Appendix A: Supplementary Material

Example 1 Simulation Model

Piasecki et al. (2005) provide an explicit formula for their statistical model in the main text, which we reproduce in Equation S1.

$$\begin{aligned} Hangover_{ij} = {} & (b_0+b_{0i}) + b_1*FHP_i + b_2*Sex_i \\ & + (b_3+b_{3i})*Year_{ij} + b_5*FHP_i*Year_{ij} + b_6*Sex_i*Year_{ij} \\ & + (b_4+b_{4i})*Drink_{ij} + b_7*FHP_i*Drink_{ij} + b_8*Sex_i*Drink_{ij} + e_{ij} \end{aligned} \quad \text{(S1)}$$

$Hangover_{ij}$ is the hangover frequency rating of person i at year j. At a given year j an individual's *Hangover* score is fit as a function of a global intercept (b_0), *FHP* status (b_1), *Sex* (b_2), *Year* of assessment (b_3), report of heavy drinking (b_4), and the four corresponding two way interactions (b_5–b_8). There are also three estimated random effects (subscripted by i) that were not originally reported; however, the lead author provided us with the random effect variances and error variances (e_{ijk}). The authors did not estimate correlations between the random effects, so we constrained them to be zero.

Example 2 Simulation Model

Trull et al.'s (2016) Supporting Information provides the formula for the estimated model, which we reproduce in Equation S2.

$$\begin{aligned} PA_{ijk} = {} & (b_0+b_{0i}+b_{0j(i)}) + (b_1+b_{1i})*Alc\_o_{ijk} + (b_2+b_{2i})*Alc\_o_{ijk-1} + (b_3+b_{3i})*Can\_o_{ijk} \\ & + (b_4+b_{4i})*Can\_o_{ijk-1} + (b_5+b_{5i})*Alc\_d_{ij} + (b_6+b_{6i})*Alc\_d_{ij-1} + (b_7+b_{7i})*Can\_d_{ij} \\ & + (b_8+b_{8i})*Can\_d_{ij-1} + b_9*Alc\_p_i + b_{10}*Can\_p_i + b_{11\text{-}25}*Covariates + e_{ijk} \end{aligned} \quad \text{(S2)}$$

PA_{ijk} is the positive affect rating of person i on day j at occasion k. At a given occasion k an individual's *PA* score is fit as a function of a global intercept (b_0), the current and previous occasion's alcohol (*Alc*) use (b_1 and b_2), an individual's average alcohol use for that day (b_5) and

the previous day (b_6), and an individual's overall person-average of alcohol use across the entire diary period (b_9). There are also parallel occasion- (b_3 and b_4), day- (b_7 and b_8), and person-level (b_{10}) effects for an individual's cannabis use (*Can*). To consolidate the model, we excluded the covariates ($b_{11\text{-}25}$) from our data generation, as the reported alcohol and cannabis effects already explicitly adjusted for them, and their exclusion minimally affected degrees of freedom (and resulting standard errors) given the large number of observations.[1]

In general, reported coefficient estimates explicitly adjust for covariances between predictors within levels of analysis. However, we observed that co-use in Trull and colleagues' (2016) data was small (r = .09; see Supplemental Materials for details) relative to previous studies (rs = .10-.60; Hartman & Huestis, 2013; Walsh et al., 2004). Therefore we chose to consider it as a parameter to vary in the sensitivity analyses.[2]

The nine estimated random effects in Equation 2 (subscripted by i and/or j) were not reported by Trull and colleagues (2016); however, the lead author provided us with a table of the random effect variances and error variances (e_{ijk}). The authors did not estimate correlations between the random effects, so we constrained them to zero. We estimated missingness consistent with Trull et al. (2016).

Estimating Alcohol and Cannabis Co-Use Correlation

One of the model parameters of interest in the current simulations was the correlation between alcohol and cannabis use. Substance co-use is frequently observed in epidemiological

[1] Caution should be taken before adopting this approach if there are relatively few total observations, as the exclusion of the covariates can have a larger impact on the degrees of freedom in that case.

