Rigor and Reproducibility: Back to Basics

>> Patricia Valdez: Hi, and welcome to the session on Rigor And Reproducibility: Back To Basics. My name is Patricia Valdez. I'm from the NIH Office of Extramural Research where I serve as the extramural research integrity officer. So today I'm going to talk to you about many of the things that NIH has implemented over the past couple of years to enhance reproducibility through rigor and transparency. We're going to talk about a couple of things today. So first we're going to talk about, why are concerned about reproducibility? Then I will talk about the NIH response. Then I'll talk about the updates to grant applications. Now these updates went into effect in 2016, and then we had a little tweaking in 2019, so I'll let you know what those are. And then finally, we're going to talk about some training and resources. Okay, so this is a figure that was taken from a nature commentary where they surveyed ... This was several years ago. They surveyed researchers or readers asking them, do they believe there was a reproducibility crisis? And as you can see here where they surveyed 1,500 researchers. A large portion of them, if you see in red and kind of the light red, a large portion of them did think that there was a significant crisis or a slight crisis. So over half thought it was significant. So at that time there was a lot of media that was coming at us, and so this slide shows you some of those issues that came to the forefront. Now we saw lots of articles that were published. This was done in the community and multiple publications across different research areas, though it was really focused on preclinical research. A lot of the attention started when a couple of pharmaceutical companies, Amgen and Bayer, were unable to reproduce some of the scientific literature, and so they published these papers indicating that many of the studies that they tried to reproduce, they were not able to. Although I will give you one caveat is that they didn't describe to us what papers they tried to reproduce, or what they did to try to reproduce, so there's a little bit of a lack of transparency there, but this caught everyone's attention, including the NIH, I would say. But actually when we look back a couple of years ago, NIH has been concerned about this issue of transparency and reporting and rigorous research for a couple of years. Now I'm just going to tell you this story that NINDS experienced and NIH experienced several years before a lot of this reproducibility, quote-unquote, crisis came into everyone's ... The media and everyone became aware of it. So back in 2002, there were several studies that were published that reported that this drug named minocycline would increase the survival of SOD1 mice. These mice serve as a model for ALS. Now based on these positive findings, NINDS funded a clinical trial in which hundreds of ALS patients were treated with minocycline for about 9 months, and in the end they found that there was no benefits, nothing. It was just absolutely no benefits, whereas increased survival was found in mice. So that was published in 2007. So NINDS decided to take a look back, and they were concerned. How could this data that they thought looked great, and they ended up funding a clinical trial ... How could that have failed so miserably? So they looked back on those papers, the preclinical studies in mice, again, and what they found is that in many of the cases, those papers didn't report blinding. They didn't report whether or not the studies were randomized, and in some cases they didn't report the sex of the mice that were being used in the studies, and also they didn't report ... Some of them didn't report the transgene copy numbers in the SOD1 mice. Now so in response to this, NINDS convened a workshop in 2012, and the workshop led to the publication of the perspective that I'm showing here on the slide, and this perspective, which was, again, written by many of the workshop attendees called for increased transparency and reporting. It was really asking for certain things to be reported for transparency, things like what were you sample ... How did you calculate your sample size? Estimations, were the animals randomized, lots of different things. So this is just really asking for people to report this information and be more transparent when they do so. Just going back to the ALS data, so a year after the clinical trial data were published, the group, ALS TDI, they published about 5 years' worth of their own rigorous studies exploring the effects of more than 70 drugs on the mouse model of ALS. So this is the SOD1 transgenic mouse, and as we can see here ... Here, let me just show you this graph. So here we're looking at the different types of drugs that were reported in the literature on the left, and then the x-axis, we're looking at the change in survival observed in the mouse study. So in blue we see that's the published data, so you can see a lot of these, these are all very positive, increase in survival. You see minocycline, over here on the left, led to the most increase in survival of the mice, over 35 percent change in survival, increase in survival. Now when ALS did their very rigorous studies using these same drugs, they could not reproduce the published data, and in many cases the drugs actually led to a decrease in survival of the mice. So this is very concerning. They, again, had a lot of mice in these cases they carefully controlled for many things, including a transgene number, and they couldn't reproduce the published data. It's likely that small sample sizes led to these erroneous results that were published. So that was a major concern. Now I want to next move to talk about, what are the challenges to rigor and transparency and reporting science? Now we often think of science as self-correcting. So we think eventually we'll get to the right answer if we're just patient, right? So the question is, does that mean that reproducibility doesn't matter? So maybe perhaps over the long-term, this is the case. But you know the short-term and medium-term, there are many factors that will short-circuit this self-correction and lead to a lot of terrible outcomes. Of course it leads to the reproducibility problem, and also you have to think about the potential loss