

CDC/MCH EPI 2008 Conference

Scientific Writing Communicating Research Investigation Effectively to Expedite Publication, Programming and Policy in Public Health Training II a

December 8-11, 2008

PAUL Z. SIEGEL: ...variable like age and then you break it down into age categories. That's an example of organizing the data starting with a numeric variable and then changing it into a categorical variable or a continuous variable to a categorical variable. And then there's this extra point here about data that are relevant to the purpose of this study. And so for a moment I want to put the paper into the context of the analysis. I mentioned this morning about the Missouri BRFS paper is an example of a paper where the study had two distinct components to it and I think it is completely appropriate to write two papers based on one project. Now that's not unusual. You might have one project. Maybe there are two papers. Well the two papers won't just be different introductions and discussions. The data will be different. So if you imagine that you've got a mountain of data in front of you it doesn't just all go into the paper. You, as the author part of your job is again to decide what's the gap in the literature that you're trying to fill, what's your approach to filling that gap and when you get to the methods section, it's very important to circumscribe the content of the methods section so that it corresponds to the purpose of the study. So you don't just present all of your data in the methods section. You present the data that are relevant to the purpose of the study.

Now going down page 40 just a few examples, clearly present and define all of the analysis variables. So we saw in that last paper that had the very long methods section, this is just a good example of presenting the analysis variables where there's a distinction, the author draws a distinction between the analysis variables and co-variants. I think that's a useful way of organizing. Now I think there's too much text in there. It doesn't need to be so long but it's a good way of organizing the presentation of the variables. So now I'll ask you to turn back to the Missouri BRFS paper...and think again about this point here...clearly present and define all of the analysis variables. So I'd like you to look at this portion and if somebody could call out the page that it's on...page...that table is on page 13...is that right? Oh, I'm sorry analysis...right, right, yeah....okay it's on the left-hand column paragraph from the bottom. So I'd like everybody to ready that paragraph under analysis and ask yourself the question do the authors clearly present and define the analysis variables in that paragraph? Okay what do you think anybody?

UNKNOWN SPEAKER: (Inaudible)...

PAUL Z. SIEGLER: Right.

UNKNOWN SPEAKER: ...you look up at a table and you see questions that lead you to the variables...category questions...and the (Inaudible)...

PAUL Z. SIEGLER: Well let me ask you the question what are the variables? For this study, what are the variables?

UNKNOWN SPEAKER: The variables were...actually the questions (Inaudible)... looking at the liability among the questions.

PAUL Z. SIEGLER: The variables are what? They are looking at what? You said it...they're looking at...?

UNKNOWN SPEAKER: (Inaudible)...is that what you were asking?

PAUL Z. SIEGLER: No, I'm asking what is the, or are the variables in this study?

UNKNOWN SPEAKER: The questions on the behavioral risk factor (Inaudible)...

PAUL Z. SIEGLER: Actually not.

UNKNOWN SPEAKER: Could be hm...

PAUL Z. SIEGLER: I'm not torturing you. What I'm getting at is the authors don't tell you...this section is labeled analysis....so what are the analysis variables in the study?

UNKNOWN SPEAKER: (Inaudible)...

PAUL Z. SIEGLER: Okay it's not clear but what are they?

UNKNOWN SPEAKER: (Inaudible)...

PAUL Z. SIEGLER: Broad categories of questions and what are the analysis variables?

UNKNOWN SPEAKER: (Inaudible)...

PAUL Z. SIEGLER: Yeah, the reliability. That's the analysis variable in this study is reliability made the agreement, we're not even exactly sure but those are the variables. Its reliability and/or agreement so we have a section that's labeled analysis and it does describe analysis but at least we don't really know what the variables are. Now we know what they are because it's obvious but the authors are not telling us what the variables are. So as opposed to the previous example where here we have the variables. Okay but here the variable that...the variable with interest is the reliability, not the questions, it's the reliability of the questions so we're of course concerned about the questions but the analysis variable is the reliability. So I would say the analysis variables are not clearly defined. We already talked about in the very beginning about whether or not the article is about reliability or about agreement and so if we scan here we see...okay kappa values greater than .75 represent excellent agreement, less than .4 show poor agreement. In the middle show fair to good agreement so it looks like our variable is

agreement. Calculated...and...so it looks here like agreement is...agreement is at least the variable that the authors are talking about here.