[2] The authors employed the disaggregation approach recommended by Curran and Bauer (2011); therefore we did not generate covariances between predictors at different levels of analysis, as the model estimates reflect independent effects and the residual variance adjusting for those covariances was provided. Researchers adopting such models should carefully consider whether to incorporate these additional parameters in their simulations, especially if they are interested in associations that occur at a particular level of analysis but that may positively covary with associations at another level.

and intensive longitudinal studies. It is also well-established that (positive) correlations between model predictors decreases unique variance - and therefore power - attributable to each predictor in relation to an outcome. Thus, naturally occurring substance co-use could undermine the power to detect unique effects. However, it is unclear to what extent this represents a substantive problem. Using the information provided by Trull and colleagues (2016), we were able to estimate an approximate value for the alcohol and cannabis co-use correlation observed in their study that we could then vary in sensitivity analyses to examine the relative impact on power. Their Table 1 provides frequencies and percentages for the amount of alcohol, cannabis, and combined use across days and occasions, which we can use to calculate a phi correlation coefficient (though acknowledging that we are ignoring the clustering of days within individuals). Given that there are 658 (24.9%) alcohol use days, 364 (13.9%) cannabis use days, and 123 (4.7%) co-use days, we can complete a 2×2 table of Yes/No substance use and estimate their correlation (Table S1). In doing so, we estimate a correlation of $r_{\Phi} = .08$. Similarly, the corresponding co-use correlation at the occasion level is estimated to be $r_{\Phi} = .09$. For our simulations we round these correlations to $r_{\Phi} = .10$.

Table S1. Estimation of daily co-use correlation from Trull et al. (2016)

Daily Co-Use

	Alcohol		
Cannabis	No	Yes	Total
No	1744	535	2279
Yes	241	123	364
Total	1985	658	2643

$$\phi = \frac{(1744 \times 123 - 535 \times 241)}{\sqrt{(2279 \times 364 \times 1985 \times 658)}} = \mathbf{.08}$$

Note. Shaded boxes are those provided in Trull et al. (2016) Table 1.

Example Sensitivity Analysis using SAS

Figures S1 and S2 provide SAS syntax for conducting an example sensitivity analysis. Here we present the sensitivity analysis for the full model described by Piasecki et al. (2005;

Figure S1) and Trull et al. (2016; Figure S2) and vary the number of participants in each sample, holding all other parameters constant. These analyses corresponds to the solid black lines in Figure 1A and 1B, respectively, and can be edited to reproduce the results reported in each of the figures. The syntax is annotated to facilitate interpretation, implementation, and editing for alternative models/parameters for interested readers.

Example Sensitivity Analysis using R

Figures S3 and S4 provide R syntax for conducting an example sensitivity analysis for the full model described by Piasecki et al. (2005; Figure S3) and Trull et al. (2016; Figure S4), respectively.

Additional References

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, *62*, 583-619.

Hartman, R. L., & Huestis, M. A. (2013). Cannabis effects on driving skills. *Clinical Chemistry*, *59*, 478-492.

