Family, Life, Activity, Sun, Health, and Eating Survey (FLASHE) Study

Webinar on Dyadic Data Analysis

cancercontrol.cancer.gov/flashe

Webinar Questions & Answers

Q. What was the purpose of the FLASHE study?

A. One of the biggest motivators for this study was to understand cancer-preventive health behaviors from a conceptually multilevel approach. We did a scan of the literature and there were not any public datasets that would allow researchers to assess these different behaviors from both the parent and child perspective. One of the main drivers of this project was to fill that gap and provide researchers with the data they need to answer unique research questions related to cancer control and cancer prevention.

Q. How do we know whether these relationships are statistically significant?

A. It turns out that all the paths in the Actor-Partner Interdependence Model (APIM) model that we presented were statistically significant. We didn't emphasize significance primarily because with a sample size of about 1,480 couples, even tiny correlations will turn out to be significant. However, all of those are statistically significant at the .05 level.

Q. If a teen's partner effect is not statistically significant, does it mean that the teens did not influence their parents for the outcome of interest?

A. No, I don't think you could draw that conclusion. There are many variables that could affect the outcome of interest and we are just showing you self-efficacy. Although in this case, it's not a problem, if you have smaller sample sizes, focusing just on the significance could be misleading. These path decomposition methods that we've been using will only work if you use the actual coefficients in the model. So you would leave all the coefficients in the model. If some are not significant, then I wouldn't trim that model. Instead of significance I would focus more on the confidence interval for that effect. You could have nonsignificant effects but their confidence intervals can include values that are large enough that you would not want to ignore them. So you can have a nonsignificant effect but it can still be consistent with a much larger effect. You just don't know.

Q. Why can I not just run separate regression models? What added value am I getting by doing an actor-partner interdependence model?

A. There are four types of effects that we talked about (two actor effects and two partner effects). The only way that we can actually get those is by having dyadic data and estimating a model where both the teen and the parent outcomes are simultaneously part of the model. That's the only way that we could get these effects because these effects are dyadic effects. If you were to do a regression model with the just the teen outcome (teen fruit and vegetable consumption) regressed on teen and parent self-efficacy, you would get the R^2 in terms of the variance explained in the teen outcome, a kind of teen actor-effect and parent-partner effect. However, you wouldn't, technically speaking, get the right actor effect. If you want to also get the corresponding effects for the parent, we will need to include the

parents' outcome. We are also trying to explain the covariation among dyad members. If we don't have both dyad members' outcomes in the same model, we cannot explain this covariation in an efficient way, so it is suboptimal. It may not be incorrect necessarily, but it's suboptimal in terms of making the most of dyadic data that we've collected.

Q. What benefit does this APIM approach have over other methods such as a repeated-measures model?

A. If you run what I now consider the modern version of repeated-measures analyses, which are in the mixed level or multilevel modeling tradition, you will get the same results. In the FLASHE Dyadic Analysis User's Guide and Sample Code we show how to do the analyses both ways -- using the equivalent of a repeated-measures analysis and using a structural equation modeling (SEM) analysis -- and you get the same results. Depending on what you're familiar with or what you are used to, you might find it easier to work with the SEM tradition. We have established, and others have too, that structural equation modeling and multilevel modeling in certain cases are the same. They're just different ways of partitioning variance.

Q. Can you compare what the differences would be if you were just using the teen data to control for the parent data? If you were just looking at one outcome, and you were using teen variables to control, what would you lose insight-wise?

A. I think it comes back to covariance. If you want to understand why is it that when a teen eats more fruit and vegetables his or her parent also tends to eat more fruits and vegetables (why these things go together), then you need to have a model that estimates both outcomes simultaneously. Again, it wouldn't be incorrect to (if you're focusing on the teen outcome but you control for the parent) put in efficacy of the parent as well as efficacy of the teen However, it wouldn't get at this idea of how much of this relationship we are explaining -- this covariation, this correlation between the fruit and vegetable consumption of the teen and parent. That is a different research question.

Q. With the number of tests you've done, do you have to be concerned about the multiple contrasts and controlling for that issue of the multiple comparisons?

A. I don't think so -- for a model like this. Everything we've drawn your attention to are fundamental parameters of the actor-partner interdependence model and things that we ought to have expectations about and planned to look at ahead of time. I don't think in this case I would worry about multiple comparisons. Obviously if you have a very complex model with several independent variables and perhaps mediation variables and dependent variables, then you might be in the situation where multiple comparisons are of concern. If you've looked at 100 of these coefficients, statistical theory tells us that five of those will be significant just by chance. I wouldn't be worried about it in this particular example or in a model that has perhaps maybe one or two more independent variables.

Q. Is parent-teen sex a moderator? If yes, how can you deal with that?

A. That was something that crossed our minds as we worked with this dataset -- are some of these effects that we see here in the APIM different depending on the kind of pairing of

parent-teen? In this case, we have identified a 2-by-2 crossing of male and female parents, and male and female teens. We have these four different kinds of dyads and it is very possible to see whether these effects differ by type of dyad. We didn't show that here because of time restrictions, however, it's exactly this kind of analysis that we explain in the FLASHE Dyadic Analysis Data User's Guide and Sample Code. It would be a moderation analysis within the context of this APIM. Learn more on the FLASHE website.

Note: We have developed sample code for users to learn from, and it is available on our FLASHE website. The sample code went through internal review, and we hope it's helpful to you as you learn more about how to analyze dyadic data.

Q. Why are you converting the response scale for the fruit and vegetable outcome from the 1-to-5-point scale to a 1-to-10-point scale? Is that something that's standard for dyadic analysis or is that something that is specific to this dataset?

A. We could have left the variables on their original scales. Both parents and teens got essentially the same questions. The efficacy item was on a 1-to-6-point scale. The fruit and vegetable consumption was on a 1-to-5-point scale. If you look at standardized results, the actual scaling doesn't matter. We've been influenced by a paper by Cohen, et al.* When we standardize things, we are subtracting the mean and we are dividing by the standard deviation. That's an ideal thing to do when you have normally distributed data, If you don't, I think it's interesting to create a dimension on which zero means the lowest and 10 means the highest. It doesn't have to be 0 to 10 -- it could be 0 to 1, it could be 0 to 100 or any other interval -- but this is an alternative to standardizing the variables perhaps before you start on the analysis. As it turns out, it doesn't matter whether you look at the data standardized or unstandardized. The significance of the tests is the same. It's a matter of choice what you do. And you notice when we were doing the path decomposition, we used the standardized results. I think there will be occasions where you're dealing with variables that are things like blood-pressure, and these are things that have meanings in their original scale. You may not want to standardize those and so you'll be working with the unstandardized estimates. You will see in the FLASHE Dyadic Analysis User's Guide and Sample Code that we show you the Mplus results in both forms. I think what's most important is if the verbal description of the effects (for example, the actor effect) is more clear and is communicated better, then that is the one you should use. The tests of significance are always the same. It's not required for dyadic data analysis.

*Cohen J, Cohen P, Aiken LS, West SG. The problem of units and the circumstance for POMP. *Multivariate Behavioral Research*. 1999;34:315-346.

If you have additional questions, please email nciflashe@mail.nih.gov.