# Career Awards, Big Data, mHealth, and Causal Inference

with Donna Coffman

May 5, 2015

In this podcast, we talk with Methodology Center Investigator [Donna Coffman](https://methodology.psu.edu/people/dcoffman) about the K01 award she received from the National Institutes of Health (NIH) [Big Data to Knowledge (BD2K) initiative](http://datascience.nih.gov/bd2k). Topics include the emergence of big data in NIH-funded research and the practical aspects of applying for a K award. Donna also explains her research on analyzing biosensor data from a parenting study and how her move into big data integrates with her research on causal inference.

Speaker 1: The Methodology Center Perspective podcast is brought to you by The Methodology Center at Penn State, your source for cutting edge research methodology in the social, behavioral, and health sciences.

Aaron: Hello, and welcome to Methodology Minutes. With us today is Donna Coffman, Senior Research Associate at The methodology Center and Research Associate Professor in the College of Health and Human Development. Last fall, Donna received a K award from the National Institutes of Health, and she agreed to come and sit down and talk to us about it. Donna, welcome back to Methodology Minutes.

Donna: Thank you, Aaron.

Aaron: So to get started, could you please explain for anyone who may not already know, what is a K award?

Donna: Well, there are different types of K awards. This one in particular is a K01, which is an early career award. There's also mid-career awards and senior career awards. But the early career award, the K01, has both a training and a research component, and it is mentored. So my mentors are Runze Li, who's in the Department of Statistics here at Penn State; Vasant Honavar, who is in the College of Information Sciences and Technology; and Joshua Smyth, who's in the Department of Biobehavioral Health. The other thing is that my K is unique in that it is through the Big Data To Knowledge initiative, which is abbreviated BD2K, is a trans-institute NIH initiative. So it cuts across all of the National Institutes of Health, it cuts across all of the Institutes. It's basically to promote software development and training in biomedical big data research.

Aaron: And so what will you research as part of this award?

Donna: Well, my goal is to develop and apply big data methods to promote and maintain health behavior change, which will have broad implications for public health, particularly for the development of adaptive individualized health behavior interventions, which can then deliver a specific intervention at the specific moment when it's needed, and increase the efficiency and effectiveness of interventions while decreasing participant burden.

Aaron: So when you say deliver a specific intervention at the moment that it's needed, could you give us an example of what that might entail?

Donna: Yeah, sure. So for example, if the participants or patients who are wearing a biosensor band that perhaps detected electrodermal activity and detected that they were stressed, then that may send a signal to their smart phone, which would then chime and lead them through an intervention that was tailored specifically to them in that moment.

Aaron: So we're leveraging off of emergent mobile technologies, basically, in order to improve people's health.

Donna: Correct.

Aaron: And you said this was a trans-NIH initiative. Why do you think NIH believes that this program, and more specifically your work, is valuable?

Donna: Well, so NIH has been moving more towards mHealth and using digital technology to promote and maintain health behavior, particularly in OBSSR. Wendy Nelson in particular has been promoting the use of digital technologies and wearable sensors and mobile applications to promote health.

Aaron: Great. And for anyone listening that doesn't know, OBSSR is the Office of Behavioral and Social Science Research.

Donna: Correct. Which is also trans-institute.

Aaron: Yeah. And really cool. And if you don't know who they are, you should check them out. So there's a training component to a K award. What skills will you be learning?

Donna: So in this particular K, as I said, this was a special funding announcement through the BD2K initiative. So there are training goals that are related to each of the three areas of data science, which are computer science, statistics, and biomedical research. And in my case, biomedical research is broadly defined, but in my case it's specifically focused on health behavior. So basically I'm going to be learning the basics of databases and Python and R and those sorts of high performance computing skills. What's most important in a K is that the training goals relate to the research aims. So basically, what you propose in the training component needs to facilitate your research aims. So basically, the types of data that I will be working with in my research include genomic data, data from wearable sensors, and data from ecological momentary assessments and ecological momentary interventions. So I'm currently taking a course in statistical analysis of genomics data, last semester I took a course in functional data analysis. Both of those were here at Penn State, through the Department of Statistics. But I've also taken some other courses through these MOOCs online. For example, I'm doing two self-paced courses right now: introduction to databases, which is through Stanford online, and Harvard's CS50 through edX.

Aaron: So it sounds like you're gathering a lot of big data methods and a lot of hands-on programming skills. For anyone who may not know, what is your background?

Donna: My background is in quantitative psychology. I have a strong background in behavioral science, and I also have a fairly strong background in applied statistics. Part of my training also involves more training in theoretical statistics. So for example, I'm taking stochastic processes. So the skills that I lacked, or were not as strong in, were the computer science and the theoretical statistics. So that is another thing about a K, for anyone who's thinking about writing one, either for a special call or just the parent K01 mechanism, is that you have to argue that there is some area in which you're lacking skills that you need more training in, so that you can carry out the research that you want to do.

Aaron: You talked a little bit about what you've been studying. Has it affected your research? And how so?

