Addictive Behaviors 94 (2019) 117–123

Contents lists available at ScienceDirect

Addictive Behaviors

journal homepage: www.elsevier.com/locate/addictbeh

Conducting sensitivity analyses to identify and buffer power vulnerabilities in studies examining substance use over time

Sean P. Lane*, Erin P. Hennes

Purdue University, West Lafayette, IN 47907, USA

HIGHLIGHTS

- Sensitivity analysis is a tool for identifying model vulnerabilities.
- Base rates, collinearity, and random slopes were considered.
- Low/high base rates, large random slopes impede power; collinearity less so.
- Increasing participants buffer power more than increasing assessments.
- Sensitivity analyses can empower researchers to optimize study design.

ARTICLE INFO

Keywords:
Alcohol
Cannabis
Longitudinal designs
Multilevel modeling
Statistical power

ABSTRACT

Introduction: A priori power analysis is increasingly being recognized as a useful tool for designing efficient research studies that improve the probability of robust and publishable results. However, power analyses for many empirical designs in the addiction sciences require consideration of numerous parameters. Identifying appropriate parameter estimates is challenging due to multiple sources of uncertainty, which can limit power analyses' utility.

Method: We demonstrate a sensitivity analysis approach for systematically investigating the impact of various model parameters on power. We illustrate this approach using three design aspects of importance for substance use researchers conducting longitudinal studies – base rates, individual differences (i.e., random slopes), and correlated predictors (e.g., co-use) – and examine how sensitivity analyses can illuminate strategies for controlling power vulnerabilities in such parameters.

Results: Even large numbers of participants and/or repeated assessments can be insufficient to observe associations when substance use base rates are too low or too high. Large individual differences can adversely affect power, even with increased assessments. Collinear predictors are rarely detrimental unless the correlation is high.

Conclusions: Increasing participants is usually more effective at buffering power than increasing assessments. Research designs can often enhance power by assessing participants twice as frequently as substance use occurs. Heterogeneity should be carefully estimated or empirically controlled, whereas collinearity infrequently impacts power significantly. Sensitivity analyses can identify regions of model parameter spaces that are vulnerable to bad guesses or sampling variability. These insights can be used to design robust studies that make optimal use of limited resources.

1. Introduction

Advances in longitudinal study design and analysis have greatly increased understanding of causes and effects of substance use with respect to a number of cognitive, affective, and behavioral processes and associated mental and physical health characteristics (e.g., Trull & Ebner-Priemer, 2013; Wilhelm, Perrez, & Pawlik, 2012). At the same time, these discoveries generally require a great deal of time, human, and monetary investment. It is critical that studies of substance use behaviors and disorders be robustly designed to maximize informational value (Witkiewitz, Finney, Harris, Kivlahan, & Kranzler, 2015).

A priori power analysis (Cohen, 1962, 1988, 1992) is one method

* Corresponding author at: Department of Psychological Sciences, Psychology Building, 703 Third Street, West Lafayette, IN 47907, USA.
E-mail address: seanlane@purdue.edu (S.P. Lane).

https://doi.org/10.1016/j.addbeh.2018.09.017
Received 1 April 2018; Received in revised form 6 August 2018; Accepted 11 September 2018
Available online 12 September 2018
0306-4603/

for enhancing empirical reliability that has recently gained renewed attention in the addiction sciences (e.g., *Addiction*, 2018; Hallgren & Witkiewitz, 2013; National Institutes of Health, 2015; Tackett et al., 2017). In brief, power analysis usually consists of estimating anticipated effect(s) of a planned study and calculating the sample size necessary to have some predetermined probability (e.g., 80%) of correctly rejecting the null hypothesis (see Cohen, 1988, 1992; Maxwell, 2004; for detailed discussions of statistical power). Appropriately powering a planned study has a number of benefits, including reducing Type II error, improving the ability to distinguish true nulls from underpowered effects, increasing precision, and ensuring optimal resource utilization.

Several tools exist for estimating power for some basic and more complex designs, such as G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) and Optimal Design (Raudenbush, Spybrook, Liu, & Congdon, 2011). In addition, we have recently elaborated a flexible simulation-based power analysis method (see also, Bolger & Laurenceau, 2013; Gelman & Hill, 2006; Muthén & Muthén, 1998-2017). This approach is flexible to any model specification and is particularly useful when closed-form power calculation tools are unavailable. A detailed discussion of this approach (and corresponding syntax) is available in Lane and Hennes (2018).