UNKNOWN SPEAKER: (Inaudible)...

PAUL Z. SIEGLER: Uh huh...I'm not sure. A variable is something that has a value. We don't know what the value is. It could be high, it could be low in the middle and in this case that's reliability. We're measuring...we're measuring the variable. The variable in this case is reliability....or maybe...

UNKNOWN SPEAKER: (Inaudible)...

PAUL Z. SIEGLER: Yeah those are measurements. The correlation is an example of the measure but in this case I think the variable is reliability.

UNKNOWN SPEAKER: (Inaudible)...?

PAUL Z. SIEGLER: Yeah, I think you're right. I think there is some ambiguity between measurement and variable.

UNKNOWN SPEAKER: (Inaudible)...

PAUL Z. SIEGLER: Right, that's exactly right and in this case I think its reliability.

You've got all these questions and then the variable for the variable is reliability. I mean for example...

UNKNOWN SPEAKER: (Inaudible)... reliability of...?

PAUL Z. SIEGLER: Yeah...the reliability of each question is the variable. But we could maybe discuss it a little bit more. There may be a nuance in there but I think that reliability is the variable. So uh...and then of course there is no definition given in here. There are analytic procedures and different ways of measuring these variables but there is no...there is no definition at all. But there are references. There is reference 16-19, reference 17 and reference 20 and 21 so as I said earlier any time there is a gap in the paper or anytime you want to avoid putting large amounts of text in the paper one good strategy for that is to provide references so that the reader can go to those references and find information that's not...that you don't make explicit in the paper so there are these five or six references here. So let's take a look at the references. They are numbered 16-20. So if we read these references what can we expect to learn about the analysis variables? Can we expect to learn how to measure and interpret agreement? Well for the Cotter's variables we use Kohn's kappa statistic as a reference kappa values greater than .75 represent X on agreement. So I think it looks clear. If we go to these references I think the text is telling us we can go to these places and learn about agreement. How about this...can we learn how to measure or interpret reliability? Can we expect to learn how to measure or interpret reliability if we go to these references?

UNKNOWN SPEAKER: (Inaudible)... it doesn't spell it out...I mean it's possible.

PAUL Z. SIEGLER: It's possible...

UNKNOWN SPEAKER: He doesn't spell it out but (Inaudible)...

PAUL Z. SIEGLER: Right, it's possible but the author is not giving us a clear indication that if what we want is to learn how to measure reliability it's not clear that we can go to any of these references and learn anything about reliability. Can we learn whether reliability and agreement mean the same thing? Can we expect to learn that if we go to these references? Not really...not at all because there's just nothing about...nothing about reliability. There's no direct reference to reliability. So we don't know if we can learn anything about how to measure reliability. We're told we can learn about agreement. Can we learn how Pearson and Spearman correlation coefficients are used to measure agreement? Yeah, I think we can because it gives an indication we can learn that. So this is just to encourage you that when you include references in your paper, try to make it clear to the reader what the reader can learn when they go to that reference. Try to be fairly explicit about that. So again, methods, purpose to describe how you collect, organize and analyze the data that are relevant to the purpose of the study and now we're going back to I think it was page 40? Yeah, page 40 and going back down this list. Organize into logical sub-sections that illustrate the steps that you took to collect, organize and analyze the data. It's very important and very helpful to

organize the method section with subsections. It makes it possible for the reader to go, oh okay here I want to learn about the analysis or they can go to the section in the methods that's labeled analysis. I want to learn about how the study subjects were selected. They can go to the section in the methods that's labeled study subjects. So some examples of subject headings, study population, sample methods, laboratory methods, epidemiologic investigation...this is also on page 40. Baseline data collection...diagnostic evaluation...these are all just examples of possible sub-sections to the methods section. And very often statistical analysis is the last one.