Figure S1. SAS syntax for Piasecki et al. power analysis presented in Figure 1A.

```
*SET UP MACRO;
%macro power_piasecki(samples=1000,participants=486,years=6,ratioFHN=.5,VSUB_B3=0.0208); *default values
are those from Piasecki et al.;
DATA EXAMPLE1;
RETAIN SUBNUM ;
RETAIN SEED 20180726;
/*estimates for the model effects*/
RETAIN B0 1.4221; RETAIN B1 0.5265; RETAIN B2 0.4222; RETAIN B3 -0.0551; RETAIN B4 1.6181; RETAIN B5 -
0.0474; RETAIN B6 0.0501; RETAIN B7 -0.3006; RETAIN B8 -.5067;
/*estimates for the random effects*/
RETAIN VSUB_B0 2.7900; RETAIN VSUB_B3 &VSUB_B3; RETAIN VSUB_B4 0.7023;
/*estimate for the error/residual variance*/
RETAIN VRESID 1.4086;
RETAIN SAMPLES &samples;
DO SAMPLE=1 TO SAMPLES;
  DO SUBNUM=1 TO &participants;
    B0i=SQRT(VSUB_B0)*RANNOR(SEED);
      B3i=SQRT(VSUB_B3)*RANNOR(SEED);
      B4i=SQRT(VSUB_B4)*RANNOR(SEED);
      *now split things in order to specify 25/75, 50/50, 75/25;
      if SUBNUM le &participants*(1/2) then SEX=1; else SEX=0; *evenly split men and women;
      if SUBNUM le &participants*((1/4)+((&ratioFHN-.5)/2)) | (SUBNUM gt &participants*(1/2) & SUBNUM le
&participants*((3/4)+((&ratioFHN-.5)/2))) then FHP=0; else FHP=1; *split family history according to user;
      DO YEAR=0 TO 10;
            DRINK=SQRT(0.22)*RANNOR(SEED);
            hangover = (B0 + B0i) + (B1)*FHP + (B2)*SEX + (B3 + B3i)*YEAR + (B4 + B4i)*DRINK +
                        (B5)*FHP*YEAR + (B6)*SEX*YEAR + (B7)*FHP*DRINK + (B8)*SEX*DRINK +
SQRT(VRESID)*RANNOR(SEED);
    OUTPUT;
END; END; END;
DROP SEED b0 b1 b2 b3 b4 b5 b6 b7 b8 vresid vsub_b0 vsub_b3 vsub_b4 samples ; RUN;

*Generate missing data;
data example1; set example1;
rand=ranuni(20180726);
if rand>(1-0.017)**YEAR then hangover=.; *accumulating missing data rate by year;
if hangover=. then miss=1; else miss=0;
*set up separate dependent variables corresponding to different amounts of waves of data;
```

```
*can specify different years with missing data to maximize/minimize variability in the year variable;
if year=0|year=1|year=2|year=3|year=6|year=8|year=9|year=10 then hangover8=hangover; else hangover8=.;
if year=0|year=1|year=2|year=3|year=6|year=9|year=10 then hangover7=hangover; else hangover7=.;
if year=0|year=1|year=2|year=3|year=6|year=10 then hangover6=hangover; else hangover6=.;
if year=0|year=1|year=2|year=3|year=6 then hangover5=hangover; else hangover5=.;
if year=0|year=1|year=2|year=3 then hangover4=hangover; else hangover4=.;
run;

*Turn off all output;
ODS GRAPHICS OFF; ODS SELECT NONE;
*Analyze the simulated data using the specified model and save the parameter estimates;
PROC MIXED DATA=EXAMPLE1 COVTEST NOCLPRINT;
BY SAMPLE;
CLASS SUBNUM ;
  MODEL hangover&years = FHP SEX YEAR DRINK FHP*YEAR SEX*YEAR FHP*DRINK SEX*DRINK / S ddfm=kr;
  RANDOM INTERCEPT YEAR DRINK / SUBJECT=SUBNUM type=vc;
ODS OUTPUT SOLUTIONF=fixedeffects covparms=randomeffects; RUN;
*Turn all output back on;
ODS SELECT ALL; ODS GRAPHICS ON; ODS LISTING;

data fixedeffects; set fixedeffects;
sig=0;
if effect="Intercept" & estimate>0 & probt<=0.05 then sig=1;
if effect="FHP" & estimate>0 & probt<=0.05 then sig=1;
if effect="SEX" & estimate>0 & probt<=0.05 then sig=1;
if effect="YEAR" & estimate<0 & probt<=0.05 then sig=1;
if effect="DRINK" & estimate>0 & probt<=0.05 then sig=1;
if effect="FHP*YEAR" & estimate<0 & probt<=0.05 then sig=1;
if effect="SEX*YEAR" & estimate>0 & probt<=0.05 then sig=1;
if effect="FHP*DRINK" & estimate<0 & probt<=0.05 then sig=1;