Despite the availability of simulation- and equation-based tools, power analyses often entail considerable uncertainty on a number of dimensions. Many analytic models employed by substance use researchers, such as multilevel (e.g., Fitzmaurice, Laird, & Ware, 2012; Raudenbush & Bryk, 2002) or robust designs (e.g., Liang & Zeger, 1986), entail observations at multiple levels of analysis, individual differences in associations across clusters/individuals, and a large number of additional parameters that may significantly impact power but are infrequently reported. Despite known mathematical associations between certain statistical constructs and power (e.g., base rates and variability; Preacher, Rucker, MacCallum, & Nicewander, 2005), the magnitude of these effects as a function of model type and other parameters is essentially unexplored. Researchers often have poor intuitions about different parameters' impact on power (Bakker, Hartgerink, Wicherts, & van der Maas, 2016), and there are few available and accessible tools that elucidate such relationships. Where available, they are generally limited to variations in between-subjects sample size and a single effect size (e.g., G*Power; Faul et al., 2007). Additionally, researchers must consider the phenomenology of processes of interest, such as how often use occurs, whether it is confounded/collinear with other behaviors (e.g., co-use), and how often use will be observed given the assessment schedule.

As a result, even with available power estimation tools, challenges in identifying appropriate point estimates across a long vector of parameters can lead to highly misleading sample size determinations. Conducting a single power analysis using a “best guess” does not reveal vulnerabilities in the estimated model where misjudgments may lead to significant under- or overestimates of required sample size (nor areas where misjudgments have limited impact). Without this knowledge, a researcher will not have the requisite information to know whether additional pilot testing or literature review is necessary to obtain more precise parameter estimates, nor whether to consider alternative study designs that buffer power. Individual power analyses, while incredibly useful, may give researchers a sense of false confidence, leading them to make design decisions that do not maximize statistical power and methodological efficiency.

To address these challenges, the current study demonstrates the utility of sensitivity analyses, in which power analyses are conducted across a multivariate range of parameter estimates (see also Bolger, Stadler, & Laurenceau, 2012; Gelman & Hill, 2006). Rather than relying on heuristic recommendations that may not generalize to their model or in which the magnitude of impact on their own data is unknown, sensitivity analyses enable researchers to explore their own empirical models to independently make efficient design decisions.

1.1. Current study

We draw on results reported in two articles that adopt different longitudinal designs to demonstrate the added benefit of a sensitivity analysis approach over individual power analysis. The first employed a multi-year longitudinal design to characterize differences in hangover trajectories as a function of sex and family history of alcoholism (Piasecki, Sher, Slutske, & Jackson, 2005). The second used an ambulatory assessment design (Trull & Ebner-Priemer, 2013) and examined temporal associations between alcohol/cannabis use and affect (Trull, Wycoff, Lane, Carpenter, & Brown, 2016). We demonstrate a series of sensitivity analyses that a researcher might initiate if they aimed to replicate findings from either article in subsequent research. We adopt the simulation approach described elsewhere to conduct all power and sensitivity analyses (e.g., Bolger & Laurenceau, 2013; Gelman & Hill, 2006; Lane & Hennes, 2018; Muthén & Muthén, 1998-2017).

We use the authors' statistical models and parameter estimates as representative starting points, and systematically vary:

- Number of participants and assessments per participant (Examples 1 and 2)
- Variability (i.e., base rates, frequency, distributions) of model predictors (Examples 1 and 2)
- Collinearity (i.e., co-use, correlation) of predictors (Example 2)
- Individual differences (i.e., random slopes) in the association between model predictors and outcomes (Examples 1 and 2)

In light of classic power formulae (Cohen, 1988), we hypothesize that, 1) base rates that maximize variability, 2) lower correlations between predictors, and 3) smaller individual differences will be associated with increased power. Based on recent research using similar designs (Lane & Hennes, 2018; see also Rast & Hofer, 2014; Rouder & Haaf, 2018), we hypothesize that 4) increasing participants would benefit power more than increasing the number of assessments per participant. However, given the nonlinear effects of varying such parameters as a function of effect size (c.f., Lane & Hennes, 2018), it was unclear the extent to which these two parameters would buffer power in the context of base rates, correlated predictors, or random slopes in the current examples of substance use over time and in daily life. Thus, the findings from the current simulations can inform studies characterized by similar design characteristics, possibly across substantive domains. At the same time, the aim of the current manuscript is not merely to report specific findings regarding these limited study characteristics, but to illustrate how researchers can adopt such practices to independently optimize their designs.

2. Method

2.1. Example 1 – Effects of family history on hangover frequency in early adulthood

Piasecki et al. (2005) examined hangover frequency (0 = no past-year hangover, 8 = 40 or more hangovers in the past year) across 6 assessments spanning 11 years (Years 1, 2, 3, 4, 7, and 11) among 486 college freshmen (at study entry). They modeled linear trajectories over time as moderated by participants' family history of alcoholism (51% family history positive [FHP; family history negative = FHN]; 1 = FHP, 0 = FHN) and sex (53% female; 0 = female, 1 = male). They also included a time-varying covariate of the number of heavy-drinking days in the past month and its interactions with family history and sex. Four hundred ten participants were assessed at Year 11, indicating 16% missing data at the final wave. We assumed a constant (1.7% per year) rate of dropout to generate an approximate pattern of missing data.

