of $\hat{\theta}$
z_{N1}	*z*-score associated with θ_{c1} under H_0 for the 1-sided test
z_{N2+}	*z*-score associated with θ_{c2+} under H_0 for the 2-sided test
z_{N2-}	*z*-score associated with θ_{c2-} under H_0 for the 2-sided test
z_{A1}	*z*-score associated with θ_{c1} under H_1 for the 1-sided test
z_{A2+}	*z*-score associated with θ_{c2+} under H_1 for the 2-sided test
z_{A2-}	*z*-score associated with θ_{c2-} under H_1 for the 2-sided test
π_1	Power for the 1-sided test
π_2	Power for the 2-sided test

[a]cdf, cumulative distribution function

[b]The term "effect" is used here to refer to any parameter that reflects a hypothesis of interest. In this short report, our discussion pertains to testing parameters for which *z*-tests are appropriate (e.g., slopes, intercepts, and mean differences)

Power of a 1-sided *z*-test

Please refer to Table 1 for a comprehensive reference list of all symbols used in equations, and their definitions. Although the following presentation assumes the alternative hypothesis $H_1 : \theta > \theta_0$—i.e., that the proposed effect θ is greater than its null-hypothesized value θ_0—analogous logic applies when $\theta < \theta_0$ such that, if indeed $H_1 : \theta < \theta_0$, our final formula (i.e., Eq. 15) still holds, as we describe later.

Under the alternative hypothesis $H_1 : \theta > \theta_0$, in a 1-sided *z*-test, the critical value of *z* under H_0 at a specified significance level, denoted z_{N1}, can be visualized in Fig. 1 Panel A and is defined as:

$$z_{N1} = \Phi^{-1}(1-\alpha) = \frac{\theta_{c1} - \theta_0}{\sigma} \tag{1}$$

where θ_{c1} is the corresponding critical value of the parameter being tested. Thus,

$$\theta_{c1} = \sigma\Phi^{-1}(1-\alpha) + \theta_0 \tag{2}$$

The value of *z* for the same θ_{c1} under H_1 (i.e., that θ is the proposed effect) is denoted z_{A1}. This z_{A1} is represented in Fig. 1 Panel B and is expressed as

$$z_{A1} = \frac{\theta_{c1} - \theta}{\sigma} = \frac{\sigma\Phi^{-1}(1-\alpha) + \theta_0 - \theta}{\sigma} = \Phi^{-1}(1-\alpha) + \frac{\theta_0}{\sigma} - \frac{\theta}{\sigma} \tag{3}$$

Thus, a quantity that will be used later in deriving the conversion formula is

$$\frac{\theta}{\sigma} = \Phi^{-1}(1-\alpha) + \frac{\theta_0}{\sigma} - z_{A1} \tag{4}$$

The probability of rejecting H_0 when H_1 is true—i.e., power for the 1-sided *z*-test—is denoted π_1:

