Sample size estimation for biomechanical waveforms: Current practice, recommendations and a comparison to discrete power analysis

Mark A. Robinson[1*], Jos Vanrenterghem[2], and Todd C. Pataky[3]

[1]School of Sport and Exercise Sciences, Liverpool John Moores University, UK

[2]Musculoskeletal Rehabilitation Research Group, Faculty of Movement and Rehabilitation Sciences, KU Leuven, Belgium

[3]Department of Human Health Sciences, Kyoto University, Japan

April 13, 2021

*Corresponding Author: Mark Robinson, m.a.robinson@ljmu.ac.uk, +44 151 904 6267

Abstract

Testing a prediction is fundamental to scientific experiments. Where biomechanical experiments involve analysis of 1-Dimensional (waveform) data, sample size estimation should consider both 1D variance and hypothesised 1D effects. This study exemplifies 1D sample size estimation using typical biomechanical signals and contrasts this with 0D (discrete) power analysis. For context, biomechanics papers from 2018 and 2019 were reviewed to characterise current practice. Sample size estimation occurred in approximately 4% of 653 papers and reporting practice was mixed. To estimate sample sizes, common biomechanical signals were sourced from the literature and 1D effects were generated artificially using the open-source *power1d* software. Smooth Gaussian noise was added to the modelled 1D effect to numerically estimate the sample size required. Sample sizes estimated using 1D power procedures varied according to the characteristics of the dataset, requiring only small-to-moderate sample sizes of approximately 5-40 to achieve target powers of 0.8 for reported 1D effects, but were always larger than 0D sample sizes (from N+1 to >N+20). The importance of *a-priori* sample size estimation is highlighted and recommendations are provided to improve the consistency of reporting. This study should enable researchers to construct 1D biomechanical effects to address adequately powered, hypothesis-driven, predictive research questions.

Keywords: Statistical power, waveform analysis, hypothesis testing, statistical parametric mapping, numerical simulation

1

1 Introduction

Testing a prediction is the foundation of frequentist science. A prediction takes the form of a scientifically meaningul effect or relationship that is determined in advance of the experiment by the researchers. It is then the responsibility of the researchers to choose thresholds of type I (false positive) and type II error (false negative) they are willing to accept for their experiment and calculate the sample size required to avoid reaching false conclusions regarding treatment effectiveness. This ensures that the experiment has sufficient power to detect hypothesised effects, and is not over-powered and thereby detects inconsequentially small effects, or effects unrelated to the hypothesis.

Undertaking *a-priori* sample size estimation is a challenge for biomechanics researchers. Meaningful effect magnitudes are often unknown and measurement variability can be large which inflates the sample size required. Logistical reasons and time also mean that recruited sample sizes are often small (Knudson, 2017, Vagenas, Palaiothodorou, & Knudson, 2018). It is therefore perhaps understandable that sample size estimation is seen somewhat infrequently in the biomechanics literature (Knudson, 2017), as in similar fields (Abt et al., 2020; Lohse et al., 2020). Encouraging good practice and appropriate sample sizes however is needed irrespective of these challenges (Harrison et al., 2020, Lakens, 2021). The accurate reporting of statistical power and the data used to conduct a sample size estimation helps the reader to judge scientific quality and ensure reproducibility. The frequency with which sample sizes are reported and the characteristics of these calculations in recent biomechanics studies is not known.

Sample size estimation typically requires specialist software. Matlab, Minitab, G*Power and JASP all facilitate sample size estimation but all are limited to 0-dimensional (0D, discrete) data. Within biomechanics, hypotheses are often not specific to discrete data (Pataky, Robinson, & Vanrenterghem, 2013) and therefore 1-dimensional (1D) continua e.g. ground reaction forces, kinematics, kinetics, EMG, muscle forces are analysed (e.g. Robinson, Donnelly, Tsao, & Vanrenterghem, 2014; De Ridder et al., 2015; Pincheira, De La Maza, Silvestre, Guzmán-Venegas, & Becerra, 2019; Kipp, Comfort, & Suchomel, 2019). Suitable 1D analysis methods such as statistical parametric mapping or functional data analysis (Pataky, 2012, Ramsay & Silverman, 2007) are therefore required for 1D hypotheses to avoid experimental biases or false positive results (Pataky et al., 2013; Pataky, Vanrenterghem, & Robinson, 2016). Open source software, *power1d*, has been developed and applied to biomechanical scenarios to perform 1D sample size estimations based on simulated idealistic experimental effects (Pataky, 2017). Our knowledge of the biomechanics literature indicates *power1d* has been used in statistical (Serrien, Goossens, & Baeyens, 2019, Naouma & Pataky, 2019) but not yet in experimental contexts. We therefore believe that specifying 1D effects, sample size estimation for 1D effects, and comparison to 0D sample sizes requires further consideration and demonstration, enabling biomechanists to better justify sample sizes concomitant to their 1D data analyses.

For 1D sample size estimation the precise 1D effect of interest has to be specified *a priori*. This perhaps conflicts with the exploratory experiments often seen within biomechanics. Whilst exploratory experiments are necessary for understanding a new research topic or question, the null hypothesis is often "no effect", implying that no specific prediction is tested, and thus that no scientific or clinical meaning is imposed on the experiment in an *a priori* manner. Confirmatory research, involving specific, non-null predictions would be scientifically valuable, and complimentary to exploratory results. Sample size estimation from a 1D predicted effect is more complex than for 0D data (Pataky, Robinson, & Vanrenterghem, 2018). To encourage researchers to specify *a priori* 1D effects and conduct 1D power analysis this paper demonstrates how artificial 1D biomechanical effects of interest can be created, then uses numerical simulation to conduct sample size estimation, and contrasts these results to a 0D power calculation.