if effect="SEX*DRINK" & estimate<0 & probt<=0.05 then sig=1;
run;

proc sort data=fixedeffects; by effect; run;
proc means data=fixedeffects noprint; by effect; var estimate stderr df sig; output out=power mean=; run;
data power; set power; N=&participants; years=&years; ratioFHN=&ratioFHN; randslope=&VSUB_B3; run;
proc export data = power file =
"C:\PATH\piasecki_n&participants._t&years._ratio&ratioFHN._slope&VSUB_B3..csv" dbms = csv replace; run;
%mend power_piasecki;

*Figure1A;
```

```
options nonotes nosource nosource2 errors=0;
/*50%*/
%power_piasecki(samples=10000,participants=486,years=4,ratioFHN=.5,VSUB_B3=0.0208); /*1944*/
%power_piasecki(samples=10000,participants=486,years=5,ratioFHN=.5,VSUB_B3=0.0208); /*2430*/
%power_piasecki(samples=10000,participants=486,years=6,ratioFHN=.5,VSUB_B3=0.0208); /*2916*/
%power_piasecki(samples=10000,participants=486,years=7,ratioFHN=.5,VSUB_B3=0.0208); /*3402*/
%power_piasecki(samples=10000,participants=486,years=8,ratioFHN=.5,VSUB_B3=0.0208); /*3888*/

%power_piasecki(samples=10000,participants=324,years=6,ratioFHN=.5,VSUB_B3=0.0208); /*1944*/
%power_piasecki(samples=10000,participants=405,years=6,ratioFHN=.5,VSUB_B3=0.0208); /*2430*/
*%power_piasecki(samples=10000,participants=486,years=6,ratioFHN=.5,VSUB_B3=0.0208); /*2916*/
%power_piasecki(samples=10000,participants=567,years=6,ratioFHN=.5,VSUB_B3=0.0208); /*3402*/
%power_piasecki(samples=10000,participants=648,years=6,ratioFHN=.5,VSUB_B3=0.0208); /*3888*/

/*25%*/
%power_piasecki(samples=10000,participants=486,years=4,ratioFHN=.25,VSUB_B3=0.0208); /*1944*/
%power_piasecki(samples=10000,participants=486,years=5,ratioFHN=.25,VSUB_B3=0.0208); /*2430*/
%power_piasecki(samples=10000,participants=486,years=6,ratioFHN=.25,VSUB_B3=0.0208); /*2916*/
%power_piasecki(samples=10000,participants=486,years=7,ratioFHN=.25,VSUB_B3=0.0208); /*3402*/
%power_piasecki(samples=10000,participants=486,years=8,ratioFHN=.25,VSUB_B3=0.0208); /*3888*/

%power_piasecki(samples=10000,participants=324,years=6,ratioFHN=.25,VSUB_B3=0.0208); /*1944*/
%power_piasecki(samples=10000,participants=405,years=6,ratioFHN=.25,VSUB_B3=0.0208); /*2430*/
*%power_piasecki(samples=10000,participants=486,years=6,ratioFHN=.25,VSUB_B3=0.0208); /*2916*/
%power_piasecki(samples=10000,participants=567,years=6,ratioFHN=.25,VSUB_B3=0.0208); /*3402*/
%power_piasecki(samples=10000,participants=648,years=6,ratioFHN=.25,VSUB_B3=0.0208); /*3888*/

/*75%*/
%power_piasecki(samples=10000,participants=486,years=4,ratioFHN=.75,VSUB_B3=0.0208); /*1944*/
%power_piasecki(samples=10000,participants=486,years=5,ratioFHN=.75,VSUB_B3=0.0208); /*2430*/
%power_piasecki(samples=10000,participants=486,years=6,ratioFHN=.75,VSUB_B3=0.0208); /*2916*/
%power_piasecki(samples=10000,participants=486,years=7,ratioFHN=.75,VSUB_B3=0.0208); /*3402*/
%power_piasecki(samples=10000,participants=486,years=8,ratioFHN=.75,VSUB_B3=0.0208); /*3888*/

%power_piasecki(samples=10000,participants=324,years=6,ratioFHN=.75,VSUB_B3=0.0208); /*1944*/
%power_piasecki(samples=10000,participants=405,years=6,ratioFHN=.75,VSUB_B3=0.0208); /*2430*/
*%power_piasecki(samples=10000,participants=486,years=6,ratioFHN=.75,VSUB_B3=0.0208); /*2916*/
%power_piasecki(samples=10000,participants=567,years=6,ratioFHN=.75,VSUB_B3=0.0208); /*3402*/
%power_piasecki(samples=10000,participants=648,years=6,ratioFHN=.75,VSUB_B3=0.0208); /*3888*/

options notes source source2 errors=20;
```

