This Week in Virology

with Vincent Racaniello, Ph.D.

Special Episode: How to Read a Science Paper

With hosts Vincent Racaniello, Dickson Despommier, Rich Condit, Alan Dove, and Michael Walsh

Originally aired 5 February 2012

Released as separate podcast 27 November 2012

http://traffic.libsyn.com/twiv/TWiVspecial2.mp3

[0:00:00] (Or beginning in TWiV 169 at 1:43:56)

Vincent: Hi everybody, I am Vincent Racaniello and you are listening to a TWiV Special. Not too long ago on TWiV 169 in response to a listener email we had a great discussion about how to read a scientific paper. Each of us gave our own opinions and practices on how we do it. That included myself, Rich Condit, Alan Dove, Dickson Despommier, and Michael Walsh, our guest on that episode.

A number of individuals have told us that they liked that part of the show so much that I've decided to split it out and make it a special on its own. If you have listened to TWiV 169, you don't need to listen to this one unless you would like to hear again how we read a scientific paper. Enjoy!

[0:01:05]

Dear twiv/twip/twim hosts (not really sure where this mail belongs).

I recently started reading a lot more papers than what I'm used to (school related) and I actually find it quite difficult to use them.

Of course it doesn't help that English is my second language, but I can't help thinking that it's more than that, so:

How do you make a paper accessible to yourself? I mean, everybody can read a paper, but actually extracting the relevant information seems more like an art than anything else.

How do you avoid getting lost in the details or missing them completely? When I read the methods for example, everything kind of runs together, especially when they repeat the same experiments just with different doses or a slightly different composition of drugs (vet school student).

In short, I guess I'm asking: How do you decode a paper, to get out the relevant information without getting lost?

Thanks again for the great podcasts :)

So we can all weigh in on this.

Alan: I think we should let Michael go first.

Michael: I do a seminar occasionally on critiquing the literature. I think I understand what she is saying about it being an art, but I also think it helps to have a systematic approach to going through. This is what I try to get across to students. In fact I can send you my… I have power point slides that I can send you if you want them.

Vincent: That would be great.

Michael: The first thing that I make sure to convey when you are starting to read through a paper is you are obviously starting with the background and the introduction. You need to thinking about how well does the introduction that you are reading synthesize and summarize the current state of the knowledge with respect to the research question that this particular group is going to ask. Does in fact the description of the background identify an area that is lacking? Does the background naturally flow into the hypothesis?

Do they state the hypothesis? More than that, do they clearly state a measurable-testable hypothesis? If you are stating the hypothesis at the end of the introduction and you say, we sought to evaluate the relationship between job-related stress and cardiovascular disease. That is not a measurable-testable hypothesis because you haven't even defined your exposure or outcome with that. I think that is a critical thing to look for, is a clear statement of a measurable and testable hypothesis.

When you get into the methods it is going to vary of course by field. The things that I am going to look for are going to be different to the things that a molecular biologist is going to look for. But ultimately what you want is to be described is the study design. Why did the investigators choose the particular methods that they used and how does this relate back to again to the hypothesis under study?

Are they using methods that are appropriate to answer the question? In epidemiology one of the things that we need to look for are what are the measures of occurrence? We had this discussion on case fatality, how are the investigators defining what it is that they are measuring? What are they using to measure the association between the exposure and the outcome?

One thing that I always look for in results are data clearly represented in tables and the figures. I don't want to read an endless list of their findings. I want all the salient… I want the key data to be placed well in tables and figures and then you in the text, you highlight the salient features. I don't want to read lists and lists and lists in the text of what you identified.

Is there internal coherence with the results? In other words, do things add up? You have to look at that. You would be amazed at how often they don't. People will report things in one table that directly contradict what they are reporting in another table. I mean it is amazing to me how these key features, like a research question that is clearly stated, is not often well articulated in a research paper. Or their description of the methods is completely vague.

Finally when it comes to the discussion, first of all do the authors succinctly summarize their findings? Let's start with that. They should be interpreting in each of the major findings that they report in the results… they should be interpreting a point-by-point in the discussion and then grounding those interpretations in the existing literature. I don't want to read just your philosophical interpretation of what may be happening. You can put forth an interpretation but it shouldn't just come out of thin air. I want to see that grounded in literature.

How honestly do the investigators discuss their strengths and limitations, often they don't. They sometimes won't even state them explicitly.

The final summary is did the authors accomplish what they set out to with the study. Do you feel as a reader, as an informed critic of the scientific literature that this is an appropriate addition to the scientific literature?

So that is just sort of a summary of the steps that I think can be useful to go through. I will send you the slides, there are more specific points that any student can go through and apply.

Vincent: That is great, that would be great. Rich, what do you think?