of time, money, people's careers and public confidence in science. So these are things that we don't want to take for granted when it comes to thinking about reproducibility in science, and the possibility that if we just wait long enough, science will self-correct. So what are some of the factors that short-circuit self-correction? So one of the factors is poor training. So perhaps individuals were not trained adequately in experimental design, how to set up experiments. In some cases, perhaps the individuals are inappropriately using statistics. This is something called p-hacking, where you basically torture your data enough so that you finally get the result that you're interested in. A lot of times people are kind of shooting for the P value of .05, which is kind of really an arbitrary number, but that can lead to several bad outcomes, and again, p-hacking, which we don't want to happen. In some cases, we also see that there is incomplete reporting of the resources used and/or unexpected variability in resources. Now this means that, for example, there are cell lines that people are using, even still today, that are misidentified. So they think it maybe a certain type of cancer cell from a certain organ, but it's actually from a different organ. So there's also concerns about specificity of antibodies, for instance. So we do need to make sure that people are reporting their resources and that they're aware of the possible variability in those resources. Now other factors that can short-circuit self-correction include publication practices. So as you know, there is difficulty in publishing negative findings, and it would be great if more people could publish the negative findings because then other individuals would not waste time coming to the same negative conclusions. Not that it's a waste of time. That's often important information that does need to be out there. Also with publication, we often see there's an overemphasis on the exciting, big picture findings that leave out the necessary details. Now I remember when I was in grad school, around that time I think method sections started becoming smaller and smaller and smaller, and this is back when most of the time most of the journal articles were paper. Maybe they were trying to save space, but you know it really led to the de-emphasis of some methods, and methods are really important. So the good thing about the publishing world today is that a lot of this is online, and so now there is hopefully more room for people to expand on the details and their methods. Another factor that can short-circuit self-correction is the current hypercompetitive environment. So as you all know, we're at historically low funding rates. This could leave people to take shortcuts, for instance, to get the data that they want to because grant review and promotion decisions often depend too much on these high-profile publications, or maybe in some cases they need the data to get the grant. So the hypercompetitive environment is definitely another thing that contributes to this problem in reproducibility. So back in 2014, Francis Collins and Larry Tabak published an article, and this article really laid out the NIH's plans to enhance reproducibility through transparency and rigor. They mentioned the updates to grant applications that were going to happen starting in 2015/2016. The quote here is really an important thing to consider, where it says,"Efforts by the NIH alone will not be sufficient to effect real change in this unhealthy environment." So we have to really appreciate the fact that this really is going to involve the researchers themselves, institutions, journals, as well as the funding agencies to help enhance reproducibility and solve this problem. So also in 2014, the NIH published this article indicating that we're going to start to implement some policies to hopefully try to balance sex in animal and cell studies. So as you're aware, there have been a lot of problems around the lack of either females or lack of males in some particular studies. For instance, in some cases females were not always included in clinical trials, or perhaps when they were, the data were aggregated and so we really couldn't see the outcome and whether they were any differences between the males and the females. Also in some fields, people use male rates or female mice, and so there was a concern that this could lead to skewing of the data because we do understand that men and women, or male and female animals and humans are different, and they can respond to treatments differently. So we wanted to make sure that we were actually considering both males and females in the studies. So what this article lays out is that over the course of 2015, we were ruling out policies that would require applicants to address the inclusion of both sexes in biomedical research. Another article that was published this year in 2014 addressed the issues of cell lines. So you've probably heard of the cell line called the HeLa cell line. The HeLa cell line is very, very fast growing, very robust cell line, and it turns out that many people were working with other types ... They thought they were working with other types of cell lines, but the cell lines actually ended being HeLa cells because they were very easy to contaminate other cell lines. So there were some studies that were done to look at cell line contaminations, and since the 1960s, there's been more than 400 widely used cell lines worldwide that have shown to be misidentified. There was a 2011 study looking at 122 different head and neck cancer cell lines, and they found that 30 percent of them were misidentified. So this is a huge problem, and it continues to be a problem surprisingly. But we want to make sure that people are authenticating and making sure that they're using the resources that they think they're using. So this was a concern that we wanted to bring to everyone's attention as well. Again, 2014, there was an NIH-sponsored meeting to bring in several journal editors from journals sponsored mostly NIH research, and so they came together and they came up with these guidelines for reporting free clinical research to help improve reproducibility. Again, as I mentioned before, this is the problem of ... or lack of reproducibility is going to take the whole