Figure S2. SAS syntax for Trull et al. power analysis presented in Figure 1B.

```
DATA STEP1;
RETAIN SUBNUM ;
RETAIN SEED 20170620;
RETAIN B0 2.33; *intercept value;
RETAIN B1 .12; *slope value; RETAIN B2 -.07; *slope value;
RETAIN B3 .02; *slope value; RETAIN B4 .02; *slope value;
RETAIN B5 .33; *slope value; RETAIN B6 -.17; *slope value;
RETAIN B7 .11; *slope value; RETAIN B8 .11; *slope value;
RETAIN B9 .41; *slope value; RETAIN B10 .58; *slope value;

RETAIN VB0i .11; *variance of intercept random effect for individuals;
RETAIN VB0j .31; *variance of intercept random effect for days within individuals;
RETAIN VB1 .02; *variance of the random slope for occasion drinking status;
RETAIN VB2 .02; *variance of the random slope for lag occasion drinking status;
RETAIN VB3 .08; *variance of the random slope for occasion pot status;
RETAIN VB4 .01; *variance of the random slope for lag occasion pot status;
RETAIN VB5 .07; *variance of the random slope for day drinking status;
RETAIN VB6 .04; *variance of the random slope for lag day drinking status;
RETAIN VB7 .34; *variance of the random slope for day pot status;
RETAIN VB8 .21; *variance of the random slope for lag day pot status;
RETAIN VRESID .29; *variance of the residual;
RETAIN VFREQ .054; *variance of individuals drinking frequency;

RETAIN SAMPLES 10000; *number of samples to generate;
ARRAY SIZE[5] (31 62 93 124 155); *sensitivity analysis varying the number of participants generated;
DO SSIZE=1 TO 5;
DO SAMPLE=1 TO SAMPLES;
  DO SUBNUM=1 TO SIZE[SSIZE];
    SUB_B0i = SQRT(VB0i)*RANNOR(SEED); *OUTCOME RANDOM EFFECT;
    SUB_B1 = SQRT(VB1)*RANNOR(SEED); *OCCASION DRINKING RANDOM EFFECT;
    SUB_B2 = SQRT(VB2)*RANNOR(SEED); *LAG OCCASION DRINKING RANDOM EFFECT;
    SUB_B3 = SQRT(VB3)*RANNOR(SEED); *OCCASION POT RANDOM EFFECT;
    SUB_B4 = SQRT(VB4)*RANNOR(SEED); *LAG OCCASION POT RANDOM EFFECT;
    SUB_B5 = SQRT(VB5)*RANNOR(SEED); *DAY DRINKING RANDOM EFFECT;
    SUB_B6 = SQRT(VB6)*RANNOR(SEED); *LAG DAY DRINKING RANDOM EFFECT;
    SUB_B7 = SQRT(VB7)*RANNOR(SEED); *DAY POT RANDOM EFFECT;
    SUB_B8 = SQRT(VB8)*RANNOR(SEED); *LAG DAY POT RANDOM EFFECT;
```

```
      ALC_P = RANNOR(SEED); *PERSON LEVEL DRINKING;
      CAN_P = RANNOR(SEED); *PERSON LEVEL SMOKING;
      DO DAY=1 TO 27;
        SUB_B0j = SQRT(VB0j)*RANNOR(SEED); *OUTCOME RANDOM EFFECT for day;
        *GENERATE DAILY DRINKING AND POT USE EVENTS;
        ALC_D=RAND('BERNOULLI',.25); LAG_ALC_D=RAND('BERNOULLI',.25);*25% of days Trull et al. - Table 1;
        CAN_D=RAND('BERNOULLI',ALC_D*.10+.115); LAG_CAN_D=RAND('BERNOULLI',LAG_ALC_D*.10+.115); *14% of
days from Trull et al. (w/ r~.10) - Table 1;
        DO MEASUREMENT=1 TO 6;
            *GENERATE DRINKING AND POT USE EVENTS;
            IF ALC_D=1 THEN ALC_O=RAND('BERNOULLI',.28); ELSE ALC_O=0;
            IF LAG_ALC_D=1 THEN LAG_ALC_O=RAND('BERNOULLI',.28); ELSE LAG_ALC_O=0; *7% (25% of days * 28%
of occasions on drinking days from Trull et al. - Table 1;