[0:07:28]

Rich: Well, I am going to… I like that as background and I am going to assume that I am a student or anybody approaching a paper that is written in a satisfactory fashion relative to all those points. I've still got a problem because I have to read this thing and the first few times around, especially, it is really difficult.

I would say that in particular as a beginning student, almost no matter what you do it is going to be difficult because it just… you need background and more than anything else what you need is practice. I think that it's real easy to get frustrated early on and say I just don't get it and then don't do a thorough job and kind of leave it.

If you find that it is difficult don't worry, you are not alone. I think for everybody starting off this is in fact a difficult task and the way to get it is like riding a bicycle or anything else, it is just to keep at it. As you acquire more experience with it and as you acquire more background knowledge it will get easier.

I find that when I read a paper I take advantage of the organization. In our field the organization is typically abstract, introduction, materials and methods in different journals goes in different places, then its results, and discussion.

Although as a traditionalist, I objected initially when people started putting materials and methods at the end of the paper, I find that less and less objectionable because as time has gone on… another thing I should say is that I think over time with practice everybody develops their own method at this.

What I have found is that I pretty much go through the paper from beginning to end. In fact the organization that we ordinarily use makes pretty much sense. Some sort of background information as to why they did the experiment and then jumping right into the results. If the results are written properly they have a summary of they did the experiment; so the methods are implied. I find with the methods at the end, as I read through the paper I'll find that I don't have enough from the results to understand how they did the experiment and then I'll use the methods as a reference section… to go to the methods and find out what they actually did.

I can understand that reading straight through the methods from start to finish there is going to be a lot of extraneous stuff in there that you don't necessarily… detail that you don't necessarily need to understand the experiment. If methods can be used more as a reference section at least initially that is okay.

The key to me is that it doesn't matter what they say, what matter are the actual data that are in the figures and in the tables. To me the focus is on that. You look at a figure and your job is to understand exactly what the methods were in order to generate the data in that figure. Then your job is to understand exactly what the data are. You have these black blobs, what are they? Is that radioactivity, is it a Western blot, what is it? If there are numbers, where do the numbers come from? Understand the data and then understand the interpretation and then understand the extrapolation from that.

When I read a paper, what I do is I go through and I kind of cruise through the introduction and get an idea. Then I cruise through the results and I get up to here is the question and here is the experiment that we did, and I get up to the point where they are going to describe the results and then I put the text aside.

I get out the figure and look at the figure and say, do I understand how this was done? If I don't then maybe I have to go to the methods or something like that and figure it out, as if I were doing the experiment. I am at the bench, I am doing the experiment and so I understand it from start to finish. I understand how they got those data and then I look at those data and I draw my own conclusion.

Then I look back at the text and see what they concluded and see if they are the same. Then I press on. I do that when I am reviewing a paper in particular. This is what I teach the students. If I am reviewing a paper I do the same process—read the methods up until the point where they are going to make a conclusion, then study the data in the figure and draw my own conclusion, and compare it to what they said. If it is the same the paper gets accepted, if different it gets rejected.

That to me is what actually makes it fun because I want to participate in the process. That way I get to do the experiment and I'm really engaged and I'm asking myself my own questions about how it was done and all that kind of stuff, not just reading what they said. Then you can compare your process with their process and then move on to the next thing.

So that is what is key to me. The data is what drives, the data are what drive it and so you have to understand how the experiment was done and provide your own interpretation for the data.

[0:13:12]

Vincent: Alan.

Alan: Okay, excellent, excellent descriptions, both of you. What I can add to this is the perspective of a mediocre student. When I got to graduate school I felt like a total idiot because I had read papers as an undergraduate but then all of a sudden I had to read bunches and bunches of papers. A lot of the time I would sit down and I'd look at these things and I wouldn't get it.

The first thing to realize is it is not because you are stupid, in my case that may have been going on but in most cases, it is not hard because you are stupid, it is hard because it is hard.

A scientific paper is not a magazine article, it is not like any other kind of reading you've ever… it's not even like a text book. It is a compressing an enormous amount of information into a very small space to transmit it over distance. The first thing you have to understand is that you are not there to read the paper so much as unpack it into your mind.

What I do is sit down with the paper… I have evolved this process to the point where I read the abstract just to kind of skim quickly through it. You are not really going to get everything that is going on in the abstract because that is taking the compressed information and compressing it even more. Then dig into the introduction. In a well written or reasonably well-written paper the introduction will start off with stuff that you probably already either know or can look up pretty easily and figure out. It will funnel down toward this is why we are doing this particular research. That sets the ground work for the whole rest of the paper, hopefully.

Understand that there are an awful lot of scientific papers that are very poorly written. That is not because the science is necessarily bad, you can see these in great journals, it is just that they had to compress too much information or they were in a hurry to write it because they were going to get scooped and for whatever reason the writing just may not be all that great.