The aims of this paper are threefold: (1) to evaluate the use of sample size estimation in recent biomechanics studies, (2) to demonstrate how to estimate sample sizes for representative 1D biomechanical effects from the literature, (3) to contrast 1D sample size estimation with traditional 0D power analysis results. These aims should enable researchers to conduct 1D hypothesis tests for confirmatory hypotheses.

2 Methods

2.1 Power reporting practice

To provide context for the main aims of this study, and to justify current practice, an audit of research articles and short communications published in the Journal of Biomechanics in 2018 or 2019 was undertaken. Specifically the audit sought to determine the frequency of statistical power analysis and the quality of the reporting. The total number of publications matching the above criteria was 922 (figure 1). An initial search in Science Direct (https://sciencedirect.com) using the terms *"hypothesis OR ANOVA OR regression OR t test"* sought to separate experimental studies from other types e.g. methodological, and reduced the total articles to 653. The boolean *"AND power"* was added to the previous search to find experimental studies conducting power analysis. All articles returned in this search (n=226) were downloaded and stored locally as a full-text PDF. Articles were then full-text indexed using Adobe Acrobat Pro DC (v.20.009.20067). A subsequent full-text search was conducted using the boolean search *power NOT (power generation OR power absorption OR positive power OR negative power OR segmental power OR power law OR joint power)* which returned 84 articles. The 84 articles were then full-text screened manually to determine if *a-priori* power analysis was undertaken (figure 1). Articles that referred to statistical power only in the references section were excluded. Of these, 29 articles contained *a-priori* power analysis and characteristics of these papers were extracted

including the type II error rate, if a clear effect was specified and from existing literature, if the description was replicable, and the software used.

2.2 Sample size estimation

All analyses were conducted using *power1d* (v. 0.1.1, https://github.com/0todd0000/power1d, Pataky, 2017) in Python 3.7.6 using the Anaconda distribution (Anaconda Inc, https://anaconda.com).

2.2.1 Process overview

To conduct 1D power analysis using *power1d* a 1D baseline effect and an alternative effect are required, representing the null and alternative hypotheses, respectively. *Power1d* provides a variety of parameterised 1D geometries for flexible creation of the baseline and alternative 1D effects. Alternatively experimentally measured data (e.g. mean curves for two groups) can be used to specify these effects. The flexibility with which specific 1D features such as minima and maxima can be specified allows precise alternative (predicted) effects to be generated (figure 2). These 1D effects are combined with 1D random noise for which the noise type (e.g. *smooth Gaussian*), mean, standard deviation and smoothness are specified. A baseline (null) and alternative model can then be created to represent the experiment of interest (figure 3 a-c). Next the power of the experiment can be determined by simulating multiple experiments and generating the 1D test statistic of interest e.g. the t statistic, from which the maximum t statistic is stored for each one-tailed simulated experiment. Across multiple experiments these form a t distribution from which the 95th percentile of the null effect, the critical threshold, can be determined. The proportion of the alternative t distribution exceeding the critical threshold is then the omnibus power for the experiment. Finally, this process can be repeated for a range of sample sizes to determine the minimum sample size required to detect the experimental effect at a given power (figure 3 d-f). The specific requirements of *power1d* for this process are described elsewhere (https://spm1d.org/power1d) and supplemented by the availability of code from this paper. We also include a numerical sample size estimation for 0D data analogous to the above description as supplementary material.

2.2.2 Modelling 1D effects from the literature

Exemplar 1D effects from the biomechanics literature across a range of signal types were collated and reported (Bakke & Besier, 2020, Barrios & Willson, 2017, Bovi, Rabuffetti, Mazzoleni, & Ferrarin, 2011, Phan et al., 2017, Gomes, Ackermann, Ferreira, Orselli, & Sacco, 2017, Robinson et al., 2014). These effects then formed the basis of subsequent 1D and 0D sample size estimations. These studies were chosen because they reported a 1D effect that could represent a 1D biomechanical hypothesis, for example clear 1D means for two experimental conditions. Where the data were not directly available,

4

1D profiles were digitised using WebPlotDigitizer (v.4.3, https://automeris.io/WebPlotDigitizer), exported as text data then linearly interpolated to 101 data points. These curves therefore display an intended likeness rather than direct reproduction of the original results (figure 4). The specific experimental context of the 1D biomechanical effects was not considered further as the aim was to exemplify 1D effects rather than make specific biomechanical conclusions about these datasets. One 1D effect per dataset was assigned as the null (or baseline) signal.

2.2.3 Simulations to estimate sample sizes

To undertake sample size estimation we created 1D alternative effects and 1D noise curves to represent the 1D effects and variance from the literature thus making a 1D prediction for these artificial experiments. All 1D effects were created using the *Gaussian pulse* and *constant* waveform shapes from *power1d* (Pataky, 2017) and were 101 data points in length. As specific pulse locations and magnitudes were required for each alternative effect, the code to reproduce all 1D effects is provided at https://github.com/m-a-robinson/sample-size. Eight smooth Gaussian noise curves also 101 data points in length were generated for each dataset with the standard deviation parameter matched as closely as possible to the original dataset based on their reported data. These eight noise curves were then combined with the 1D effects to form the data for one artificial experiment. The same smoothness (full width at half maximum, fwhm=20) was used for all datasets as this is a realistic approximation of smoothness in biomechanical data (Pataky et al., 2016). 1D sample size estimation was conducted by using Monte Carlo simulation for 5000 one-sample, one-tailed, t-test experiments at sample sizes from 5-50 in increments of 5 for all datasets. 0D sample sizes were calculated by generating the 1D standardised effect size, selecting the maximum effect, then calculating the sample size required to observe the 0D effect with alpha at 0.05, and the desired power at 0.8.