            IF CAN_D=1 THEN CAN_O=RAND('BERNOULLI',ALC_O*.10+.40); ELSE CAN_O=0;
            IF LAG_CAN_D=1 THEN LAG_CAN_O=RAND('BERNOULLI',LAG_ALC_O*.10+.40); ELSE LAG_CAN_O=0; *6% of
occasions from Trull et al. (w/ r~.10) - Table 1;
            *NOW GENERATE MODEL;
            POS = (B0 + SUB_B0i + SUB_B0j)
                  + (B1 + SUB_B1)*ALC_O
                  + (B2 + SUB_B2)*LAG_ALC_O
                  + (B3 + SUB_B3)*CAN_O
                  + (B4 + SUB_B4)*LAG_CAN_O
                  + (B5 + SUB_B5)*ALC_D
                  + (B6 + SUB_B6)*LAG_ALC_D
                  + (B7 + SUB_B7)*CAN_D
                  + (B8 + SUB_B8)*LAG_CAN_D
                  + (B9)*ALC_P
                  + (B10)*CAN_P
                  + SQRT(VRESID)*RANNOR(SEED);
        OUTPUT;
 END;  END;  END;  END;  END;
 DROP SEED b0 b1 b2 b3 b4 b5 b6 b7 b8 b9 b10 vb0i vb0j vb1 vb2 vb3 vb4 vb5 vb6 vb7 vb8 vresid samples ;
 RUN;
DATA STEP1; SET STEP1;
if ssize=1 then ssize=size1; else if ssize=2 then ssize=size2; else if ssize=3 then ssize=size3; else if
ssize=4 then ssize=size4;
else ssize=size5;
DROP size1--size5;
run;

*generate random missing data;
```

```
data step1; set step1;
rand=ranuni(20170620);
POS_miss=POS; if rand ge .892 then POS_miss=.; *corresponds to ~90.5% data from Trull et al.;
if POS_miss=. then count=0; else count=1;
run;

*estimate observed power for all parameters;
ODS GRAPHICS OFF; ODS SELECT NONE;
PROC MIXED DATA=STEP1 COVTEST NOCLPRINT; BY SSIZE SAMPLE; CLASS SUBNUM ;
MODEL POS_miss = ALC_O LAG_ALC_O CAN_O LAG_CAN_O ALC_D LAG_ALC_D CAN_D LAG_CAN_D ALC_P CAN_P / S
DF=.,.,.,.,.,46,.,.,.,.; *use degrees of freedom from Trull et al. for effect of interest – others are
freely estimated;
RANDOM INTERCEPT ALC_O LAG_ALC_O CAN_O LAG_CAN_O ALC_D LAG_ALC_D CAN_D LAG_CAN_D / SUBJECT=SUBNUM type=vc;
RANDOM INTERCEPT / SUBJECT=DAY(SUBNUM) type=vc;
ODS OUTPUT SOLUTIONF=fixedeffects COVPARMS=randomvariances; RUN;
ODS SELECT ALL; ODS GRAPHICS ON; ODS LISTING;

data power; set fixedeffects; sig=0;
if effect='Intercept' & probt le .05 & estimate gt 0 then sig=1;
if effect='ALC_O' & probt le .05 & estimate gt 0 then sig=1;
if effect='LAG_ALC_O' & probt le .05 & estimate lt 0 then sig=1;
if effect='CAN_O' & probt le .05 & estimate gt 0 then sig=1;
if effect='LAG_CAN_O' & probt le .05 & estimate gt 0 then sig=1;
if effect='ALC_D' & probt le .05 & estimate gt 0 then sig=1;
if effect='LAG_ALC_D' & probt le .05 & estimate lt 0 then sig=1;
if effect='CAN_D' & probt le .05 & estimate gt 0 then sig=1;
if effect='LAG_CAN_D' & probt le .05 & estimate gt 0 then sig=1;
if effect='ALC_P' & probt le .05 & estimate gt 0 then sig=1;
if effect='CAN_P' & probt le .05 & estimate gt 0 then sig=1;
run;
proc sort data=power; by SSIZE effect; run; proc means data=power noprint; by SSIZE effect; var estimate
sig; output out=powermeans mean=; run;
proc export data = powermeans file = "C:\PATH\VARY_people25.csv" dbms = csv replace; run;
```