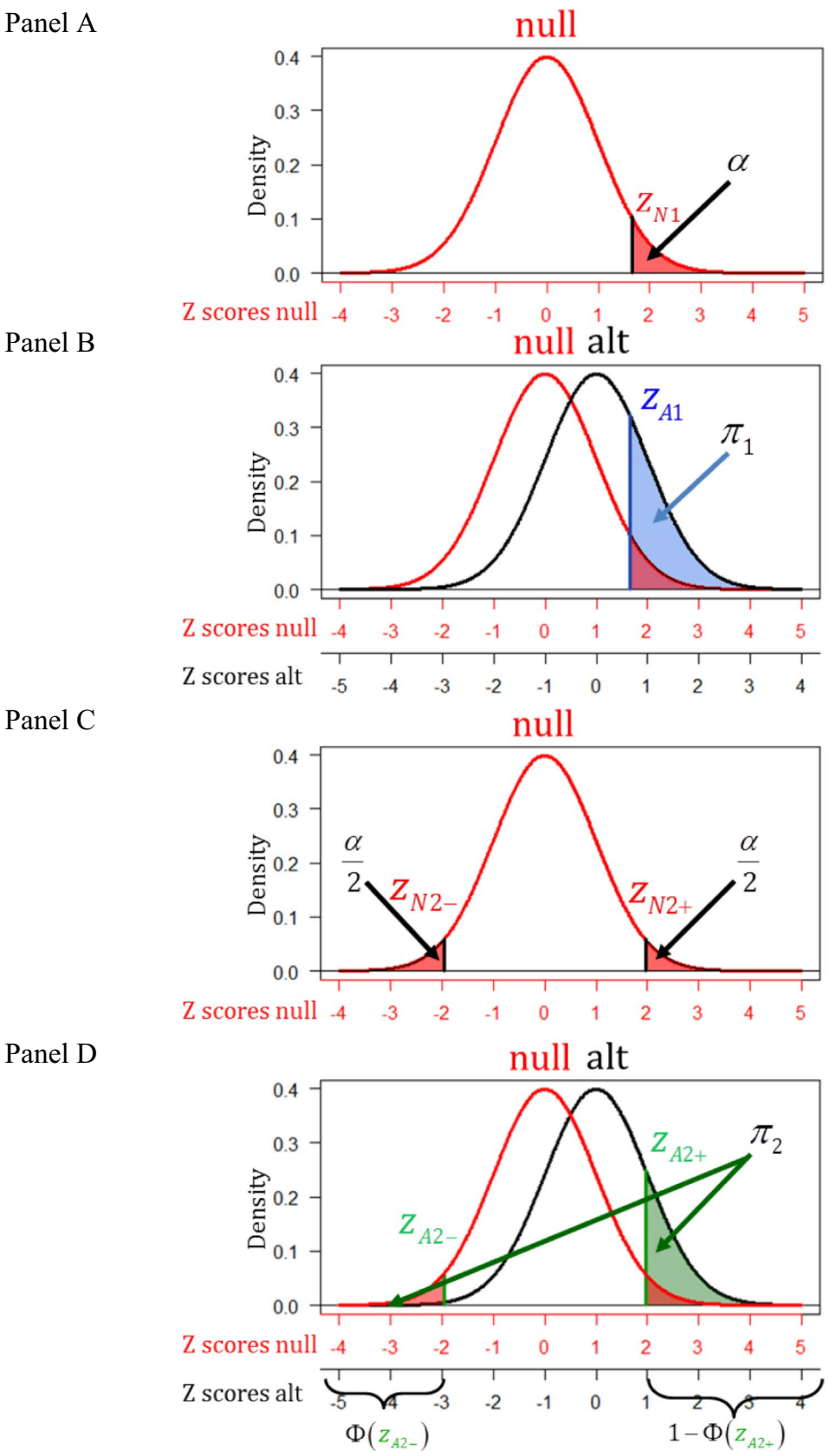


Fig. 1 Diagrams to visualize 1-sided power (π_1) and 2-sided power (π_2) of a z-test, where H_1: $\theta > \theta_0$

$$\pi_1 = 1 - \beta_1 = 1 - \Phi\left(z_{A1}\right) \tag{5}$$

π_1 can be visualized in Fig. 1 Panel B.

Power of a 2-sided *z*-test

Unlike 1-sided tests, a 2-sided *z*-test involves both left-side (θ_{c2-}) and right-side (θ_{c2+}) critical values of the parameter being tested, under H_0. The right-side critical value of *z* under H_0, denoted z_{N2+} is depicted in Fig. 1 Panel C and is defined as:

$$z_{N2+} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = \frac{\theta_{c2+} - \theta_0}{\sigma} \tag{6}$$

Thus, the right-side critical value of the parameter being tested, θ_{c2+}, is:

$$\theta_{c2+} = \sigma\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + \theta_0 \tag{7}$$

The value of *z* for the same θ_{c2+} under H_1, denoted z_{A2+}, is illustrated in Fig. 1 Panel D and can be expressed as:

$$\begin{aligned} z_{A2+} &= \frac{\theta_{c2+} - \theta}{\sigma} = \frac{\sigma\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + \theta_0 - \theta}{\sigma} \\ &= \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + \frac{\theta_0}{\sigma} - \frac{\theta}{\sigma} \end{aligned} \tag{8}$$

Thus, a quantity to be used later when deriving the conversion formula is:

$$\frac{\theta}{\sigma} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + \frac{\theta_0}{\sigma} - z_{A2+} \tag{9}$$