We focus on the interaction between family history and year, which the authors report to be negative and significant ($b = -0.05$, $p < .05$; Table 1). This intriguing finding warrants replication. While the authors replicate previous findings indicating that FHP individuals are at higher

Table 1
Model estimates, observed power, and *N* required for 80% power to replicate each effect using the same assessment schedule for the linear growth model of hangover frequency reported by Piasecki et al. (2005).

Main effects	Estimate	Observed power	N for 80% power
Intercept	1.42***	100%	
Family History	0.53**	89%	368
Sex	0.42*	72%	588
Linear Slope	−0.06***	90%	363
Linear Slope * Family History	**−0.05***	**67%**	**655**
Linear Slope * Sex	0.05*	71%	609
Heavy-drinking Covariate	1.62***	100%	40
Heavy-drinking * Family History	−0.30*	57%	827
Heavy-drinking * Sex	−0.51***	95%	294

Random effects	Variance
Person intercept	2.79***
Person linear slope	0.02***
Person heavy-drinking covariate	0.70***

Note. Bold-type indicates the effect of interest for the simulation analyses.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

risk for hangover in college (e.g., Newlin & Pretorious, 1990), their data suggests that this risk diminishes as individuals enter adulthood, consistent with an interpretation of FHP as a "developmentally limited" risk factor of hangover (Sher & Gotham, 1999).

Researchers interested in conducting replications or extensions of this study would need to make a number of potentially costly design decisions, such as number of participants, number and timing of assessments, distributions of gender, family history, and heavy-drinking within the sample, and inclusion of additional covariates or moderators. However, the magnitude of impact of such decisions on power may not be obvious. Here we illustrate sensitivity analyses exploring the impact of a few of such factors, (a) sample size, (b) number of assessments (c) FHP proportion, and (d) individual differences in the linear slopes over time (that are unexplained by FHP or sex), on power to replicate the linear slope by family history interaction effect. Before examining these factors, we first simulated datasets that reproduced the reported effect sizes of Piasecki et al. (2005) as closely as possible. Syntax and details about this approach are provided in Appendix A.

2.2. Example 2 – Effects of alcohol and cannabis use on positive affect among individuals with borderline personality or depressive disorder

Trull et al. (2016) report results from a sample of 93 psychiatric outpatients diagnosed with either borderline personality or depressive disorders who reported using alcohol and/or cannabis at least once (on average) over the course of 27 days. Individuals were randomly prompted six times each day and asked to report, in part, on their felt positive affect (1 = very slightly or not at all, to 5 = extremely) and their alcohol and cannabis use (1 = yes, 0 = no) since the last prompt. On average, individuals completed 144.5 prompts each. The authors report parameter estimates for a three-level multilevel model in which current and previous occasion substance use (level-1), current and previous day substance use (level-2), and person average substance use (level-3), along with covariates, were used to predict positive affect.

The authors find statistically significant positive associations between positive affect and alcohol use on the current day, but significant negative associations between positive affect and alcohol use on the

Table 2
Model estimates, observed power, and *N* required for 80% power to replicate each effect using the same assessment schedule from Trull et al. (2016) of the associations between concurrent and lagged alcohol and cannabis use and positive affect.

Main effects	Positive affect		Observed power	N for 80% power
	Estimate	95% CI		
Intercept	2.33***	[2.16, 2.50]	100%	
Occasion level				
Current occasion alcohol use	0.12***	[0.06, 0.18]	98%	48
Previous occasion alcohol use	−0.07*	[−0.14, −0.01]	66%	131
Current occasion cannabis use	0.02	[−0.11, 0.15]	6%	8716
Previous occasion cannabis use	0.02	[−0.05, 0.09]	8%	3134
Day level				
Current day alcohol use	0.33***	[0.16, 0.49]	97%	49
Previous day alcohol use	**−0.17***	**[−0.33, −0.02]**	**61%**	**143**
Current day cannabis use	0.11	[−0.21, 0.43]	10%	1596
Previous day cannabis use	0.11	[−0.16, 0.39]	13%	1221
Person level				
Degree of alcohol use	0.41	[−0.90, 1.73]	10%	1596
Degree of cannabis use	0.58	[−0.15, 1.32]	36%	266

Random effects	Variance
Person intercept	0.11***
Person(day) intercept	0.31***
Current occasion alcohol use	0.02
Previous occasion alcohol use	0.02
Current occasion cannabis use	0.08**
Previous occasion cannabis use	0.01
Current day alcohol use	0.07
Previous day alcohol use	0.04
Current day cannabis use	0.34
Previous day cannabis use	0.21

Note. *95% CI* = 95% confidence interval. Bold-type indicates the effect of interest for the simulation analyses.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

previous day ($b = -0.17$, 95% confidence interval [−0.33, −0.02]; Table 2). In light of the concurrent positive effect of alcohol use, the negative lagged effect could be interpreted in line with negative reinforcement models and would be important to replicate given established inconsistencies in observing such a mechanism across studies (Baker, Piper, McCarthy, Majeskie, & Fiore, 2004).