3 Results

3.1 Power reporting practice

From 653 experimental journal articles published in the Journal of Biomechanics in 2018 or 2019, 29 articles or approximately 4% conducted *a-priori* power analysis. Of the 84 papers that contained the term "power" 28 papers referred to (statistical) power in the discussion, often in the context of being underpowered, and 10 papers performed post-hoc power analysis. Extracted characteristics of the studies using *a-priori* power analysis are provided (table 1). Overall most studies reported clearly the effect of interest and this was usually from either unpublished pilot data or previous work from the literature. Around half of the sample size estimations undertaken were clearly replicable from the information provided. G*power was the most frequently reported software. From 17 papers that

reported 1D data, none conducted 1D statistical analysis or 1D power analysis. Elements of poor practice included providing the required sample size but no indication of how this was determined, providing effect sizes without sufficient context about their origin, powering the study based on only one variable from many being reported, and calculating the required sample size but then overpowering the study. A bibliography containing all papers reviewed is provided (see supplementary material).

Table 1: Reporting evaluation of *a-priori* power analysis articles (* = one study reported a power of 0.1)

power	clear effect	published effect	replicable	software	1D measures	1D power
0.8 (24)	Yes (22)	Yes (13)	Yes (16)	G*power (10)	Yes (17)	Yes (0)
0.95 (1)	No (7)	No (16)	No (13)	unknown (19)	No (12)	No (17)
unknown* (3)						

3.2 Sample size estimates

Representative alternative 1D effects from a variety of biomechanical signals were created using *power1d*. Numerical simulation of artificial experiments across a range of sample sizes found that the sample sizes required to achieve a power of 0.8 varied substantially (figure 5). Datasets where the effect size was large (>1) achieved >0.8 power with small sample sizes (≤ 10). Where effect sizes were comparatively smaller (<1), 0.8 power was achieved at sample sizes of 25-50. 0D sample size estimation resulted in a smaller required sample size to achieve 0.8 power compared to the 1D results in all datasets with the greatest difference being >20 participants.

4 Discussion

The first aim of this paper was to evaluate the use of sample size estimation in recent biomechanics studies. *A-priori* sample size estimations were performed infrequently in the biomechanics studies we evaluated. In those that did perform sample size estimations many lacked the appropriate detail to be fully transparent or replicable. Other aims of this paper were to estimate sample sizes for artificial 1D effects from the biomechanics literature and to contrast 1D and 0D power analysis results. By generating alternative 1D effects and simulating artificial experiments across sample sizes we demonstrated novel sample size estimation results for a variety of 1D biomechanical signals, and illustrated the disparity between 1D and 0D sample size estimation.

4.1 Power reporting practice

The observation of infrequent sample size estimations further validates previous calls for researchers to justify their sample sizes (Knudson, 2017, Harrison et al., 2020, Vagenas et al., 2018, Lakens, 2021). As *a-priori* sample size estimations are encouraged as part of a rigourous scientific process (Harrison et al., 2020, Pataky et al., 2018, Abt et al., 2020), possible explanations for poor adherence to this are: (1) biomechanics studies are often exploratory, rather than confirmatory and they test a null-effect rather than an alternative prediction, making specific predictions difficult, (2) the complexity of the biological system being observed and the number of variables that can be derived from biomechanical data analysis typically means that a distinction between a primary outcome measure and other data is not made, (3) fears that the required sample size would conflict with what is possible due to the temporal, financial or logistical constraints of the experiment, (4) not having to justify sample sizes has become normalised behaviour in the reporting biomechanics studies, (5) the methods or tools to calculate required sample sizes for complex biomechanical data do not yet exist, (6) biomechanists often come from applied scientific fields where training on statistical data analysis may be inadequate. Regardless of the possible justifications, many of these explanations can be mitigated with appropriate planning and careful consideration of biomechanics experiments. Researchers, reviewers, and editors should also require biomechanics studies to meet higher academic standards in respect to sample size justification. Many insightful resources exist to help researchers justify their sample sizes (e.g. Batterham & Atkinson, 2005, Knudson, 2017, Lakens, 2021).

Elements of poor practice within studies conducting sample size estimations were found. As inferential statistics and null-hypothesis significance testing are a primary analysis approach within biomechanics (Vagenas et al., 2018) the accurate reporting of sample size estimations ensures the maximum reproducibility and transparency of an experiment. To improve reporting practice we recommend reporting the following when *a-priori* sample size estimation is performed: the statistical test/design for which the sample size estimation was performed, whether one or two tailed, the variable of interest, the effect size of interest, the rationale for the effect size of interest, the variability of the effect (if using a non-standardised effect size), alpha and beta (error rates), the software used and any design specific requirements. We do not believe this to be unnecessarily complex or prohibitive to researchers. More specifically we would also encourage researchers to distinguish their primary outcome measure from other secondary hypotheses and analysis. This allows researchers to power their study appropriately and avoid any ambiguity with other variables and analyses. Specifying a primary outcome measure may also then allow future studies to make more precise predictions, progressing towards more confirmatory/ predictive studies. Once common outcome measures are agreed, core outcome measurement sets can be developed as is commonplace in clinical trials (Boers et al., 2014).

For *a-priori* sample size estimation, researchers are required to determine an appropriate effect size.

One common way to do this is to use published studies (as was done in 45% of reviewed papers) and/or pilot data. There are however concerns raised about various biases with this approach which leads to the new study being underpowered (Albers & Lakens, 2018). Alternative justifications of effect sizes are possible (Lakens, 2021), and include the smallest effect size of interest (Albers & Lakens, 2018), the minimal detectable effect, or conducting a sensitivity power analysis. At the moment *power1d* only permits the latter. Whilst this paper only considers *a-priori* power analysis, researchers may consider justifying their sample size based on alternative reasoning such as resource constraints (see Lakens, 2021 for others). Justification of both the sample size and the effect size will improve the transparency of biomechanics studies.