The power for the 2-sided *z*-test, denoted π_2, is thus given in Eq. (10) and is illustrated in Fig. 1 Panel D:

$$\pi_2 = 1 - \beta_2 = 1 - \Phi\left(z_{A2+}\right) + \Phi\left(z_{A2-}\right) \tag{10}$$

where z_{A2-} is the value of *z* corresponding to θ_{c2-} under H_1. Under the assumption of normality necessary for applying *z*-tests, and when[2] $H_1 : \theta > \theta_0$, the $\Phi\left(z_{A2-}\right)$ becomes negligible (as demonstrated later, and as illustrated in Fig. 1 Panel D), adding only trivially to power.

[2] Our derived conversion formula in Eq. (15) holds whether $H_1 : \theta > \theta_0$ or $H_1 : \theta < \theta_0$, as described later; for simplicity, our derivations are presented under the assumption $H_1 : \theta > \theta_0$.

Derivation of a formula converting power for a 2-sided *z*-test to power for a 1-sided *z*-test

Setting both of the above expressions for $\frac{\theta}{\sigma}$, in Eqs. (4) and (9), to be equal, we can then derive the *z*-score of the critical value θ_{c1} under H_1, denoted z_{A1}, in terms of z_{A2+}:

$$\begin{aligned} &\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + \frac{\theta_0}{\sigma} - z_{A2+} = \Phi^{-1}\left(1 - \alpha\right) + \frac{\theta_0}{\sigma} - z_{A1} \\ &z_{A1} = \Phi^{-1}\left(1 - \alpha\right) - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + z_{A2+} \end{aligned} \tag{11}$$

Now we can take our earlier Eq. (5) expression for 1-sided power π_1 and substitute Eq. (11) for z_{A1}:

$$\begin{aligned} \pi_1 &= 1 - \beta_1 = 1 - \Phi\left(z_{A1}\right) \\ &= 1 - \Phi\left(\Phi^{-1}\left(1 - \alpha\right) - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + z_{A2+}\right) \end{aligned} \tag{12}$$

But we would like π_1 to be expressed in terms of two-sided power π_2 (rather than in terms of z_{A2+}), so we solve for z_{A2+} using an approximation further discussed in the next section:

$$\begin{aligned} &\pi_2 = 1 - \Phi\left(z_{A2+}\right) + \Phi\left(z_{A2-}\right) \\ &\Phi\left(z_{A2+}\right) = 1 - \pi_2 + \underbrace{\Phi\left(z_{A2-}\right)}_{\text{essentially } 0} \\ &\Phi\left(z_{A2+}\right) \approx 1 - \pi_2 \\ &\Phi^{-1}\left(\Phi\left(z_{A2+}\right)\right) \approx \Phi^{-1}\left(1 - \pi_2\right) \\ &z_{A2+} \approx \Phi^{-1}\left(1 - \pi_2\right) \end{aligned} \tag{13}$$

Finally, substituting Eq. (13) into Eq. (12) yields the following approximation formula for converting power for a 2-tailed *z*-test, π_2, to power for the 1-tailed *z*-test, π_1:

$$\pi_1 \approx 1 - \Phi\left(\Phi^{-1}\left(1 - \alpha\right) - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + \Phi^{-1}\left(1 - \pi_2\right)\right) \tag{14}$$

A mathematically equivalent, but slightly briefer, formula is given in Eq. (15):

$$\pi_1 \approx \Phi\left(\Phi^{-1}\left(\alpha\right) - \Phi^{-1}\left(\frac{\alpha}{2}\right) + \Phi^{-1}\left(\pi_2\right)\right) \tag{15}$$

In other words, for a given significance level α and a power estimate π_2 from a 2-sided *z*-test, we can use Eq. (15) to approximate the power of the corresponding 1-sided *z*-test π_1.