As in Example 1, a number of parameters in such a replication are under the control of the researcher. We illustrate sensitivity analyses examining the impact of (a) sample size, (b) number of assessments, (c) alcohol use base rate, (d) cannabis co-use, and (e) individual differences in the association between alcohol use and affect. As in Example 1, we first simulated datasets that reproduced the effect sizes reported in Trull et al. (2016). Syntax and details are provided in Appendix A.

2.3. Simulation procedure

For each model parameter of interest, we simulated hypothetical data using a range of possible population parameter values. This process was repeated 10,000 times for each combination of model values and subsequently analyzed using the models depicted in Eqs. S1 and S2. In each case, we coded the dichotomous statistical significance of each effect ($\alpha = 0.05$), and aggregated the dichotomous codes across the 10,000 simulated samples to create estimates of power for each parameter (see Lane & Hennes, 2018, for more information about this approach). Analyses were conducted using SAS 9.4 (SAS Institute, 2014; see Appendix A for specific syntax and parallel scripts in R).

3. Results

3.1. Initial power analyses

We first report the results of two individual a priori power analyses for replicating the negative effect of 1) FHP status on yearly linear trajectories of hangover (Example 1), and 2) previous day's alcohol use on positive affect (Example 2). Observed power for the respective effects was approximately 67% (Example 1) and 61% (Example 2). We would need approximately 655 individuals observed across 6 waves to achieve 80% power to replicate the negative FHP by linear slope interaction (Table 1). We would need approximately 143 individuals observed across 27 days to achieve 80% power to replicate the negative effect of previous day alcohol use on positive affect (Table 2).

3.2. Sensitivity analyses: Number of individuals vs. number of assessments

We now illustrate how conducting additional sensitivity analyses can offer insight into optimizing design efficiency and buffering against "bad guesses". We first report analyses examining the impact of adding participants (*N* - solid gray line) versus assessments (*n* - solid black line) for both studies (Fig. 1), holding total number of person-assessments constant. In Example 1, we would achieve 80% power (holding number of individuals constant at 486) by increasing the number of assessments up to 8 in the 11-year span (Panel A). In contrast, holding the number of individuals in Example 2 constant at 93 but increasing the number of assessment days shows that power begins to asymptote and does not reach 80% until 95 days. This indicates that even hundreds of assessments per person may be insufficient to power an effect given an insufficient number of participants (Panel B; c.f., Lane & Hennes, 2018; Rast & Hofer, 2014; Rouder & Haaf, 2018). The difference between Examples 1 and 2 is primarily due to the fact that increasing the number of assessments in Example 1 also increases the variance of the linear predictor of year. Together, these findings demonstrate the utility of sensitivity analyses to assess the benefit of increasing assessments given particular design characteristics. While in some cases increasing assessments can be valuable, often such efforts will not lead to appreciable benefits.