I want to show you a couple of examples. This one begins on page 41...for this I'm not going to ask you to read the paper I just want to highlight how the method section is organized. In this case you have these sub-sections in the methods. So there's an example of one paper that has five sub-sections in the methods. No, six sections because there's a section on ethical considerations. So now if you go ahead to the article on page 45...you have another example and this one has more detail. There are more sub-sections like maybe eight or nine of them. So there's an example of a paper that has one...one, two, three, four, five, six, seven, eight...sub-sections to the methods section. One more beginning on page 48, methods participants, intervention, process and outcome measures statistical analysis. But then notice actually one of the sections...one of the sub-sections is divided into sub-sub-sections so you can create this kind of hierarchy when you organize the methods section. And then here this is just a practical example of another paper that has a flow diagram with those four elements. The enrollment, allocation, follow-up, and analysis. So now we're going back again to

page 40 going through our list on page 40...in the methods section describe what you did and not what you found. So we looked at an example earlier that had a lot of results that was put into the methods section so be careful not to do that. And as a general rule describe things in chronological order. Now that won't always work, there will always be some exceptions...some reasons why it makes more sense to describe things in an order that is not strictly chronological but start out describing things in chronological order. And then if for some reason it doesn't make sense, well then you can shift it around but at least make sure that if it's not in chronological order that you're doing that for a good reason. And then as we mentioned describe original methods in detail otherwise give references. And methods section will vary a lot. The length of the method section will vary a lot if you've done mostly original work. If the methods that you used are mostly original well you'll need to describe them in detail. If you're borrowing methods from previous studies in particular if you're doing secondary analysis of the data set that's already been collected and the analysis that you're doing is fairly standard the method section might be two short paragraphs. Okay now let's go ahead to page 51. I'd like you to do a little exercise here. So this is an excerpt from the methods section of the paper, so it's brief and I'd like you to look at that. And then think about the question at the bottom of the page. Does all of the information above belong in the methods section and if not, where would it fit better. So just take...just take two minutes and see what you think about that. Go for it, there's no right or wrong answers just...

UNKNOWN SPEAKER: I have the same issue (inaudible)...

PAUL Z. SIEGLER: Speak up please...

UNKNOWN SPEAKER: (Inaudible)... okay...I had the same issues when I did my dissertation a couple of years ago. I had one committee and everything but I had the same issue where I had (Inaudible)... and one committee member said methods the other one said (Inaudible)... it was never really....

PAUL Z. SIEGLER: They couldn't agree on it...

UNKNOWN SPEAKER: They couldn't agree.

PAUL Z. SIEGLER: Okay, do you have an opinion about this?

UNKNOWN SPEAKER: (Laugh) Uh, so I'm still confused.

PAUL Z. SIEGLER: You're still confused. Okay does anybody have an opinion?

UNKNOWN SPEAKER: (Inaudible)...

PAUL Z. SIEGLER: It's results yeah. So what do you think? Is it in the wrong place? Okay let me ask you this. I mentioned this morning. Your goal is to...make the reader feel like they're on a train and the train is moving forward.

UNKNOWN SPEAKER: Well I think it's relative to the method that you used so I would leave it in methods.

PAUL Z. SIEGLER: You would leave it in methods okay...oh wait if you have something or you come up with something and you're not sure where to put it always try to put yourself in the shoes of the reader and ask yourself, am I writing this in a way so that the reader will read this and then move forward and not read it and think did I miss something? Do I need to jump ahead? You don't want the reader to feel lost. So this is an example where the authors started out on a pathway...they did some preliminary analysis and then they made an analysis decision midway in the process. They did not make a decision in the beginning to do stratified analysis by sex. They made a decision to do that after they observed preliminary results. So I think to make the readers feel like they're on the train moving forward the author has to give a little explanation for why they're doing an analysis that's stratified by sex. So I think this is an example where putting a little bit of results into the methods section accomplishes that goal of helping the reader to feel that they're on the track and they're moving forward. If you simply say we did all analysis performed separately by sex. What do you think the reaction of many readers will be? They'll ask why and they won't know why unless you tell them. So this is an example I think where it's a good idea to put a little bit of results in the methods section. General rule of thumb is if you think it's essential to put results in the methods section try to present the minimum amount of results that the reader needs to stay on the track. So in this case I think it could be a little briefer. It could be something like this

simply because sex specific death rates were unequal, male equals, female equals all analysis were performed separately. So what I'm saying is I think it's the difference in the rate. I don't think it's really essential...oops sorry. I don't think it's...well I think it's less important to give the death count. More important to give the rate, less important to give the count is what I think.