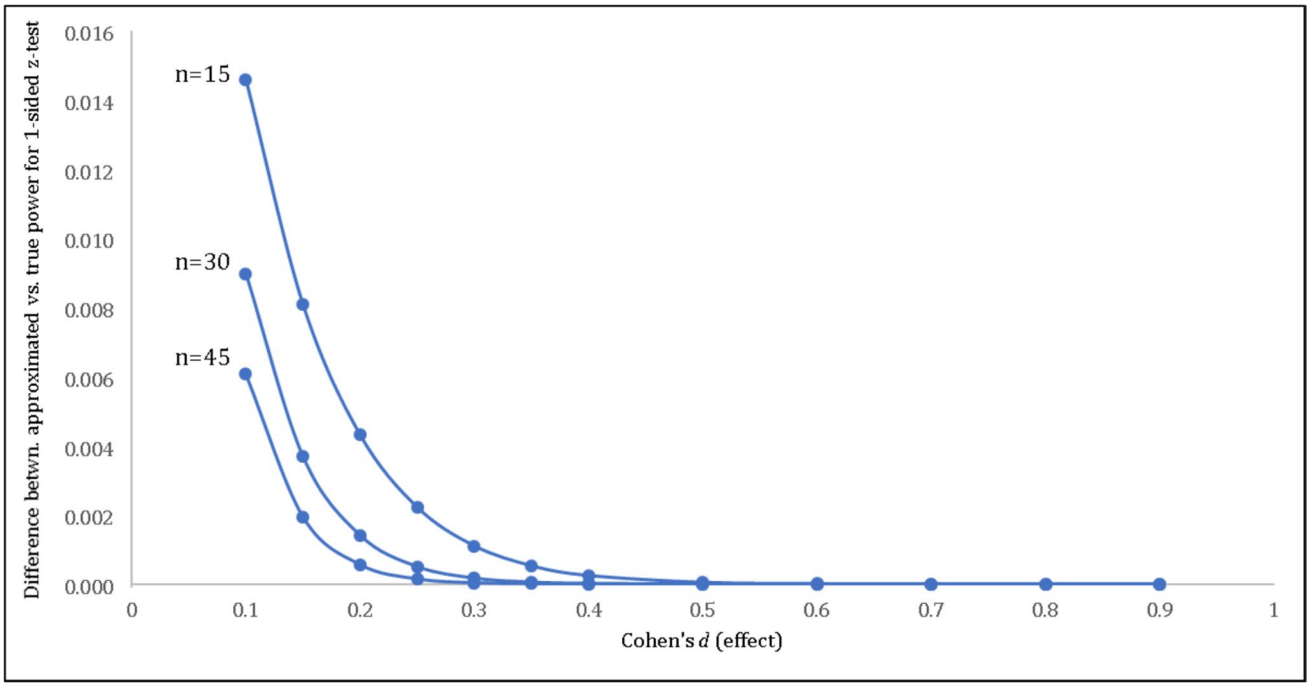


Fig. 2 Relationship between approximated (using conversion formula) vs. true power for a 1-sided *z*-test, across sample sizes and target effects at α=.05

Adequacy of the approximation used in the conversion formula

As a consequence of ignoring the trivial contribution of $\Phi(z_{A2-})$ (i.e., the tiny probability under the opposing side of the alternative distribution from the 2-sided test) in Eq. (13), our estimate of $\Phi(z_{A2+})$ and z_{A2+} in Eq. (13) will be negligibly decreased. This implies that, in the approximation formula for 1-tailed power π_1 (Eq. (14)), the argument to the outer Φ will be negligibly decreased, so our estimate of π_1 will be negligibly inflated, to a degree we here show is trivial—under the conventional normality assumption necessary for the use of *z*-tests. Figure 2 shows that, assuming normality, this inflation is inconsequential as it alters estimates for 1-tailed power π_1 only trivially (inflation of ~.01, on the vertical axis) for small target effects on the horizontal axis (such as Cohen's *d* of .10) and—for more typically-conjectured effect sizes—alters power estimates by ~.00, on the vertical axis, for modest to large effects on the horizontal axis (such as Cohen's $d \geq .20$) across sample size (*n*), at α=.05. One may consider how Fig. 2 would look for a different α. Values of $\alpha > .05$ are uncommonly used and $\alpha < .05$ would result in even smaller $\Phi(z_{A2-})$ and thus even smaller amounts of inflation. If an unusually large α=.10 were used—which is uncommon—the inflation would still be ~.00 for Cohen's $d \geq .25$ but for Cohen's d=.10 would range from.029 for n=15 to.013 for n=45.