4.2 Sample size estimation

Analytical sample size estimation is commensurately more complex for 1D data than for 0D data. Indeed, analytical 1D power analysis is possible only for simple 1D effects and is impossible for arbitrary 1D effects. This paper however has demonstrated that arbitrary 1D effects can be specified relatively easily to form an alternative (predicted) 1D signal and numerical 1D power analyses, including the numerical procedures in *power1d*, are relatively easy to conduct, and are substantially more flexible than analytical procedures. Moreover, numerical power analysis converges to analytical results when the simulated noise follows the analytical approaches presumed variance model (usually Gaussian), provided a sufficient number of artificial experiments are iteratively conducted, usually requiring 1,000 to 10,000 iterations for numerical convergence (Pataky, 2017). Considering the variety of biomechanical datasets and signals chosen, the flexibility with which *power1d* can specify the effect shape, magnitude and timing led to convincing artificial representations of these signals. For biomechanics experiments however, any predicted 1D effect should be considered as biomechanically plausible not just synthetically possible. The estimated sample sizes (≤ 50) were not of a magnitude prohibitive for biomechanics experiments. All sample sizes reflected the interactions between the 1D effect magnitude and the 1D noise as expected. The specification of all 1D noise models was based on standard deviations matched closely to the original studies. Often however standard deviations were numerically reported for a 0D effect e.g. a peak, rather than as an average standard deviation across a 1D curve or as a pooled standard deviation to consider multiple experimental conditions. So whilst the standard deviations used are informed by the biomechanics experiments they do not exactly match the actual standard deviations required to precisely replicate the experiments. This is not thought to be a problem for demonstrative purposes as in this study, but illustrates that care should be taken when choosing the input parameters for 1D sample size estimation and sensitivity analyses should be undertaken where necessary. As different standard deviations would affect the sample size estimates additional consideration of the effect of altered standard deviations only is provided (figure

6). Smooth Gaussian noise was chosen to represent the noise present within biomechanical data as in previous studies (Woltring, 1985, Wachowiak, Rash, Quesada, & Desoky, 2000, Pataky, Robinson, Vanrenterghem, & Challis, 2019 however the noise present within biomechanical data is likely more complex and signal dependent. This is therefore a reason to consider required sample sizes as sample size estimates rather than calculations, thereby implying some indication of uncertainty remains present.

Comparing 0D estimated sample sizes to 1D estimated sample sizes showed that 0D sample sizes were always smaller. This means that if 0D sample sizes are estimated to test a 1D hypothesis the experiment would almost certainly be underpowered, except if the specified effect was very large. The interaction between 0D and 1D effect sizes, and sample size is exemplified for one dataset (figure 7). Researchers are therefore encouraged to specify a 1D effect for their experiment to avoid a type II error. Ensuring an appropriate sample size should negate type II errors, but multiple testing within biomechanics studies often will inflate the type I error (Pataky et al., 2016). Careful consideration of the suggestions in this paper should mitigate the potential for incorrect conclusions within biomechanics experiments. In this study the 0D sample sizes were only compared to the omnibus power from a 1D experiment. The omnibus power fully considers the whole of the 1D continuum, that is, it considers the probability of rejecting the null hypothesis at any continuum location. This therefore means that the null hypothesis may not always be rejected at the location of the maximum alternative signal. Power estimates for precise locations within a continuum are possible but were not considered within this study (see "centre of interest power" and "point of interest power" in Pataky, 2017, Pataky et al., 2018). Power for hypotheses that pertain to less than the full 1D continuum would further complicate theoretical comparisons between 0D and 1D power analysis.

Whilst many software options are available for calculating 0D power analysis, options for 1D analysis are at the moment restricted to the Python package *power1d*. Although *power1d* is open-source it does require some familiarity with basic Python programming. We are currently developing a graphical interface to *power1d*, which requires no Python knowledge, and with which full 1D power analysis can be conducted. Whilst other nD power assessment software is available this is more suited to neuroimaging applications and it does not provide a means to numerically simulate data to estimate sample sizes (see Pataky, 2017 for a comprehensive software comparison). Numerical simulation as used in this study allows researchers to consider both the signal and noise in their original units.

4.3 Limitations

Elements of complexity of 1D power not considered in this study include research design dependence as all samples in this study were estimated using a one-sample t-test design. We nevertheless note that power1d supports power analysis for arbitrary experimental designs, including two-sample t-tests,

9

regression and ANOVA. This study also considered a single uniform smoothness value (fwhm=20). In reality datasets have arbitrary smoothness (Pataky, Vanrenterghem, Robinson, & Liebl, 2019), and greater smoothness will generally cause 1D power results to converge to 0D power results. However, the smoothness parameter can also be regarded as an implicit component of the effect; just as in 0D analysis it is typical to specify the effect and not explicitly the standard deviation, in 1D analysis it may be similarly suitable to regard the modelled effect as implicitly embodying both standard deviation and the fwhm. In other words, the *a-priori* modelled effect (and standard deviation) will never exactly match an experimentally observed effect (or standard deviation). Power analysis for 1D data can in our view be regarded similarly, with an approximate effect used that is used to guide experimental sample sizes. Regardless, further consideration of the relationship between the standard deviation, data smoothness and resulting effects should be considered in future work. We regard these limitations as essential future considerations, but also as largely irrelevant to this paper's conclusions, which pertain to a literature-wide need for more robust *a-priori* sample size estimation.

5   Conclusion

Recent biomechanics studies conducted sample size estimations infrequently and reported them inadequately. We used 1D biomechanical effects from the literature to demonstrate the utility of 1D versus 0D sample size estimation. Using 0D sample size estimation for a 1D hypothesis will almost certainly result in studies being underpowered. We have provided recommendations for reporting sample size estimations unambiguously. To facilitate greater clarity in reporting and to encourage predictive, hypothesis-driven research, researchers should also consider defining a primary outcome measure. This study should enable researchers to justify sample sizes for 1D biomechanical effects to address adequately powered hypothesis-driven research questions.