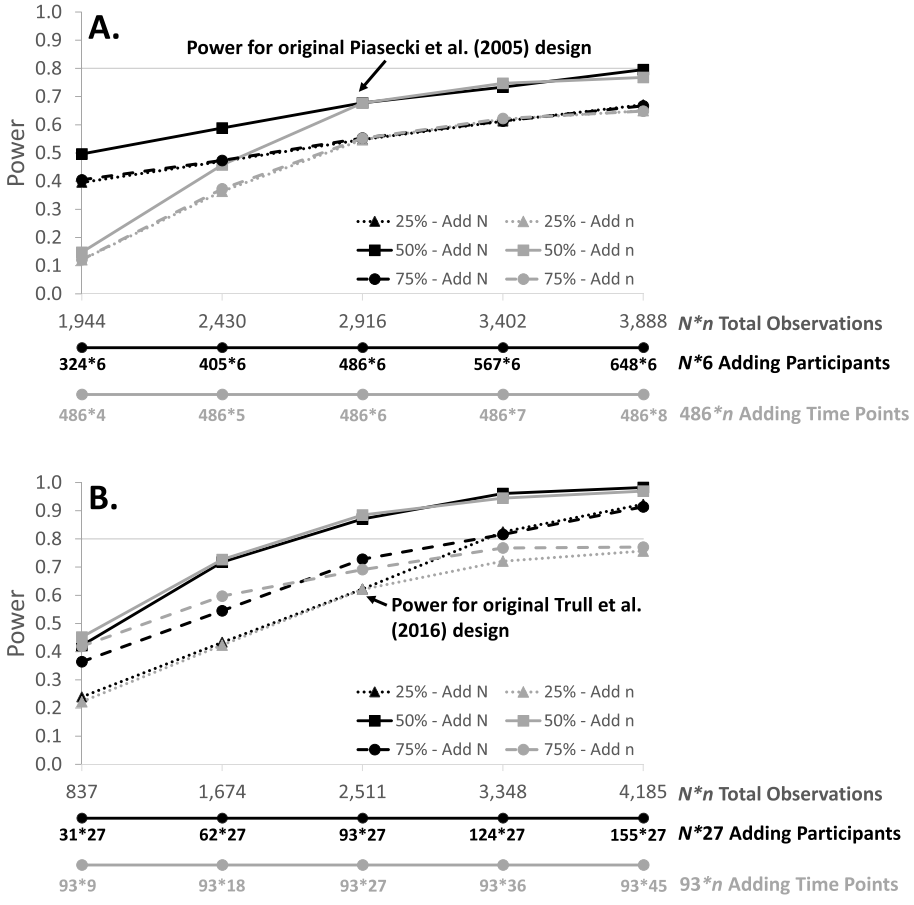

Fig. 1. Power for the Piasecki et al. (2005) FHP by year interaction effect (Panel A) and the Trull et al. (2016) lagged effect of day-level alcohol use on positive affect (Panel B), assuming 25% (triangle - short dash) 50% (square - solid), and 75% (circle - long dash) proportions of FHP/drinking as a function of increasing participants (*N*) or assessments (*n*).

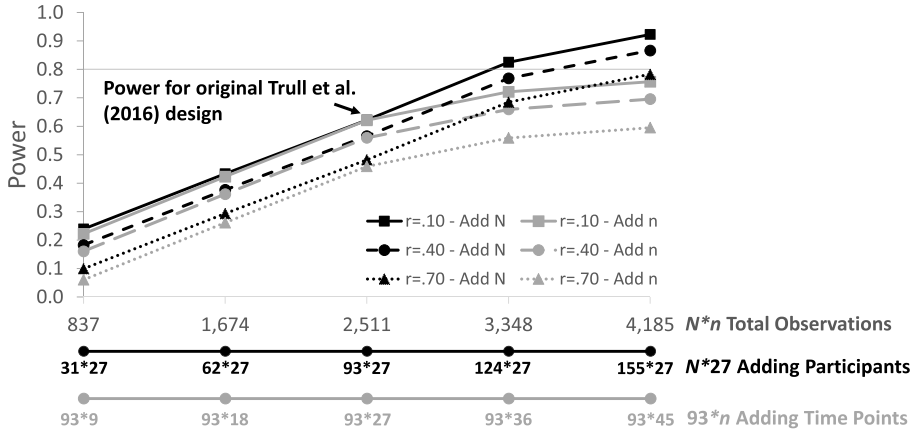

Fig. 2. Power of Trull et al. (2016) lagged alcohol effect on positive affect as a function of the correlation between daily alcohol and cannabis use when increasing participants (black lines) or assessment days (gray lines).

3.3. Sensitivity analyses: Base rates

Piasecki et al.'s (2005) sample was specifically recruited to have equal proportions of FHP and FHN individuals, which maximizes power (Fig. 1A). Trull et al. (2016) indicated that their sample reported drinking on approximately 25% of days, which is comparable to epidemiological estimates among drinkers in the United States (Dawson, Goldstein, Saha, & Grant, 2015). Fig. 1B indicates that if a researcher instead recruited participants who drank on average 50% of days, 80% power is achieved with only 77 participants or only 23 assessment days. However, increasing average drinking further to 75% of days does not further enhance power, but rather achieves similar power as using 25%-of-days drinkers.[1] This demonstrates that base rates increase power as a function of variability, not frequency, in substance use, and suggests that researchers could reduce study time and expense by developing a targeted recruitment strategy or altering assessment frequency to achieve a 50% rate of use.

3.4. Sensitivity analyses: Correlation between predictors (e.g., co-use)

The results reported by Trull et al. (2016) indicate minimal alcohol and cannabis co-use (see Appendix A for details on extracting this correlation from their reported results). Nevertheless, the larger literature suggests that daily alcohol and cannabis co-use can be quite large ($r \approx 0.60$) in certain subpopulations (e.g., Walsh et al., 2004). Therefore, it may be unwise to assume the same low correlation in subsequent power analyses. But what estimate should be used, and how much does it matter? Fig. 2 shows that although larger correlations between daily alcohol and cannabis use are associated with lower power, the effect is small. There is less than a 15% difference in power between correlations of 0.00 and 0.70. Increasing participants is associated with consistent gains (Fig. 2 – black lines), but increasing assessments is associated with smaller and smaller gains (Fig. 2 – gray lines). This indicates that researchers may need not worry about acquiring an accurate estimate of collinearity in their power analysis, nor about the impact on power of recruiting co-users unless co-use is very high and/or the fixed effect size of interest is small (although they should also consider the substantive implications of co-use inclusion criteria).[2]

3.5. Sensitivity analyses: Individual differences in the association between predictors and outcomes (i.e., random slopes)

Individual differences (e.g., in linear trajectories [Example 1]; between lagged alcohol use and positive affect [Example 2]) are estimated by modeling random slopes, but these parameters are infrequently reported. Therefore, a researcher may not have a good estimate of the random slope (s), intuitions about its impact on power, or strategies for buffering it. Fig. 3 shows that as the random slope magnitude increases, power decreases nonlinearly. Increasing participants (N) buffers against large random effects, however, only to a point. Increasing assessments (n) provides much less buffer, and the benefit it does have dissipates quickly. This illustrates that random slopes can substantially impact power, so it is worthwhile to obtain good estimates (e.g., via contacting authors or conducting pilot studies). Faulty estimates can be somewhat buffered by large samples; however, the number of observations may be unfeasible. The researcher might also consider reducing unexplained variability with moderators.