What if the conventional normality assumption that underlies the use of *z*-testing is markedly violated? Then it may be useful to consider alternatives to *z*-tests entirely, avoid consulting *z*-test power tables generated under normality assumptions, and avoid converting 2-tailed power values—obtained assuming normality—to 1-tailed power values that also assume normality. Nonnormality stemming from multimodality, outliers, skewness, etc. each could have different impacts on accuracy of power calculations, and such patterns could be investigated via a simulation study like the following. First, a researcher could simulate hundreds or thousands of samples from a population consistent with the null hypothesis, with the anticipated kind/degree of nonnormality in the outcome. Each simulated sample could be used to produce a test statistic—cumulatively forming an empirical null distribution. For, say, α=.05, a right-side empirical critical value $\hat{z}^{\bullet}_{N1}$—where the superscript • denotes that a normal distribution is no longer assumed—can be determined by identifying the test statistic value at the 95th percentile of this empirical null distribution. Second, the researcher could simulate thousands of samples consistent with the alternative hypothesis—again with the anticipated kind/degree of outcome nonnormality, to produce an empirical distribution of test statistics consistent with that alternative hypothesis. Third, the researcher then could calculate the area beyond $\hat{z}^{\bullet}_{N1}$ under that empirical alternative distribution to empirically determine 1-tailed power without having to assume normality (see also Table 4

Approach 3 from Liang & Sterba, in press). Fourth, the latter result could then be compared to 1-tailed power obtained using our conversion formula under conventional normality assumptions—in order to ascertain the impact of nonnormality on power estimates obtained using our conversion formula.

Software implementation of the conversion formula and illustrative example

The power conversion described in this report can be accomplished using simple commands in widely available software, such as R and Excel, as shown here.

Implementation in Excel. For example, in an empty Microsoft Excel spreadsheet, one could insert (i.e., enter) α in cell B1 of the spreadsheet, and insert (i.e., enter) the power of the 2-sided test (π_2) in cell B2 of the spreadsheet. In a third cell of the same Excel spreadsheet the researcher could then copy and paste the following code—which refers to cell B1 (where the researcher has already inserted their desired α) and to cell B2 (where the researcher has already inserted their π_2)—in order to obtain the approximate power of the 1-sided *z*-test (π_1):

```
=NORM.S.DIST(NORM.S.INV(B1)
-NORM.S.INV(B1/2)+NORM.S.INV(B2),TRUE)
```

Simplifying further, if we can assume $\alpha = .05$, the right-side critical value of *z* for a 2-sided *z*-test is 1.95996, and the critical value for a 1-sided *z*-test is 1.64485, then the researcher could instead paste into the same Excel spreadsheet the following simplified equation code:

```
=NORM.S.DIST(0.31511+NORM.S.INV(B2),
TRUE)
```

which requires only π_2 as input (where again π_2 is 2-tailed power that the researcher previously inserted into cell B2 of the same Excel spreadsheet).

Implementation in R. Similarly, π_1 can be readily computed in R software by first defining alpha as the significance level (e.g., if the desired level were.05 this could be done using R code: `alpha=.05)` and then defining `two.tail.pi` in R as the input π_2, i.e., the 2-tailed power. Next, the below code could be copied and pasted into R:

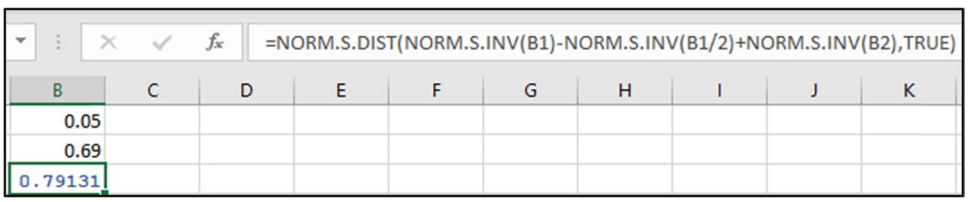
```
pnorm(qnorm(alpha)
-qnorm(alpha/2)+qnorm(two.tail.pi))
```

as will be demonstrated with a screen shot in the next section. Importantly, the approximation in Eq. 15 and the Excel and R implementations work both when H_1: $\theta > \theta_0$ *or* when H_1: $\theta < \theta_0$.