research enterprise to help out with, and journals were part of that as well. Now as we were rolling out these updates, the 2016 updates, these were the guiding principals for rigor and transparency. So first, we really wanted to clarify our longstanding expectations regarding rigor and transparency in applications. We think we were funding rigorous studies in the past, but we wanted to make sure that everyone is made aware that we still expect that you're proposing and conducting rigorous studies in a transparent way. We were also hoping to raise awareness and begin the culture shifts in the scientific community. So I think the more people that started talking about it, I think when awareness is raised, that does go a long way for people starting to improve the practices. We also wanted to make sure that applicants were describing their work in a way that reviewers could actually assess that proposed work for rigor and transparency. So that was important to us as well, and of course we wanted to make sure that we're always investing in the best science and minimizing unnecessary burdens for applicants. So in 2016, we rolled out these four areas for clarification. The first one used to be called scientific premise, and we made a couple of tweaks over that over the past couple of years. Now we refer to it as rigor of the prior research, and that was because there was some misunderstanding about what we meant by premise, and I'll talk more about this in a second. So the second area is scientific rigor of the proposed research. This is actually what you're proposing in your research plan, and what your experimental details are. The third area is relevant biological variables, such as sex. Again, we want to make sure that you're considering biological variables, including sex, when you're proposing your studies. Then finally, there's authentication of key biological and/or chemical resources. This is, for example, we want to make sure that you're using the cell line that you think you use, and we want to make sure that you have a plan to authenticate those cell lines and other biological chemical resources. So what grants does this apply to? So we have rigor research, which is those four areas that I just described to you, so these apply to our regular research grants. These are like RR1s, et cetera, career development awards, individual fellowships, center grants, people-based grants like R35s, mirror awards, program projects, small business grants, and resource-related grants, and I'm going to talk a little bit later about the most recent updates that we've made to training grants. So we also now have what I'm calling training and rigor. So this a requirement now for institutional training grants and institutional career development awards. It's required for fellowships, and it's required for career development awards. If you want to see a full list of the activity codes that are affected by rigor policy, either the rigor research portion or the rigor training portion, you can check out this list of activity codes that I have linked here at the bottom. So first, I'm going to focus a little bit on the rigor in research. This is showing you where each element appears in your application and how it's scored. So here we have ... On the left, we have the element of rigor, and we have rigor of the prior research. This is found in the research strategy. The criterion score is scored under significance, and it contributes to the overall impact score. Then secondly we have, again, rigor of the prior research. So in the first part we're asking you to describe any weaknesses in the rigor of the prior research, and then the second part of this, we're giving you the ability to describe how you're going to address the lack of rigor in the prior research, if there is any. So the second part of this is scored in the research strategy under approach. This does contribute to the overall impact score. The third one I have is ... This is actually the second element is scientific rigor. This is really your experimental design. It's in the research strategy, and it is scored under approach, and it contributes to the overall impact score. Then we have consideration of relevant biological variables, such as sex, which is found in the research strategies, scored under approach as well, and will also contribute to the overall impact score. Then finally, we have the authentication of key biological and/or chemical resources. Now this is an attachment requiring a plan. We don't want the data, and I'll talk more about that in a second. It's not scored. It doesn't contribute to the overall impact score, but it is an additional review consideration where the reviewers will say whether or not this plan is acceptable or unacceptable. We'll talk about more about that in a second. The first element as we mentioned earlier is the rigor of the prior research. We're asking applicants to assess the strength and weaknesses in the rigor of their prior research that serves as key support for the proposed research project. Now this prior research could include your observations, your own preliminary data or the published literature for any of these things. Now, again, this is under significance and scored under significance. Now when we get to the approach section, we then ask you to describe how the proposed research will address the weaknesses that you described previously, okay? So some of the FAQs, we get a couple of FAQs around this particular element. People often ask, when I'm assessing the rigor of the prior research, what should I consider? Now some of the things that you can consider, again, you can talk about your own preliminary data, or you can talk about some published data. You can talk about whether there was an appropriate sample size, whether or not the other studies randomized or blinded their studies, if applicable. You can talk about whether there was adequate positive or negative controls. You can also discuss the statistical tests that were used. Also biological variables, so perhaps one study was done only in males and you would like to assess this in females as well. That could be something that could seen initially as a weakness in the rigor, but you could then propose to address that. Authentication of key resources. For instance, maybe a previous study