4. Discussion

Power analyses represent a valuable tool for optimizing limited resources to conduct informative research. However, the impact of inaccurate parameter estimates on power calculations can vary dramatically across model dimensions and is rarely intuitive. Sensitivity analysis can complement power analyses by buffering study designs to vulnerabilities due to uncertainty. To be sure, sensitivity analyses fall prey to the same uncertainty limitations as individual power analyses, as they are simply iterative power analyses. However, mapping a landscape of possible scenarios enables researchers to identify opportunities to optimize study methodology (e.g. assessment schedule, recruitment restrictions, moderators) and provides insight into parameters for which effort to identify precise estimates is more and less necessary. Table 3 provides a list of (incomplete) recommendations for conducting sensitivity analyses that can help researchers understand and mitigate factors that influence the power of their studies.

Although not the primary goal of the current research, the simulations illustrated here also provide specific insights along several dimensions for studies that assess binary indicators (e.g., group membership [Example 1] or frequency of substance use [Example 2]) across time. Consistent with previous research (Lane & Hennes, 2018; Rast & Hofer, 2014; Rouder & Haaf, 2018), we find that, holding person-assessments constant, increasing participants is more valuable for power than increasing assessments. This is likely to be true more generally when there are expected individual differences in (within-cluster) associations of interest. However, the degree to which this is true depends on other factors, some of which we illustrate, that are difficult to infer despite well-known statistical rules. Conducting sensitivity analyses in these cases can be extremely helpful for determining if investigators are better off recruiting a few

[1] Power is not identical for the 25% and 75% base rate models because increasing alcohol use to 75%, holding all other parameters constant, induces a negative correlation between alcohol and cannabis co-use that increases power. If the base rate of cannabis use was similarly increased to maintain the correlation the results would be identical.

[2] Smaller effect sizes are proportionately more affected by correlations with other predictors, and thus the impact of collinearity should be seriously considered if effects of interest are expected to be small.

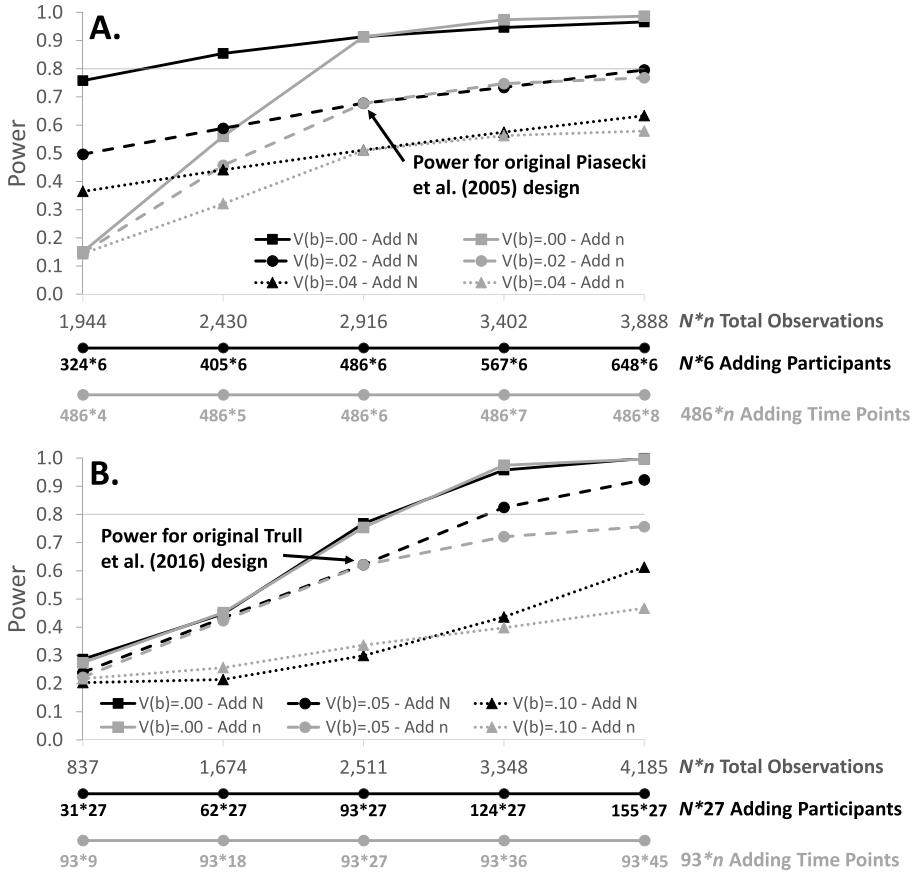

Fig. 3. Power of Piasecki et al. (2005) Year*Family History interaction (Panel A) and Trull et al. (2016) lagged alcohol effect (Panel B) as a function of the random slope magnitude when increasing participants (black lines) and assessment days (gray lines).