6   Conflict of interest statement

The authors have no conflicts of interest to declare.

References

Abt, G., Boreham, C., Davison, G., Jackson, R., Nevill, A., Wallace, E., & Williams, M. (2020). *Power, precision, and sample size estimation in sport and exercise science research.* Routledge.

Albers, C., & Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of experimental social psychology*, *74*, 187–195.

Bakke, D., & Besier, T. (2020). Shape model constrained scaling improves repeatability of gait data. *Journal of Biomechanics*, 109838.

Barrios, J., & Willson, J. (2017). Minimum detectable change in medial tibiofemoral contact force parameters: derivation and application to a load-altering intervention. *Journal of applied biomechanics*, *33*(2), 171–175.

Batterham, A. M., & Atkinson, G. (2005). How big does my sample need to be? a primer on the murky world of sample size estimation. *Physical Therapy in Sport*, *6*(3), 153–163.

Boers, M., Kirwan, J. R., Wells, G., Beaton, D., Gossec, L., d'Agostino, M.-A., . . . others (2014). Developing core outcome measurement sets for clinical trials: Omeract filter 2.0. *Journal of clinical epidemiology*, *67*(7), 745–753.

Bovi, G., Rabuffetti, M., Mazzoleni, P., & Ferrarin, M. (2011). A multiple-task gait analysis approach: kinematic, kinetic and emg reference data for healthy young and adult subjects. *Gait & posture*, *33*(1), 6–13.

De Ridder, R., Willems, T., Vanrenterghem, J., Robinson, M. A., Palmans, T., & Roosen, P. (2015). Multi-segment foot landing kinematics in subjects with chronic ankle instability. *Clinical Biomechanics*, *30*(6), 585–592.

Gomes, A. A., Ackermann, M., Ferreira, J. P., Orselli, M. I. V., & Sacco, I. C. (2017). Muscle force distribution of the lower limbs during walking in diabetic individuals with and without polyneuropathy. *Journal of neuroengineering and rehabilitation*, *14*(1), 111.

Harrison, A. J., McErlain-Naylor, S. A., Bradshaw, E. J., Dai, B., Nunome, H., Hughes, G. T., . . . Fong, D. T. (2020). Recommendations for statistical analysis involving null hypothesis significance testing. *Sports Biomechanics*.

Kipp, K., Comfort, P., & Suchomel, T. J. (2019). Comparing biomechanical time series data during the hang-power clean and jump shrug. *Journal of Strength and Conditioning Research*.

Knudson, D. (2017). Confidence crisis of results in biomechanics research. *Sports biomechanics*, *16*(4), 425–433.

Lakens, D. (2021). *Sample size justification.* PsyArXiv. Retrieved from `psyarxiv.com/9d3yf` doi: 10.31234/osf.io/9d3yf

Lohse, K., Sainani, K., Taylor, J. A., Butson, M. L., Knight, E., & Vickers, A. (2020). Systematic review of the use of "magnitude-based inference" in sports science and medicine.

Naouma, H., & Pataky, T. C. (2019). A comparison of random-field-theory and false-discovery-rate inference results in the analysis of registered one-dimensional biomechanical datasets. *PeerJ*, *7*, e8189.

Pataky, T. C. (2012). One-dimensional statistical parametric mapping in python. *Computer Methods in Biomechanics and Biomedical Engineering*, *15*(3), 295–301.

Pataky, T. C. (2017). Power1d: a python toolbox for numerical power estimates in experiments involving one-dimensional continua. *PeerJ Computer Science*, *3*, e125.

Pataky, T. C., Robinson, M. A., & Vanrenterghem, J. (2013). Vector field statistical analysis of kinematic and force trajectories. *Journal of biomechanics*, *46*(14), 2394–2401.

Pataky, T. C., Robinson, M. A., & Vanrenterghem, J. (2018). A computational framework for estimating statistical power and planning hypothesis-driven experiments involving one-dimensional biomechanical continua. *Journal of biomechanics*, *66*, 159–164.

Pataky, T. C., Robinson, M. A., Vanrenterghem, J., & Challis, J. H. (2019). Smoothing can systematically bias small samples of one-dimensional biomechanical continua. *Journal of biomechanics*, *82*, 330–336.

Pataky, T. C., Vanrenterghem, J., & Robinson, M. A. (2016). The probability of false positives in zero-dimensional analyses of one-dimensional kinematic, force and emg trajectories. *Journal of Biomechanics*, *49*(9), 1468–1476.

Pataky, T. C., Vanrenterghem, J., Robinson, M. A., & Liebl, D. (2019). On the validity of statistical parametric mapping for nonuniformly and heterogeneously smooth one-dimensional biomechanical data. *Journal of biomechanics*, *91*, 114–123.

Phan, X., Grisbrook, T. L., Wernli, K., Stearne, S. M., Davey, P., & Ng, L. (2017). Running quietly reduces ground reaction force and vertical loading rate and alters foot strike technique. *Journal of sports sciences*, *35*(16), 1636–1642.

Pincheira, P. A., De La Maza, E., Silvestre, R., Guzmán-Venegas, R., & Becerra, M. (2019). Comparison of total hip arthroplasty surgical approaches by statistical parametric mapping. *Clinical biomechanics*, *62*, 7–14.

Ramsay, J. O., & Silverman, B. W. (2007). *Applied functional data analysis: methods and case studies*. Springer.

Robinson, M. A., Donnelly, C. J., Tsao, J., & Vanrenterghem, J. (2014). Impact of knee modeling approach on indicators and classification of anterior cruciate ligament injury risk. *Medicine and science in sports and exercise*, *46*(7), 1269–1276.