used a misidentified cell line. So many things you could describe here. Again, it doesn't have to be all negative. You can also describe the strengths of the prior ... The rigor of the prior research as well. The second element is scientific rigor. So we define scientific rigor as the strict application of the scientific method to ensure robust and unbiased experimental design, methodology, analysis, interpretation and reporting their results. We asked applicants to describe the experimental design in methods proposed, and how they're going to achieve robust and unbiased results. So one of the FAQs that we received is a question about how much detail should I include to address scientific rigor. So we ask that you succinctly state what is planned. So you're going to want to include information on how you calculated your sample size, how did you determine the effect size, or whether or not you're going to be blinding, whether there will be randomization in your studies. We also want to talk about what statistical analysis you'll be using. You want to make sure that you describe your experimental animal numbers here. This will be your power calculation because I think sometimes people include that in the vertebrate animal section, but we want to make sure that you include the description of that in your research plan as well. Of course, be transparent about your plans for analysis. We don't want a situation where you don't propose what your analysis plans are, and in the end, people start to analyze the data in a way that they can get results, which I guess is kind of like parking. Maybe it's not parking. But this is not a good practice, of course. You want to make sure you really lay out how you're going to do your analysis before you conduct your studies. So then you also want to stay within your page limits. We didn't increase page limits. Again, that's why we say you don't need to throw the whole kitchen sink in there, but definitely you want to include some of these things that I mentioned here, and just be as succinct as possible. Now the third item or element is consideration of relevant biological variables such as sex. Again, this doesn't only include sex, so we know that different things can also affect health or disease, including age, weight and underlying health conditions. So applicants are asked to explain how relevant biological variables, such as sex, are factored into research design and analysis for studies in vertebrate animals and humans. So it's important to pay attention to the actual ... what your requested to do. So you need to talk about your research design. You also need to talk about how you're going to do your analysis. It will be important that you specifically state that you will just aggregate your data to report that data is segregated by sex. So we want to see what does this look in males and what does this look like in females, for instance? So when you're reporting this data, you should report that in your progress reports to the NIH in your aggregated analysis. We don't expect that everyone needs to use both males and females. It is possible to propose to study one sex, but if you are proposing to study one sex, you need to include a strong justification in the scientific literature. You need to include preliminary data, or other relevant considerations. So make sure that this is very easy to understand for your reviewers. You want to just really lay out your reason for only studying one sex, okay? The FAQs that we most often get around this biological variable, such as sex, is this question about whether they will have to double their animal numbers. We want to make sure people understand that this is a continuum we're talking about, so we have very initial, early stage research, very basic science. Then you can then transition into becoming perhaps more pre-clinical, and then you may then translate it into the clinic, for instance. As you're moving towards the clinic, you definitely want to make sure you're looking at males and females, and you're doing that in a way that is ... You have a powered study. Early on, we understand that they're not may not be sex differences. We don't know what's going to happen, and so there's not really a need to do a robust study, I would say, to determine the sex differences in your animals early on because we're not asking everybody to now become sex-difference researchers. What we're asking you to do is to start to collect the data. So you may not need to power your initial studies to detect those sex differences, but we want you to start collecting that data, and maybe as you do several experiments, you may start to see a preliminary sex difference. So again, we're not asking that you become a sex-difference researcher. We're asking that you start to collect that data when you're early on in this research continuum, okay? As I mentioned before, if you're going to propose to study only one sex, you need to provide an explanation and justification. One justification could be that this is a sex-specific condition or phenomenon. For instance, you're studying ovarian or prostate cancer. Obviously, ovarian cancer, you would be studying females. Prostate cancer, you would be studying males. In some cases, there are acutely scarce resources. For instance, if you have a very complicated genetic class where you only get a few animals of a certain genotype, then you can include that as a justification for studying one sex. If you start collect that data, and eventually you start to see differences, perhaps then at that point you can start to make sex-specific hypotheses. So now you can say, "I've seen this particular effect in male mice, and I want to follow this up, and now I'm going to follow this up with the male mice." So now you can make a sex-specific hypothesis and continue with that line of research. That's another possible justification for including only one sex. Now cost and absence of known sex differences is an inadequate justification for not addressing sex, okay? The fourth element is authentication of key biological and/or chemical resources. Again, as I mentioned before, the quality of your established resources is critical. These are cell lines, antibodies, chemicals and mice as well. What we're asking applicants to do is provide a plan to regularly