Figure S3. R syntax for Piasecki et al. power analysis presented in Figure 1A.

```
#############################################################################
#packages we will use
install.packages("arm")
library(arm)
install.packages("lme4")
library(lme4)
install.packages("lmerTest")
library(lmerTest)
#############################################################################
# power analysis

#generate data
Ex1.fake <- function (J,K){
 YEAR <- rep(seq(0,K-1,length=K),J) # K measurements per person
 person <- rep(1:J, each=K)      # J person IDs
 SEX <- rnorm(J*K,0,.5)         # Sex
 FHP <- rnorm(J*K,0,.5)         # Family History Positive
 DRINK <- rnorm(J*K,0,.22)         # Heavy Drinking
 #fixed effects
 b0 <-  1.4221    # true intercept value
 b1 <-   .5265    # true FHP estimate
 b2 <-   .4222    # true SEX estimate
 b3 <-  -.0551    # true linear slope estimate
 b4 <-  1.6181    # true heavy drinking estimate
 b5 <-  -.0474    # true year * FHP estimate
 b6 <-   .0501    # true year * SEX estimate
 b7 <-  -.3006    # true DRINK * FHP estimate
 b8 <-  -.5067    # true DRINK * SEX estimate
 #random effects
 vsub.b0 <- 2.7900    # true between person variance in the intercept (i.e. hangover)
 vsub.b3 <- .0208    # true between person variance in the YEAR slope
 vsub.b4 <- .7023    # true between person variance in the DRINK slope
```

```
 vresid <-  1.4086    # true within person variance in hangover
 #combine fixed and random effects per person
 b0.int    <- rnorm(J,b0,sqrt(vsub.b0))  #generate an intercept for each person
 b3.YEAR <- rnorm(J,b3,sqrt(vsub.b3))  #generate a slope for YEAR for each person
 b4.DRINK    <- rnorm(J,b4,sqrt(vsub.b4))  #generate a slope for DRINK for each person
 hang <- rnorm(J*K, b0.int[person]        #use the person's intercept
               +b1*FHP              #FHP effect
               +b2*SEX              #SEX effect
               +b3.YEAR[person]*YEAR   #average plus person-specific yearly slope effect
               +b4.DRINK[person]*DRINK #average plus person-specific DRINK effect
               +b5*YEAR*FHP          #average YEAR by FHP effect
               +b6*YEAR*SEX          #average YEAR by SEX effect
               +b7*DRINK*FHP         #average DRINK by FHP effect
               +b8*DRINK*SEX         #average DRINK by SEX effect
               ,sqrt(vresid))        #residual
 return(data.frame(person,YEAR,FHP,SEX,DRINK,hang))
}

data <- Ex1.fake(486,6) #generate an example dataset to make sure it worked
lme.power <- lmer(hang ~ (1 | person) + (-1 + YEAR | person) + (-1 + DRINK | person)
             + FHP + SEX + YEAR + DRINK + FHP*YEAR + SEX*YEAR + FHP*DRINK + SEX*DRINK,
             data=data)
summary(lme.power)

# this function loops the data generation and analyzes it
Ex1.power <- function (J,K,n.sims=1000){     #default to 1000 simulations if not specified
 signif <- rep(NA,n.sims)                  # a vector that will record if the effect of interest is sig
 for (s in 1:n.sims){
  fake <- Ex1.fake(J,K)                  #generate a fake dataset
  lme.power <- lmer(hang ~ (1 | person) + (-1 + YEAR | person) + (-1 + DRINK | person)
               + FHP + SEX + YEAR + DRINK + FHP*YEAR + SEX*YEAR + FHP*DRINK + SEX*DRINK, data=fake) #analyze it
  est <- fixef(lme.power)["FHP:YEAR"]      #save the parameter estimate
  se <- se.fixef(lme.power)["FHP:YEAR"]    #save the standard error
  signif[s] <- (abs(est)-2*se)>0      #calculate significance - returns TRUE/FALSE
```

```
  }
  power <- mean(signif)
  return(power)
}

Ex1.power(J=486,K=6,n.sims=10000)

##############################################################################

## create a wrapper to incorporate multiple scenarios (i.e. sensitivity analyses) ##

## Specify sample sizes to examine ##
Jvals <- c(324, 405, 486, 567, 648)
Kvals <- c(4, 5, 6, 7, 8)
sample.sizes <- expand.grid(Jvals, Kvals)
names(sample.sizes) <- c("J", "K")
## To hold results ##
powvals <- rep(NA, nrow(sample.sizes))
## Calculate power for each combination of J, K ##
for(i in 1:nrow(sample.sizes)){
  ## Get the next set of sample sizes and calculate power ##
  tmpss <- as.numeric(sample.sizes[i,])
  tmppow <- Ex1.power(tmpss[1], tmpss[2], 50)
  powvals[i] <- tmppow
}

powvals

library(ggplot2)
ggplot(sample.sizes, aes(x=J, y=powvals)) + geom_line() + facet_wrap(~ K) +
  xlab("Participant Sample Size") + ylab("Power")
```