Wade through the introduction before you get on to the rest of the paper and figure out what the context of all of this is.

Then I kind of follow the pattern that Rich described although I am actually, again it's the bad student syndrome, I am likely to skip to the discussion and give it a quick read through to see where they are going with all this.

The introduction and the discussion are the bookends of the whole thing. The introduction sets the scene and concludes with the hypothesis—or as we use to call it, the my research, the rescue line. You know, this is a problem for this reason and now our research to the rescue. We are going to explorer this hypothesis. Then in the discussion you should see where they are really headed with this. There is going to be discussion of a bullet type summary of the major conclusions and then here is what we further hypothesis from it.

From there I will frequently go to the figures and tables before even getting into the results section. I find the results section is not terribly useful to me. The figures and tables are the data. As Rich said, that's the real science in the thing. The thing to understand about figures and tables in a scientific paper is they are not like figures and tables in a power point presentation, even at a meeting. Sometimes they are but usually power point people will separate out the experiments a little bit more.

Again, in the paper it is all compressed together. So you have one figure that has six panels and each of those panels represents a few different experiments and controls that are all fitting together into this thing. Don't expect to glance at it and figure it out in sixty seconds. When you look at a figure you set aside a little time. It can take 15, 20, 30 minutes to figure out a figure in a paper if it's a big one.

As Rich said, this is what is the real heart of the matter and I will sit there and wade through—okay what are they doing here; where did these samples come from; what did they do to them? That's when I'll refer to the result and sometimes even to the methods. Although I have to say, I seldom find a whole lot of reason to refer to the methods if I am in a hurry.

Figure out what the experiment was and then look at the results and form your own opinion about what those results actually show because you are now… when you go back and you actually read the results section you are going to be told there interpretation of these results which, as Rich pointed out, you may not want to agree.

That is my general entry into a paper—the introduction to figure out what the heck is supposed to be going on, skim the discussion to see where they are headed with this, and then dig into the figures and grind through those. Understand that the whole process you are going through is to unpack this information that has been presented in front of you in this very dense format.

Vincent: Thank you. D-Cubed… You don't read papers do you?

[0:19:00]

Dickson: Being the oldest member of this group I can go back before you had… when we have rapidographs (?), well we pounded our own papyrus.

Rich: Before we had paper?

Dickson: I can tell you something that has happened over the last thirty years that a lot of you are also aware of because you have that literature to go back and look at and that is the quantitative expression of results now has gotten exceptionally sophisticated. There are programs now available on your computer for massaging data into visual imagery. So it depends on the kind of paper as to how easy it is to understand.

If you look at the ecological literature for instance, where you have these global studies of climate change or sea ocean temperature changes or glaciation, event that occurred over millions of years, or ice core sample data, points in oxygen levels indicating whether the earth is getting warmer or cooler, that data, those data have been expressed in a way that allows your optic lobes maximum opportunity to catch it right away. Those are easy to see things.

I think even looking at DNA chip data in which you have little mountains that come up expressing certain levels of expression for certain genes during certain times of a cell cycle for instance, etcetera, if you just go, like Rich said, to the results and you can see these things expressed in visual imagery that your brain can interpret it makes the paper very easy to understand. Then you can go back and fill in all the results and all the data that are missing in your mind as how that was done.

I start with the abstract, no question about this. I want to know if these people think they have made a contribution that worth publishing that is worth me taking the time to understand. So I will see what the big issues were and what their conclusion was from their data that they got.

Then, just like Alan said, the introduction is key because you define the problem that exists; you talk about the missing parts, and then talk about how you've filled in the missing parts with some of your own experiments. Then of course Rich had… you said it best, just go to the results and pretend they're yours and envision how you would have done that. I have the most fun doing that too by the way Rich.

I review different literature than you do I'm sure because I am looking at a lot of ecological data to be honest, right now at least. To look at vegetation during certain seasons or during certain disasters, etcetera, etcetera, those are easier to express in some ways. I think the ecologists have had an easier time of it because they started out with graphs that express quantitation as well as qualitation. You could watch the parameters change over time in a single graph that was absolutely beautifully expressed. I'm talking about energy flow diagrams and things of this sort.

When you have experiment after experiment after experiment, and I varied this and I varied this by quantitation, and then I went on to the next one and I varied that, and you have these thick six or seven or ten paneled expressions of the same thing over time and your eye has to make a movie out of it almost and start at the top and work at the bottom and see what the event was that you are even trying to describe, it is harder to place yourself into those experiments when you are just starting out.

This is a person who is just starting out. They haven't done the work yet, so how would they be able to do a Rich Condit approach? They would do an Alan approach first—I am so stupid I don't even know where to look first. But as time goes on and your own data start to come out and you are starting to express them yourselves to get your first publication out of the way, for instance, then you start to see, oh there is some logic to all of this.