Okay let's go ahead now and move along to next...we're on page 52 now. Influence of social factors on avoidable mortality a hospital based case control study. So it begins...124 patients died of some type of avoidable mortality cause. Now hoping that some of you will immediately see that that is asking for editing. Does anybody see some editing that needs to be done there?

UNKNOWN SPEAKER: Maybe just define what some type of avoidable cause.

PAUL Z. SIEGLER: Well do they need to define it? Now this is out of context so try and think of it as something...I'm not asking you to figure out what's the context...I'd just like to edit this so that it flows and that there's no unnecessary words in there.

UNKNOWN SPEAKER: (Inaudible)...

PAUL Z. SIEGLER: Right, why sub-type up, why not just 124 patients died of an avoidable mortality cause.

UNKNOWN SPEAKER: (Inaudible)...

PAUL Z. SIEGLER: What's that?

UNKNOWN SPEAKER: Is avoidable mortality cause standard lingo? It's in the title, I guess it is.

PAUL Z. SIEGLER: Avoidable...yeah actually that is. Yeah it's a little bit of a lingo thing, yeah. Okay so we could say that 124 patients died of an avoidable mortality cause, but what's the redundancy there?

UNKNOWN SPEAKER: Died of...

PAUL Z. SIEGLER: Okay...we all when we die it's from a mortality cause. So it could simply be 124 patients died of an avoidable cause so you've saved about five words right there. Now of the 124 eligible cases 42 were excluded from the analysis. Here is our friend Winnie the Pooh, he is back. He has a question and what do you think his question is?

UNKNOWN SPEAKER: ...same thing and it was very confusing writing so it was really hard for us to figure out how to write it. The editors kept sending it back saying it doesn't make sense, it doesn't make sense because the kind of total that was excluded. And when we wrote it down as percentages to show that of the whole 50 5% were this

reason, 10% were that reason and we just couldn't get the wording right and it's as confusing here as it was when we started. You're adding the 22 up so it's really...it's the 42 (Inaudible)... by 11 and then 24 and then 7, but the way it's written it makes it look...

PAUL Z. SIEGLER: Okay well you're jumping ahead to the next step. So please just...we'll get to that point but for right now...we talked this morning about the reliability in the agreement that who was confused. What do you think he is confused by now?

UNKNOWN SPEAKER: Eligible men excluded.

PAUL Z. SIEGLER: Yes, so we have avoidable cause...we have eligible case here and died of an avoidable cause...so are these the same? Is an eligible case the same as dying of an avoidable cause? Now the numbers are the same 124, 124...so it would be logical to think...it would be an indication that they're probably used as synonyms but you don't really know for sure. So we could fix that in this way...124 patients died of an avoidable cause, 42 were excluded from the analysis. Again avoiding terminology that looks like it might be two terms that mean the same thing but the reader...the author knows if they mean the same thing but the reader may not know if they mean the same thing. So try to avoid that. Try to use consistent terminology.

Okay so that's just the beginning. Who is happy now? 124 patients died of an avoidable cause, 42 are excluded from the analysis. That means that 82 were included in the analysis he's comfortable. And now comes this...and poor little Pooh is really confused

now. And can you figure out why he's really confused? Are you confused? Or do you see anything in there that's confusing? What's confusing?

UNKNOWN SPEAKER: That's what I was saying. The way they have it spelled out is very confusing. You feel that it's 42 plus 11 plus 24, but it's not. It's the 42 broken down by those...

PAUL Z. SIEGLER: Oh okay...you are saying. Make sure that the denominator is clear. When you give a group of numbers that add up to something, make sure that that total equals some denominator that you've made explicitly clear. So I agree with you there, but there's actually something I think is much more confusing here.

UNKNOWN SPEAKER: The last part (Inaudible)...?

PAUL Z. SIEGLER: Right. Why is that confusing?

UNKNOWN SPEAKER: Because you're on top of all the 124 patients that died.