Figure S4. R syntax for Trull et al. power analysis presented in Figure 1B.

```
################################################################################
#packages we will use
install.packages("arm")
library(arm)
install.packages("lme4")
library(lme4)
install.packages("lmerTest")
library(lmerTest)
################################################################################
# power analysis

#generate data
Ex2.fake <- function (J,K,L){
  time <- rep(seq(1,K*L,length=K*L),J) # K measurements per person
  person <- rep(1:J, each=K*L)      # J person IDs
  alc.o <- rnorm(J*K*L,0,.08)       # occasion level alcohol
  lalc.o <- rnorm(J*K*L,0,.08)       # occasion level lag alcohol
  can.o <- rnorm(J*K*L,0,.08)       # occasion level cannabis
  lcan.o <- rnorm(J*K*L,0,.08)      # occasion level lag cannabis
  alc.d <- rnorm(J*K*L,0,.25)       # day level alcohol
  lalc.d <- rnorm(J*K*L,0,.25)       # day level lag alcohol
  can.d <- rnorm(J*K*L,0,.24)       # day level cannabis
  lcan.d <- rnorm(J*K*L,0,.24)      # day level lag cannabis
  alc.p <- rnorm(J*K*L,0,1)        # person level alcohol
  can.p <- rnorm(J*K*L,0,1)        # person level cannabis
  #fixed effects
  b0 <- 2.33    # true intercept value
  b1 <-  .12    # true occasion level alcohol
  b2 <- -.07    # true occasion level lag alcohol
  b3 <-  .02    # true occasion level cannabis
  b4 <-  .02    # true occasion level lag cannabis
  b5 <-  .33    # true day level alcohol
```

```
 b6 <- -.17    # true day level lag alcohol
 b7 <-  .11    # true day level cannabis
 b8 <-  .11    # true day level lag cannabis
 b9 <-  .41    # true person level alcohol
 b10 <- .58    # true person level cannabis
 #random effects
 vsub.b0 <- .11    # true between person variance in the intercept (i.e. positive affect)
 vsub.b1 <- .07    # true between person variance in the occasion alcohol slope
 vsub.b6 <- .04    # true between person variance in the day lagged alcohol slope
 vresid <-  .29    # true within person variance in positive affect
 #combine fixed and random effects per person
 b0.int    <- rnorm(J,b0,sqrt(vsub.b0))      #generate an intercept for each person
 b1.alc.o <- rnorm(J,b1,sqrt(vsub.b1))       #generate a slope for occasion alcohol for each person
 b6.lalc.d    <- rnorm(J,b6,sqrt(vsub.b6))    #generate a slope day lag alcohol for each person
 pa <- rnorm(J*K*L, b0.int[person]          #use the person's intercept
              +b1.alc.o[person]*alc.o    #occasion alcohol effect
              +b2*lalc.o              #occasion alcohol lagged effect
              +b3*can.o               #occasion cannabis effect
              +b4*lcan.o              #occasion cannabis lagged effect
              +b5*alc.d               #day alcohol effect
              +b6.lalc.d[person]*lalc.d  #day lagged alcohol effect
              +b7*can.d               #day cannabis effect
              +b8*lcan.d              #day lagged cannabis effect
              +b9*alc.p               #person alcohol effect
              +b10*can.p              #person cannabis effect
              ,sqrt(vresid))          #residual
  return(data.frame(person,time,alc.o,lalc.o,can.o,lcan.o,alc.d,lalc.d,can.d,lcan.d,alc.p,can.p,pa))
}

data <- Ex2.fake(93,27,6) #generate an example dataset to make sure it worked
lme.power <- lmer(pa ~ (1 | person) + (-1 + alc.o | person) + (-1 + lalc.d | person)
             + alc.o + lalc.o + can.o + lcan.o + alc.d + lalc.d + can.d + lcan.d + alc.p + can.p,
             data=data)
summary(lme.power)
```

```
# this function loops the data generation and analyzes it
Ex2.power <- function (J,K,L,n.sims=1000){     #default to 1000 simulations if not specified
  signif <- rep(NA,n.sims)                  # a vector that will record if the effect of interest is sig
  for (s in 1:n.sims){
    fake <- Ex2.fake(J,K,L)                 #generate a fake dataset
    lme.power <- lmer(pa ~ (1 | person) + (-1 + alc.o | person) + (-1 + lalc.d | person)
                  + alc.o + lalc.o + can.o + lcan.o + alc.d + lalc.d + can.d + lcan.d + alc.p + can.p, data=fake) #analyze it
    est <- fixef(lme.power)["lalc.d"]     #save the parameter estimate
    se <- se.fixef(lme.power)["lalc.d"]    #save the standard error
    signif[s] <- (abs(est)-2*se)>0     #calculate significance - returns TRUE/FALSE
  }
  power <- mean(signif)
  return(power)
}

Ex2.power(J=93,K=27,L=6,n.sims=10000)

#############################################################################

## create a wrapper to incorporate multiple scenarios (i.e. sensitivity analyses) ##

## Specify sample sizes to examine ##
Jvals <- c(31, 62, 93, 124, 155)
Kvals <- c(9, 18, 27, 36, 45)
sample.sizes <- expand.grid(Jvals, Kvals)
names(sample.sizes) <- c("J", "K")
## To hold results ##
powvals <- rep(NA, nrow(sample.sizes))
## Calculate power for each combination of J, K ##
for(i in 1:nrow(sample.sizes)){
  ## Get the next set of sample sizes and calculate power ##
  tmpss <- as.numeric(sample.sizes[i,])
  tmppow <- Ex2.power(tmpss[1], tmpss[2],6,50)
```

```
  powvals[i] <- tmppow
}

powvals

library(ggplot2)
ggplot(sample.sizes, aes(x=J, y=powvals)) + geom_line() + facet_wrap(~ K) +
  xlab("Participant Sample Size") + ylab("Power")
```