Empirical example. Next we consider an illustrative application of the conversion formula. Recently, Geminiani et al. (2021) published a Monte Carlo simulation that provided a power table for 2-sided *z*-tests to detect across-group differences in factor loadings in a multilevel factor mixture model. In one design cell, an absolute difference between loadings of ~.30 was detectable with.69 power at $\alpha = .05$ using a 2-tailed *z*-test. Yet, a future behavioral medicine researcher designing a study may want to know 1-tailed power under this condition. Because the published table provided only 2-tailed power, the future behavioral medicine researcher would be faced with running a computationally-intensive simulation to determine 1-tailed power for this test. Instead, in this situation 1-tailed power π_1 for the *z*-test can be quickly calculated by that researcher to be.79 by implementing our conversion formula using the above R or Excel code. In the below screen shot, the three lines of R input are followed by the output, on the fourth line.

```
> alpha=.05
> two.tail.pi=.69
> pnorm(qnorm(alpha)-qnorm(alpha/2)+qnorm(two.tail.pi))
[1] 0.7913059
```

Or, the same 1-tailed power for the *z*-test can be readily calculated using our conversion formula implemented in Excel as described above:



Discussion

In conclusion, power of 2-sided *z*-tests is readily available from published power tables or Monte Carlo power analysis software default output; however, this often is not the case for power of 1-sided *z*-tests. For situations where theory and study goals lead to an interest in power for 1-tailed tests (e.g., Williams et al., 2022), this study introduces a direct and easy method for behavioral medicine researchers to approximate power for 1-sided *z*-tests given an estimate of the power of a 2-sided test.

Our approximation method relies on the same conventional normality assumptions underlying *z*-tests. As mentioned earlier, our approximation method can be less accurate for the combination of very low $n \leq 15$, very small effect size $<.10$, and unusually large $\alpha \geq .10$; sample sizes this small combined with effect sizes this small and α levels this large are less common in practice.

Our approximation method can allow behavioral medicine researchers to efficiently evaluate directional hypothesis tests without incurring costs in terms of time, money, and computational resources. Although this study focuses on power computation in the context of *z*-tests, future research could investigate the feasibility of this strategy for other tests.

Funding This work was supported in part through a grant from the US National Institute of Mental Health (P50MH113840) to the University of Pennsylvania P50 ALACRITY Center for Transforming Mental Health Care Delivery through Behavioral Economics and Implementation Science. The views expressed herein are those of the authors and do not necessarily reflect those of NIMH.

Data availability N/A.

Code availability Included in manuscript.

Declarations

Conflicts of interest The authors hereby state and affirm that they have no competing interests, financial or otherwise, that might jeopardize the scientific integrity of this work.

Ethics approval N/A.

Consent to participate N/A.

Consent for publication N/A.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Baumann, S., Guertler, D., Weymar, F., Bahls, M., Dörr, M., van den Berg, N., John, U., & Ulbricht, S. (2020). Do accelerometer-based physical activity patterns differentially affect cardiorespiratory fitness? A growth mixture modeling approach. *Journal of Behavioral Medicine, 43*, 99–107. https://doi.org/10.1007/s10865-019-00069-6

Bosnjak, M., Fiebach, C. J., Mellor, D., Mueller, S., O'Connor, D. B., Oswald, F. L., & Sokol, R. I. (2022). A template for preregistration of quantitative research in psychology: Report of the joint psychological societies preregistration task force. *American Psychologist, 77*, 602–615. https://doi.org/10.1037/amp0000879

Burke, C. J. (1953). A brief note on one-tailed tests. *Psychological Bulletin, 50*(5), 384–387. https://doi.org/10.1037/h0059627

Cho, H., & Abe, S. (2012). Is two-tailed testing for directional research hypotheses tests legitimate? *Journal of Business Research, 66*, 1261–1266.