Table 3
General guidelines for conducting sensitivity analyses.

Step	Description
1.	Determine the complete statistical model intended to test the study hypotheses.
2.	Generate best guesses for all model parameters, based on published data, pilot studies, etc.
3.	Conduct a traditional power analysis, ensuring that the chosen method (i.e., hand calculations, software, simulation) fully accommodates the predicted model.
4.	Consider the parameters of the model in terms of (a) their centrality to the research question, (b) your certainty about their magnitude, and (c) their controllability by the researcher.
5.	Construct a range of possible values for each parameter of central interest, with endpoints corresponding to the smallest effect of interest and the largest effect that could practically be expected. Use your "best guess", minimum, and maximum as the iterations of the sensitivity analysis. Additional values can be included if more resolution is desired.
6.	Repeat Step 5 for secondary parameters (e.g., random effects), with endpoints informed by their controllability and certainty.
7.	Conduct iterative power analyses across the permutations of secondary parameter values for each primary parameter value.
8.	Identify parameters based on these sensitivity analyses that have the largest impact on power to detect the hypothesized effects.
9.	Use this information to optimize study design (e.g., increase participants, increase assessments, adjust sampling frame, create inclusion criteria, increase reliability, add moderators).

more individuals and scaling back on repeated assessments or sampling fewer participants many times.

We also demonstrate that base rates of substance use, because they directly translate to variability, can be critical for observing hypothesized associations. A sample of 93 participants similar in their drinking frequency to the average American drinker (i.e., 25% of days) did not drink enough to achieve 80% power to observe the association of interest even after 27 days of assessment. Increasing assessments had little benefitwhile increasing participants was more powerful. Alternatively, recruiting the same number of participants who instead drank on 50% of days would achieve adequate power (but recruiting those who drank on 75% of days would not). Because power increases as predictor variance increases, power is often optimized when participants are assessed approximately twice as often as use occurs.[3] Researchers might thus adjust their inclusion criteria and/or their assessment schedule (or

[3] The relationship between use frequency and observation frequency is affected by whether the predictor varies by person, within person, or displays individual differences (in random intercepts or slopes). If random slopes are not modeled or use is not centered within-person, some individuals can have a much lower and others a much higher base rate as long as the average is 50%. If random slopes or within-person centering is used, it is important to recruit individuals with little variability around the 50% base rate, as increases in individual differences can increase the random slope variance and decrease power.

implement event-contingent assessments) to maximize power.[4]

In contrast, the correlation between predictors can be quite large with relatively little decrement in power, assuming that the fixed effect of interest is not small. However, faulty random slope estimates can have a strong impact on power. Because such effects are infrequently reported in the literature, researchers should make efforts to estimate them as accurately as possible or control their impact with large sample sizes or other methods (e.g., moderators; restrictive inclusion criteria).

5. Conclusions

Collecting the rich data made available by advancements in intensive longitudinal designs comes with an array of practical and conceptual design considerations that, when ignored or left to intuition, can sabotage a study's likelihood of identifying true associations. We advocate for the use of power and sensitivity analyses to more fully understand the nuances of such designs and take preemptive action to ensure the robustness and reliability of studies conducted in addictions science.

Role of funding sources

This research was partially supported by a Purdue Research Foundation Summer Faculty Grant awarded to the first author. Neither author received any specific grant from additional funding agencies in the public, commercial, or not for profit sectors.

Contributors

Authors S.P.L. and E.P.H. generated the research design, conducted the literature search, and drafted the manuscript. S.P.L. conducted the statistical analyses. Both authors approved the final manuscript.

Conflict of interest

All authors declare that they have no conflict of interest.

Acknowledgements

A portion of these results were presented at the 2017 annual meeting of the Research Society on Alcoholism in Denver, CO.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.addbeh.2018.09.017.

References

Addiction (2018). Author guidelines. Retrieved from http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1360-0443/homepage/ForAuthors.html#TypesofArticlePublishedinAddiction.

Baker, T. B., Piper, M. E., McCarthy, D. E., Majeskie, M. R., & Fiore, M. C. (2004). Addiction motivation reformulated: An affective processing model of negative reinforcement. *Psychological Review, 111*, 33–51.