I will begin with the simple experiments that are easy to explain and work my way down into the parts that I don't have a clue as to what I just did, but it did show something and maybe I will reserve that for another paper. That's how you start sorting this data out.

Remember this, that research is an ongoing process. That is what we just devoted this whole discussion to today. It is not a story that is self-contained, although you try to write it as if it is. So you write I did A, and then finding A, I got B and finding B, I got C and finding C, I got D…. That is not how research goes. I found A, therefore I went to F. Now if F didn't work I am going to go back to E, and if E didn't work, I am going to try to save some steps on this thing.

Michael: Then you go WTF.

Dickson: Yes, you are exactly right. Or even worse, I could spell out a lot of other things here guys that I'm not going to do.

The point is we know to go from A to B we go from A to Z if we could, and then Z starts another A. That's how you do research but when you write it up, you write it up as though you did A, B, C, D, and E. It doesn't make any sense when you read it; it makes sense when you talk to a scientist and how they conduct laboratory research. The more experience Sophie, the more research you do, the more you will see the logic system that human brains apply to the expression of stories. You are going to tell us a story. The story will be self-contained, it won't be complete, but it will be self-contained. It is enough of the larger picture to write a small short story of a novel that could be your whole life. Your novel is what you do at the end of your career, not at the beginning. So you are writing chapters to this novel if you are lucky enough to work on the same thing, unless they tell you not to work on H5N1. You can write this novel for yourself by the time you get to be thirty years down the road in the lab.

So that is how I look at these things and you review grants the same way of course, only on a different scale. But you are still looking for the story. What is that story? How did they tell it? Did they tell it in a way that made it clear because if they don't either the paper doesn't get published or you don't get your grant. So it forces you to express yourself clearly, eventually.

Vincent: Alright, thank you Dickson.

Dickson: You're welcome.

Vincent: I don't read papers so I don't have anything to add.

Dickson: He doesn't read them, he writes them.

Vincent: I remember when I was a graduate student, there was a really smart professor in the department, he used to go get his mail at the department office, get his Nature and by the time he reached his office he would have read an article and understood it and he'd come in the lab and tell me all about it. This freaked me out, how does this guy in two minutes understand this. The key is what Rich said and everybody else, you'll get used to it. You'll develop your own way of doing it. Now, I can do that, although we don't pick up journals in the office anymore. You can do it.

What I do is that I read the abstract because I think it summarizes the whole paper. It tells me what to expect. The intro is really important because it—as Michael said—it frames what they are going to do. I skip the methods because these days I pretty much know how all this works. So Sophie you may want to go do it, but I do it at the end of the paper because I think it's just going to interfere at the beginning.

Then I tend to focus on the figures too. I don't really read that much unless I need to. I like to see if I can figure out what their conclusions are ahead of time. Then, depending on the journal, so the Journal of Virology which I read a lot, in the results section they have subheadings. I am looking at one here, Mirror 122 Enhances Production of Virus through Interactions with RNA. So it summarizes the findings in that paragraph. Then at the end of each of these sections, if they are written well, [it says] these results indicate that, and it summarizes all the results. So I kind of go to the heading and the last sentence. To summarize I go to the figures and I go through the results very quickly. That is what I do for TWiV often.

Then the discussion I will read if I find that there are some unanswered questions, like what did this result mean, where is it going, and so forth. I don't always read the whole thing because I have the figures and I know what it means to me. Those are just some things that I do.

Other journals, like Science, they don't make subheadings so you can't do that. I think you have a lot of suggestions.

Rich: I have one more comment on this and that is that one of the things that I do here is I talked to the students on how to give a talk. I've talked about how I take apart a figure in terms of the methods, and the data, and the interpretation, and then the conclusions from that.

For me, giving a talk should be the same process. The talks that I like the best, the talks that I like to give, are where the focus is really on the data. When somebody shows a slide with data on it, they take the time to tell me what the experiment was, and how it was done, and what the data are in the conclusion, and point out the different parts of the figure, describe all the elements so that I have time to assimilate it. So once again, I can participate in the process. That does not seem to be the trend. That is how I was taught. Or at least those were my role models.

Now there is a great tendency to take Alan's eight panel figure and throw it up and not even describe the experiment but just say what the conclusion is and move on. I am not getting anything out of that.

Vincent: Yeah, that's right.

Alan: I think power point has had a major effect on that.

Rich: Yeah.

Vincent: TWiV is on iTunes and at the Zune Marketplace and at Microworld.org. Check out our app at Microworld.org. Send your questions and comments to TWiV at TWiV.tv.

I am Vincent Racaniello and I can be found at Virology.ws.

You have been listening to This Week in Virology. Thanks for joining us. We will be back next week.

Another TWiV is viral.

Transcribed by Steve Stokowski