Davidson, K. W., Goldstein, M., Kaplan, R. M., Kaufmann, P. G., Knatterud, G. L., Orleans, C. T., Spring, B., Trudeau, K. J., & Whitlock, E. P. (2003). Evidence-based behavioral medicine: What is it and how do we achieve it? *Annals of Behavioral Medicine, 26*(3), 161–171. https://doi.org/10.1207/s15324796abm2603_01

de Vries McClintock, H. F., Morales, K. H., Small, D. S., & Bogner, H. R. (2014). Patterns of adherence to oral hypoglycemic agents and glucose control among primary care patients with Type 2 diabetes. *Behavioral Medicine, 42*, 63–71. https://doi.org/10.1080/08964289.2014.904767

Deng, Q., Wang, Y., & Liu, W. (2022). Using multilevel structural equation modeling (MSEM) to identify the predictors and influencing mechanism of technology use among Chinese physicians: An example from des-gamma-carboxy prothrombin (DCP). *Risk Management and Healthcare Policy, 15*, 59–70. https://doi.org/10.2147/RMHP.S344923

Egbewale, B. E., Lewis, M., & Sim, J. (2014). Bias, precision and statistical power of analysis of covariance in the analysis of randomized trials with baseline imbalance: A simulation study. *BMC Research Methodology, 14*, 49. https://doi.org/10.1186/1471-2288-14-49

Ford, I., & Norrie, J. (2016). Pragmatic trials. *New England Journal of Medicine, 375*(5), 454–463. https://doi.org/10.1056/NEJMra1510059

Gaito, J. (1977). Directional and nondirectional alternative hypotheses. *Bulletin of the Psychonomic Society, 9*, 371–372. https://doi.org/10.3758/BF03337026

Geminiani, E., Ceulemans, E., & De Roover, K. (2021). Testing for factor loading differences in mixture simultaneous factor analysis: A Monte Carlo simulation-based perspective. *Structural Equation Modeling, 28*, 391–409. https://doi.org/10.1080/10705511.2020.1807351

Hales, A. H. (2024). One-tailed tests: Let's do this (responsibly). *Psychological Methods*, *29*(6), 1209–1218. https://doi.org/10.1037/met0000610

Hernández, M. Á., Andrés, A. M., & Tejedor, I. H. (2018). One-tailed asymptotic inferences for the difference of proportions: Analysis of 97 methods of inference. *Journal of Biopharmaceutical Statistics, 28*(6), 1090–1104. https://doi.org/10.1080/10543406.2018.1452028

Hick, W. E. (1952). A note on one-tailed and two-tailed tests. *Psychological Review, 59*(4), 316–318. https://doi.org/10.1037/h0056061

Kaplan, D., & George, R. (1995). A study of the power associated with testing factor mean differences under violations of factorial invariance. *Structural Equation Modeling, 2*, 101–118. https://doi.org/10.1080/10705519509539999

Kessels, R., Moerbeek, M., Bloemers, J., & van der Heijden, P. G. M. (2021). A multilevel structural equation model for assessing a drug effect on a patient-reported outcome measure in on-demand medication data. *Biometrical Journal, 63*(8), 1652–1672. https://doi.org/10.1002/bimj.202100046

Kimmel, H. D. (1957). Three criteria for the use of one-tailed tests. *Psychological Bulletin, 54*(4), 351–353. https://doi.org/10.1037/h0046737

Lawler, I., & Zimmermann, G. (2021). Misalignment between research hypotheses and statistical hypotheses: A threat to evidence-based medicine? *Topoi, 40*, 307–318. https://doi.org/10.1007/s11245-019-09667-0

Levitt, E. E. (1994). The one-tailed test: A statistical editorial. *International Journal of Clinical and Experimental Hypnosis, 42*(1), 4–6. https://doi.org/10.1080/00207149408409336

Liang, A., & Sterba, S. K. (2024). Power analysis to detect misfit in SEMs with many items: Resolving unrecognized problems, relating old and new approaches, and "matching" power analysis approach to data analysis approach. *Psychological Methods*. https://doi.org/10.1037/met0000684

Loudon, K., Treweek, S., Sullivan, F., Donnan, P., Thorpe, K. E., & Zwarenstein, M. (2015). The PRECIS-2 tool: Designing trials that are fit for purpose. *British Medical Journal, 350*, Article h2147. https://doi.org/10.1136/bmj.h2147

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*, 83–104. https://doi.org/10.1037/1082-989x.7.1.83