Serrien, B., Goossens, M., & Baeyens, J.-P. (2019). Statistical parametric mapping of biomechanical one-dimensional data with bayesian inference. *International Biomechanics*, *6*(1), 9–18.

Vagenas, G., Palaiothodorou, D., & Knudson, D. (2018). Thirty-year trends of study design and statistics in applied sports and exercise biomechanics research. *International Journal of Exercise*

Science, *11*(1), 239–259.

Wachowiak, M. P., Rash, G. S., Quesada, P. M., & Desoky, A. H. (2000). Wavelet-based noise removal for biomechanical signals: A comparative study. *IEEE Transactions on biomedical engineering*, *47*(3), 360–368.

Woltring, H. J. (1985). On optimal smoothing and derivative estimation from noisy displacement data in biomechanics. *Human Movement Science*, *4*(3), 229–245.
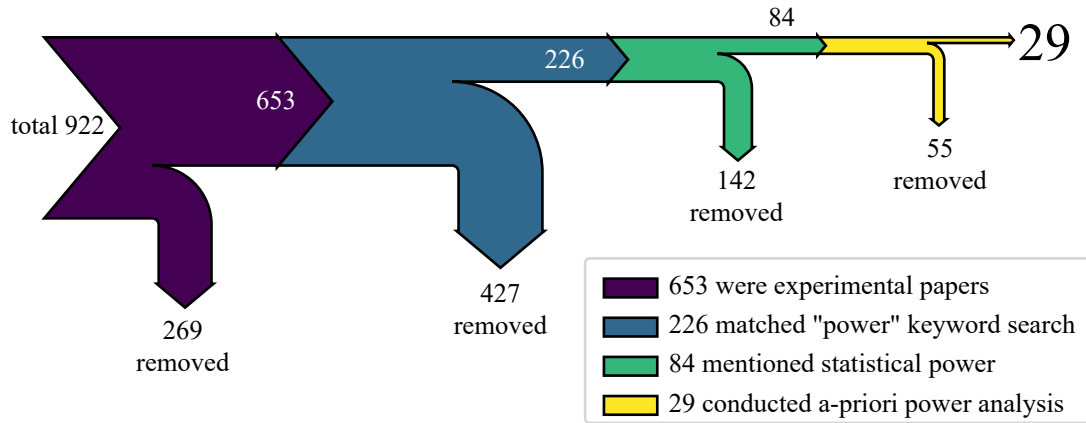
Figure 1: Narrowing down articles in the Journal of Biomechanics 2018-2019 to those containing *a-priori* power analysis.
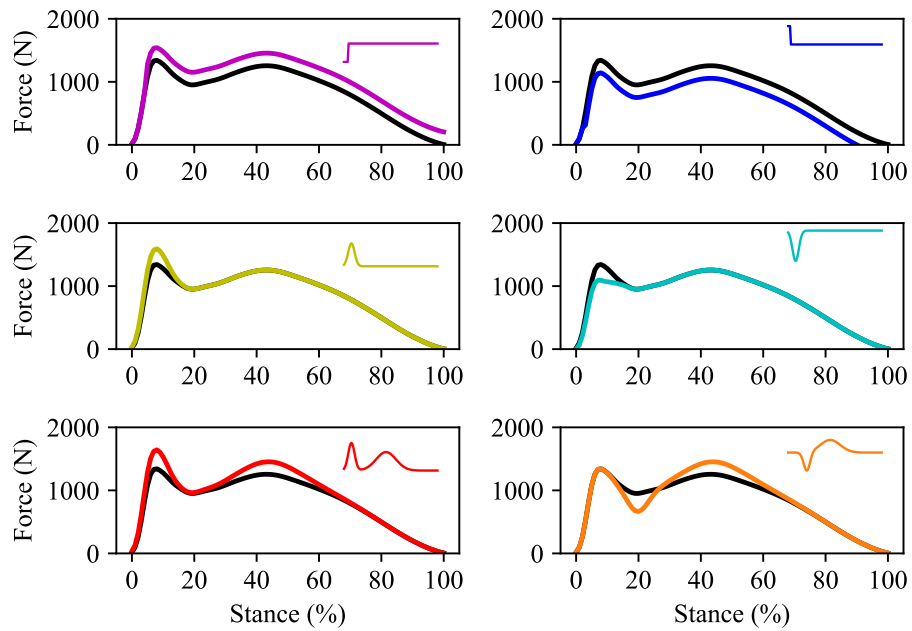
Figure 2: Alternative 1D effects created from a vertical ground reaction force curve. The baseline (black) ground reaction force is combined with the 1D artificial effect (inset) to produce the new 1D (alternative) "effect".
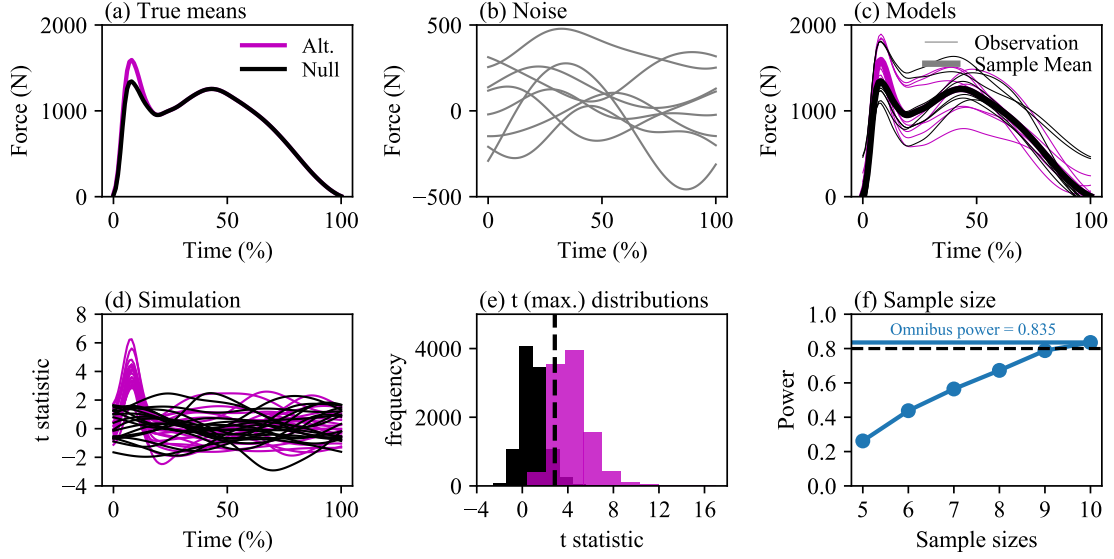
Figure 3: The workflow used to conduct sample size estimation for a 1D effect. 1D effects representing the null and alternative hypothesis are combined with 1D random noise to represent the artificial experiment of interest with sample means and multiple observations (a-c). Multiple artificial experiments are simulated (n=10,000), and 20 example t-statistic curves shown for the null and alternative experiments (d). The maximum t statistic is stored for all simulated experiments (e). The t distribution these experiments create is formed, and the 95th percentile of the null effect is determined (dashed vertical line). The proportion of the alternative t distribution exceeding the critical threshold is the omnibus power for the experiment. Repeating this process across a range of sample sizes (n=5:10) determines the minimum sample size required to achieve 0.8 power (f).