Donna: Well, so for example, last semester, I took functional data analysis, and I'm going to be applying that method to some data with a collaborator at NYU, Noelle Leonard. It's called the Power Source Parenting Project, and in that study, adolescent single mothers who live in group homes in Massachusetts, they wear biosensors. And the example that I gave before actually was drawn from that study. The biosensor detects their electrodermal activity or galvanic skin response, and when it goes above a threshold, then it sends a signal to their smart phone, which then chimes and then initiates an app called Calm Mom. And then the app asks them some questions about the context that they're in, and it makes suggestions based on the context. So for example, if it's about a fight with their boyfriend, then it might suggest that they leave the situation. But if the source of the stress is their baby, it's not gonna suggest that they leave the situation. So it does adapt to the context, to some degree, but once you go past that point, then it suggests deep breathing and cool thoughts and things like this. So the biosensor actually measures electrodermal activity four times per second, and the girls wear the band anywhere from one to three months. Basically, that's a lot of data.

Aaron: Yes.

Donna: I've already begun to use the programming skills and some of the methods that I learned in this course on functional data analysis. Basically, the goal is to examine how quickly the girls return to their baseline electrodermal activity following the intervention. In other words, that's a measure of their stress reactivity. But we're also going to look at how high does the peak go initially.

Aaron: Yeah, okay. So at The Methodology Center, your research project here, which you just alluded to in terms of the Center Grant, is on causal inference. And how does the work you are doing for your K tie into your research on causality?

Donna: So to date, a lot of the algorithms that have been used in the analysis of big data have mostly focused on prediction, rather than studying mechanisms, or to infer causality. So for recommending movies and books, prediction is really the goal. However, in science, the goal is to understand the underlying processes. So causal inference is important, regardless of the size of the data. My research component on the P50 Center Grant is on causal inference, and when I wrote the K, I tied that in, in that eventually I want to get to the point of bringing causal inference into the big data methods and algorithms.

Aaron: You referred to books and movies. I assume you're talking about big data methods being put into action by places like Netflix and Amazon, is that right?

Donna: Right, yeah.

Aaron: And they don't need to worry about cause, they just need to know what move you're gonna watch. So why are you moving towards big data with your research? And what do you think the impact of big data will be on the social, behavioral, and health sciences in the future?

Donna: Well, I think that the idea that wearable sensors or smart phone apps can help people change and maintain healthy behaviors is exciting. I moved in this direction cause I felt excited about the prospect of being able to deliver what are called just in time adaptive interventions. In particular, I was inspired by Susan Murphy's work on the topic. And I also like that I am working more directly on applied problems than I was previously. Causal inference can be a bit philosophical. So in other words, i like that I have an applied problem that I'm very excited about, and that's driving the methodological intervention, rather than the other way around. And then as far as what do I think the impact of big data will be on the social, behavioral, and health sciences in the future, I think that big data has already impacted the behavioral and health sciences. Through the Brain Initiative and most recently President Obama's Precision Medicine Initiative ... Recently, he announced the Precision Medicine Initiative, which involves combining electronic medical record data, genomics data, data from wearable sensors, as well as other available sources, to provide individuals with optimal tailored therapies, treatments, and prevention interventions. So I think that this is already happening.

Aaron: Yeah, it's very exciting. So jumping back to the topic of Ks more generally again, what is the application process for a K like?

Donna: This is true of both the BD2K K01 that I did and also the parent K01. The application itself is 12 pages, and six of those are devoted to the training plan, and six to the research plan. In addition, the specific aims count as a separate page. The application includes several sections that are not usually a part of an NIH grant application. So for example, it must include a mentoring plan; three letters of reference, and these are not from your mentors, these are just outside recommendation letters; and a one page description of training in responsible research; and letters of support from the mentors and from the applicant's institution. In addition to the usual facilities and environment description that is typically submitted with NIH grants, there's also a separate one page description of the environmental and institutional commitment to the candidate. So the K requires a lot more sections than the usual R mechanisms.

Aaron: And who should apply for a K award?

Donna: Well, as I said before, the purpose of a K is to provide training for those who want to change directions in their research career, but when applying for a K, you basically have to argue that you are moving into a new area of research, but that your prior experience and research have led you to this new area, so that it is the next logical step in your career, but you need some additional training to get you there. So those who should apply for this mechanism is those who are in that situation, where they feel like they just need a little bit of training to get them to the point where they can take their research into the direction that they want.

Aaron: And what do you think the most important attribute of your K application was, in terms of getting funded?

Donna: I think that the mentors are the most important attribute of the K application, actually. As long as the applicant is clearly good and has a record of publishing, then I think that a lot of it probably comes down to who your mentors are, and whether they have NIH funding or have had NIH funding in the past. That's really important.

Aaron: What did you learn from the application process?

Donna: One of the sections that's part of the training plan in the application is on career goals and objectives. And writing that section has given new meaning to my work, and really clarified what I want in terms of a career. Once it was clear to me what my primary goals and objectives are, then I began eliminating those things that are not in service to those goals and objectives. But I really learned a lot about what I want to be in three to five years, and who I want to be.

Aaron: That's really cool, actually. It sounds like anyone should sit down and write their career objectives out, whether or not they're applying for a K. Anyhow, Donna, thank you very much for taking time to talk to us today, and we all really look forward to seeing more fruits of your work from your K. Thanks.

Donna: Thank you, Aaron.

Speaker 1: You have been listening to the Methodology Center Perspective podcast, brought to you by The Methodology Center at Penn State, your source for cutting edge research methodology in the social, behavioral, and health sciences. This podcast is available on The Methodology Center's website at Methodology.PSU.edu.