Bakker, M., Hartgerink, C. H., Wicherts, J. M., & van der Maas, H. L. (2016). Researchers' intuitions about power in psychological research. *Psychological Science, 27*, 1069–1077.

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research.* New York: Guilford.

Bolger, N., Stadler, G., & Laurenceau, J.-P. (2012). Power analysis for intensive longitudinal studies. In M. R. Mehl, & T. S. Conner (Eds.). *Handbook of research methods for studying daily life* (pp. 285–301). New York: Guilford.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *The Journal of Abnormal and Social Psychology, 65*, 145–153.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, New Jersey: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159.

Dawson, D. A., Goldstein, R. B., Saha, T. D., & Grant, B. F. (2015). Changes in alcohol consumption: United States, 2001–2002 to 2012–2013. *Drug & Alcohol Dependence, 148*, 56–61.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied longitudinal analysis. Vol. 998*. Hoboken, NJ: John Wiley & Sons.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models.* Cambridge: Cambridge University Press.

Hallgren, K. A., & Witkiewitz, K. (2013). Missing data in alcohol clinical trials: A comparison of methods. *Alcoholism: Clinical and Experimental Research, 37*, 2152–2160.

Lane, S. P., & Hennes, E. P. (2018). Power struggles: Estimating sample size for multilevel relationships research. *Journal of Social and Personal Relationships, 35*, 7–31.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*, 13–22.

Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods, 9*, 147–163.

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide* (Eighth edition). Los Angeles, CA: Muthén & Muthén.

National Institutes of Health (2015). *Implementing rigor and transparency in NIH & AHRQ research grant applications (NOT-OD-16-011)*.

Newlin, D. B., & Pretorious, M. B. (1990). Sons of alcoholics report greater hangover symptoms than sons of nonalcoholics: A pilot study. *Alcoholism: Clinical and Experimental Research, 14*, 713–716.

Piasecki, T. M., Sher, K. J., Slutske, W. S., & Jackson, K. M. (2005). Hangover frequency and risk for alcohol use disorders: Evidence from a longitudinal high-risk study. *Journal of Abnormal Psychology, 114*, 223–234.

Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: A critical reexamination and new recommendations. *Psychological Methods, 10*, 178–192.

Rast, P., & Hofer, S. M. (2014). Longitudinal design considerations to optimize power to detect variances and covariances among rates of change: Simulation results based on actual longitudinal studies. *Psychological Methods, 19*, 133–154.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd Ed). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Spybrook, J., Liu, X.-F., & Congdon, R. (2011). *Optimal design (Version 3.01)*. Ann Arbor, MI: HLM Software. Retrieved from *http://hlmsoft.net/od/*.

Rouder, J. N., & Haaf, J. M. (2018). Power, dominance, and constraint: A note on the appeal of different design traditions. *Adv. Methods Practices Psychol. Sci. 1*, 19–26.

SAS Institute (2014). *SAS/STAT 9.4 user's guide.* Cary, NC: Author.

Sher, K. J., & Gotham, H. J. (1999). Pathological alcohol involvement: A developmental disorder of young adulthood. *Development and Psychopathology, 11*, 933–956.

Tackett, J. L., Lilienfeld, S. O., Patrick, C. J., Johnson, S. L., Krueger, R. F., Miller, J. D., ... Shrout, P. E. (2017). It's time to broaden the replicability conversation: Thoughts for and from clinical psychological science. *Perspectives on Psychological Science, 12*, 742–756.

Trull, T. J., & Ebner-Priemer, U. (2013). Ambulatory assessment. *Annual Review of Clinical Psychology, 9*, 151–176.

Trull, T. J., Wycoff, A. M., Lane, S. P., Carpenter, R. W., & Brown, W. C. (2016). Cannabis and alcohol use, affect and impulsivity in psychiatric out-patients' daily lives. *Addiction, 111*, 2052–2059.

Walsh, J. M., Flegel, R., Cangianelli, Atkins, R., Soderstrom, C. A., & Kerns, T. J. (2004). Epidemiology of alcohol and other drug use among motor vehicle crash victims admitted to a trauma center. *Traffic Injury Prevention, 5*, 254–260.

Wilhelm, P., Perrez, M., & Pawlik, K. (2012). Conducting research in daily life: A historical review. In M. R. Mehl, & T. S. Conner (Eds.). *Handbook of research methods for studying daily life* (pp. 62–86). New York: Guilford.

Witkiewitz, K., Finney, J. W., Harris, A. H. S., Kivlahan, D. R., & Kranzler, H. R. (2015). Recommendations for the design and analysis of treatment trials for alcohol use disorders. *Alcoholism: Clinical and Experimental Research, 39*, 1557–1570.

[4] Such requirements may impact the processes or population of interest, such as social drinkers versus alcoholics or single- versus multi-substance users, so researchers should be thoughtful in designing studies that ensure that the phenomenology of specific processes of interest can be adequately assessed, and the generalizability of subsequent findings is appropriate.