(a) JCF
Barrios et al. (2017)

(b) GRF
Phan et al. (2017)

(c) EMG
Bovi et al. (2011)

(d) Angle
Bakke & Besier (2020)

(e) Moment
Robinson et al. (2014)
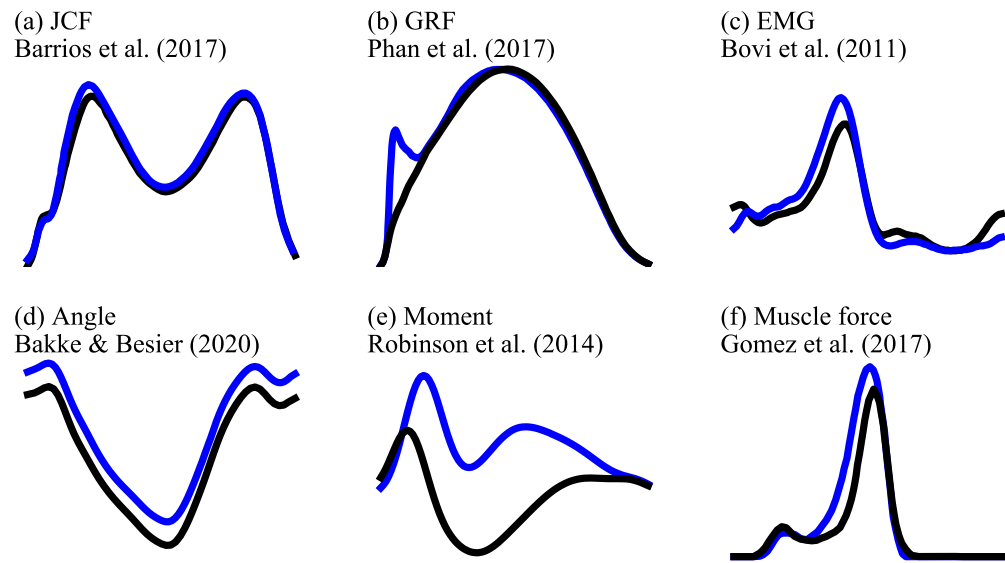
(f) Muscle force
Gomez et al. (2017)