authenticate and ensure the identity and validity of those resources. Again, we don't need the data. We are asking for a plan. So one question we got often is, "What resources do I need to authenticate?" And so each investigator is going to have to determine which of their own resources are key. Now we had a lot of questions around this when we first rolled out this requirement, and we published a notice as NOT-OD-17-068 to really, again, reiterate what the definition of key biological and/or chemical resources is. These are things that could vary from lab to lab or over time. Again, some of the examples are cell lines, chemicals, genetically modified animals or cells. We again stated that you do not need to improve plans for authentication of data sets, databases, machinery or electronics. We don't include plans for the development in authentication of new key biological and/or chemical resources. So if you're planning to develop a resource, you need to describe that in the research plan and how you're going about that. This is specifically for existing resources, okay? Don't include plans for the authentication of standard laboratory agents that are not expected to vary, and do not include any authentication data or other data in that attachment, okay? That could get your application thrown out actually. Then I want to remind you that we do expect that investigators are continuing to tell us what they're doing, and what the approaches that they're taking to make sure that they're enhancing reproducibility through rigor and transparency. So there are questions in the RPPRs that specifically ask you to describe how your ... to emphasize the approaches taken to ensure robust and unbiased results. We also want you to talk about your future plans, including your efforts to ensure that the approach is scientifically rigorous and the results are robust and unbiased. So what about training? So this a new requirement that was rolled out throughout NIH for applications that come in with due dates on May 25th, 2020, and beyond. There's a notice that was published, NOT-OD-20-033. This is really explaining how institutional training and career development awards, individual career development awards and fellowships will require a description of training and rigor. So this is a description of training and the techniques for rigorous and reproducible research. What's interesting here is that here that is going to contribute to the overall impact score, okay? This is something to definitely keep in mind if you're applying for career development awards, or fellowships, or if you have institutional training grants of any type of career development, or a T-32, make sure that you're aware of this requirement. Then I want to go ahead and also just make sure that you're aware of some resources that we have. So NIGMS has a nice clearinghouse for training modules around data reproducibility. These were funded by R25s, and they were made available to the public. These were created by your peers, and these were some really good modules that include really helpful information that I think can go into any type of training plans around rigor. Some of the modules include information like Statistical Topics for Reproducible Animal Research, Controls in Animal Studies for Rigor and Reproducibility, the BD2K Guide for the Fundamentals of Data Science series, Cell Line Authentication Training, Society for Neuroscience Rigor and Reproducibility Training Webinars, and then we have A Guide for Scientists Working at the Bench. So there's a lot of good resources here, and you check it out. Here's our website for our enhancing reproducibility through rigor and transparency policy. Here you can click through the different resources. We do have a whole page for resources for preparing your application. We have an infographic that you can see here where it really points out what are the four elements that we're interested in and we're talking about, and where you need to report that information, okay? There's also a one pager that links to different blog posts and FAQs for each of these four elements, and that also tells you where in the application you need to describe this information. I think that has been very useful for people as well. Then we also made available this reviewer guidance document that recreated a report for sex as a biological variable, also called SABV. This is kind of a flow chart to kind of walk you through whether or not you're proposing to study vertebrate animals or humans, and then if you are, are you going to be including both sexes, and if not, what to do. So this is a very useful flow chart. I recommend that you definitely go take a look at it, and it'll help walk you through kind of like what the guidance is that NIH is giving to reviewers. I think it's helpful for applicants to know this. It can also help you walk you through what kind of things you need to include in your application. The last thing I want to mention is this upcoming ... So there is an ACD, or advisory council to the director, working group that was formed, I think it was about last year. This group is called ... It's a working group for enhancing reproducibility and rigor in animal research. This working group is charged with identifying gaps in opportunities to improve rigor and transparency in animal model studies. They're going to be assessing the current state of science for validating alternative models to animal research. They're going to look into, what is the feasibility of asking people to register their animal studies that to aim to lead to a first in human trials. They're also going to be looking at the financial implications, and they're also going to be assessing training and rigor of animal research. You can learn more about it at this link here. So this is something that we'll be getting more information from this ACD working group. It should be very interesting to see what their reviews and assessments bring to the table. There will be definitely be a lot more discussion and provide a lot of thought I think for this issue. Finally, I just wanted to thank you, and if you have any questions, I think we're going to take those questions later, but in the meantime you can e-mail at reproducibility@nih.gov, and we will get back to you then, okay? Thank you so much.