Mun, C. J., Karoly, P., Okun, M. A., Kim, H., & Tennen, H. (2016). Affect, work-goal schemas, and work-goal striving among adults with chronic pain: A multilevel structural equation analysis. *Journal of Behavioral Medicine, 39*(2), 288–299. https://doi.org/10.1007/s10865-015-9696-4

Murphy, R. (2018). On the use of one-sided statistical tests in biomedical research. *Clinical and Experimental Pharmacology & Physiology, 45*(1), 109–114. https://doi.org/10.1111/1440-1681.12754

Muthén, L. K., & Muthén, B. O. (1998–2025). Mplus user's guide: Statisticalanalysis with latent variables (8th ed.). Los Angeles

Panter, A. T., & Sterba, S. K. (Eds.). (2011). *Handbook of ethics in quantitative methodology*. Multivariate applications series. Taylor & Francis/Routledge.

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling, 18*, 161–182. https://doi.org/10.1080/10705511.2011.557329

Pornprasertmanit, S., Miller, P.,Schoemann, A. M., & Jorgensen, T. D. (2025). simsem: SIMulated structural equation modeling (Rpackage version 0.5-17) [Computer software]. CRAN. Retrieved fromhttps://cran.r-project.org/web/packages/simsem/index.html

Rafat, S., & Srivastav, A. (2024). Developing hypotheses for scientific research in musculoskeletal conditions: A guide for health-care professionals. *Journal of Musculoskeletal Surgery and Research, 8*, 189–197. https://doi.org/10.25259/JMSR_117_2024

Rice, W. R., & Gaines, S. D. (1994). 'Heads I win, tails you lose': Testing directional alternative hypotheses in ecological and evolutionary research. *Trends in Ecology & Evolution, 9*(6), 235–237. https://doi.org/10.1016/0169-5347(94)90258-5

Romano, K. A., Stamates, A., Heron, K. E., Braitman, A. L., & Lau-Barraco, C. (2020). Sex and racial differences in patterns of disordered eating and alcohol use. *Behavioral Medicine, 47*(4), 272–284. https://doi.org/10.1080/08964289.2020.1748862

Ruxton, G. D., & Neuhäuser, M. (2010). When should we use one-tailed hypothesis testing? *Methods in Ecology and Evolution, 1*(2), 114–117. https://doi.org/10.1111/j.2041-210X.2010.00014.x

St Fleur, R. G., Ream, M., Walsh, E. A., & Antoni, M. H. (2025). Cognitive behavioral stress management affects different dimensions of benefit finding in breast cancer survivors: a multilevel mediation model. *Psychology & Health, 40*, 47-66. https://doi.org/10.1080/08870446.2023.2184840

Van Horn, M. L., Fagan, A. A., Jaki, T., Brown, E. C., Hawkins, J. D., Arthur, M. W., Abbott, R. D., & Catalano, R. F. (2008). Using multilevel mixtures to evaluate intervention effects in group randomized trials. *Multivariate Behavioral Research, 43*, 289–326. https://doi.org/10.1080/00273170802034893

Watanabe, M., & Yamauchi, K. (2016). Psychosocial factors of overtime work in relation to work-nonwork balance: A multilevel structural equation modeling analysis of nurses working in hospitals. *International Journal of Behavioral Medicine, 23*(4), 492–500. https://doi.org/10.1007/s12529-016-9563-x

Wi, D., Ransom, J. C., Flynn, D. M., Steffen, A. D., Park, C., Burke, L. A., & Doorenbos, A. Z. (2024). Role of pain catastrophizing in the effects of cognitive behavioral therapy for chronic pain in different subgroups: An exploratory secondary data analysis using finite mixture models. *Military Medicine, 189*, e2600–e2607. https://doi.org/10.1093/milmed/usae288

Wike, E. L. (1971). *Data analysis: A statistical primer for psychology students* (1st edn.). Routledge. https://doi.org/10.4324/9780203738938

Williams, N. J., Preacher, K. J., Allison, P. D., Mandell, D. S., & Marcus, S. C. (2022). Required sample size to detect mediation in 3-level implementation studies. *Implementation Science, 17*(1), 66. https://doi.org/10.1186/s13012-022-01235-2

Zhang, Z., & Yuan, K.-H. (2018). Practical statisticalpower analysis using WebPower and R (Eds.). Granger, IN: ISDSA Press

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.