Figure 4: Example 1D biomechanical effects from the biomechanics literature. Each panel shows two example 1D effects from the same experiment. JCF: Joint Contact Force, GRF: Ground Reaction Force, EMG: Electromyography
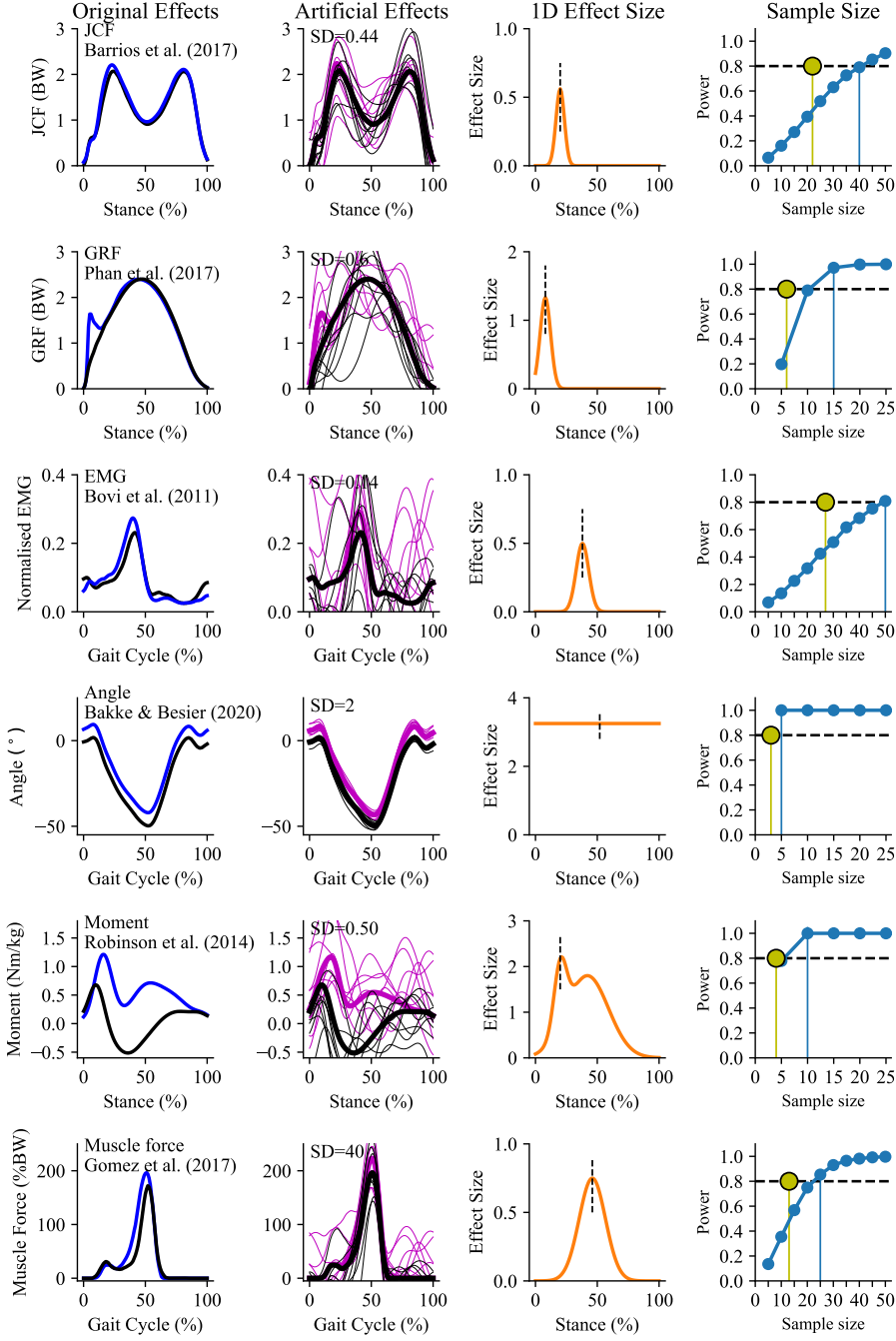
Figure 5: Real effects from biomechanical datasets (column 1) reproduced artificially in *power1d*, with noise (column 2). The 1D effect shows the maximum point (vertical dashed line) at which 0D sample size was estimated (column 3). The omnibus power from multiple 1D sample size simulations shows where the sample size exceeds a power of 0.8 (column 4). The 0D sample size required for 0.8 power is annotated as a yellow dot.
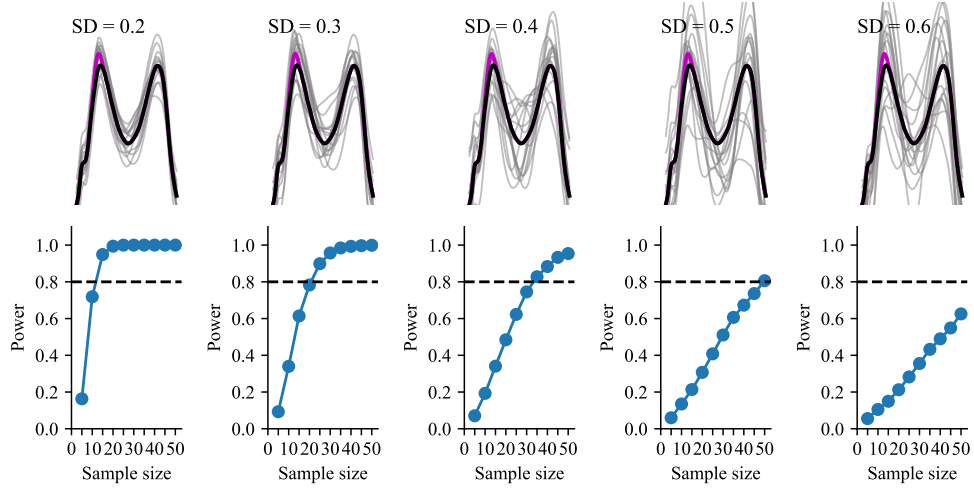
Figure 6: The biomechanical effect from Barrios et al. (see figure 4a, figure 5 row 1) with varying noise. The standard deviation (SD) for the null and alternative model is increased from 0.2-0.6 (row 1). Noise from both models has the same colour. The effect on omnibus power and sample size estimation is shown underneath (row 2).
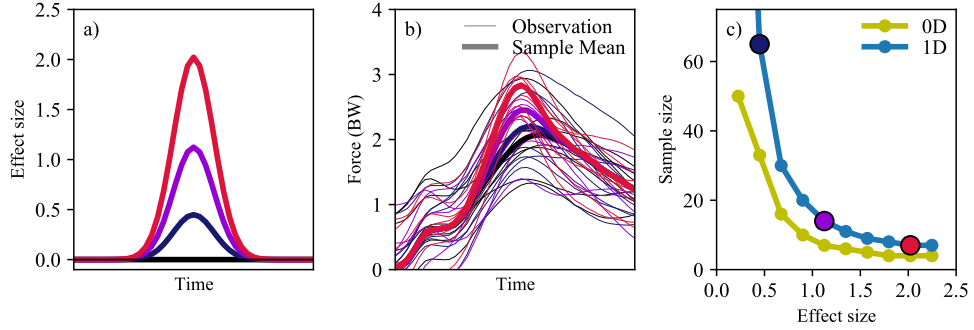

Figure 7: A comparison of 0D vs 1D sample size estimation as a function of effect size (alpha = 0.05, target power 0.8). (a) Three 1D effect sizes for illustration and the null effect (black), (b) a 1D experimental context with these effect sizes, (c) sample size estimates for 0D vs 1D effects plotted for a range of 10 different effect sizes each, the sample sizes that correspond to the effect sizes in (a) are shown as larger markers.