PAUL Z. SIEGLER: Exactly. So this is telling us that they interviewed dead people. It's not even ambiguous. It's explicit here that they interviewed dead people. So let's talk about how to fix that problem. So let's look at it in a different way. Before I show you that let me ask you...okay there are how many different reasons for excluding those

subjects from the study...or including. 42 of them excuse me...oh excuse me...42 of them were excluded. How many different reasons were there for excluding subjects? From the study? Three. Okay if you have three or more of something and you write them down that's called what? Just an ordinary word...It's a list. Okay so you have three reasons for excluding subjects and so if you have three or more of something that is an opportunity to present it to the reader as a list which is not done here. It is not presented as a list but you could present it as a list. 42 of them were excluded from the analysis. It gets a little confusing to say they were excluded including...I think I don't like that. I think I will try and change that. But there you have a clear list of three reasons why study subjects were excluded from the study and so to come to the grammatical part of this. What's the proper way, what's the proper punctuation when you use a list?

UNKNOWN SPEAKER: Colon.

PAUL Z. SIEGLER: You use a colon. The two dots... Okay so you could present it this way...42 of them were excluded from the analysis: The colon tells the reader what's coming up is a list. So think about using colons as a good way to help your readers follow.

UNKNOWN SPEAKER: Here's a comma after the and?

PAUL Z. SIEGLER: Semi-colon?

UNKNOWN SPEAKER: (Inaudible)... I just wanted to find out the rule for the comma after the and.

PAUL Z. SIEGLER: The comma...?

UNKNOWN SPEAKER: After and. A-n-d.

PAUL Z. SIEGLER: Yeah, I don't think you need a comma there.

UNKNOWN SPEAKER: (Inaudible)...

PAUL Z. SIEGLER: Yeah people have got it...and, in that case I think you...

UNKNOWN SPEAKER: The comma after the and.

PAUL Z. SIEGLER: I don't think you need it there. Some people might...I think it's wrong to put a comma there. We'll talk a little bit more about punctuation that may help with that some. So there you have how it can be useful to use a list. Now I want to step back here. I had actually asked them to get these books for you and I think it wasn't in the budget so I had some extra copies so here you go.

UNKNOWN SPEAKER: That's cool.

PAUL Z. SIEGLER: Oh...there's one more person, right somebody...?

UNKNOWN SPEAKER: I already had one.

PAUL Z. SIEGLER: Oh, you already had that. Okay great. So if you go to rule 7 on page 7 at the bottom it says use a colon after an independent clause to introduce a list of particulars and a positive which I don't know what that is, an amplification or illustrative quotation. So those are the places when you would use a colon and introducing a list is one of the very basic situation where the colon is the best punctuation. Okay so this is one way to rewrite this...well it certainly is better because at least it's not bizarre. So now...

UNKNOWN SPEAKER: Question?

PAUL Z. SIEGLER: Yeah.

UNKNOWN SPEAKER: (Inaudible)...?

PAUL Z. SIEGLER: Could you use commas here...well...eleven had no known telephone or address and seven cases the relatives declined to participate and 24 cases no family member could be contacted within five call attempts...I think you could use a comma there, but...

UNKNOWN SPEAKER: (Inaudible)... number there, comma, (Inaudible)... put the number in the beginning (Inaudible)...

PAUL Z. SIEGLER: It's okay to do it that way. The semi-colon is helpful if each of the element in the list if it's long.

UNKNOWN SPEAKER: (Inaudible)... the element there is a series.

PAUL Z. SIEGLER: If there is a comma within the element then you would have to use a semi-colon. And there could be a small list within one element, separated by commas then you would have to use the semi-colon. In this case I think it might be optional but I think just because it's long...each of the elements is long I think it helps to use the semi-colon because it emphasizes to the reader to take a pause.

UNKNOWN SPEAKER: What about I thought that...when you have a double digit number or a single digit number you're supposed to actually write it out so eleven would technically be eleven and seven would be written out like it is. The number 24 is not written out...are there different.

UNKNOWN SPEAKER: (Inaudible)...

PAUL Z. SIEGLER: There are different conventions. This is one of those questions where the answer is...

UNKNOWN SPEAKER: (Inaudible)... not black and white?

PAUL Z. SIEGLER: The question is...is there a rule, the answer is yes there is a rule but there is no agreement on what the rule is so this is one of those cases where it's good to look at a recent article from your target journal and just do it the way they do it.

Okay so now we have...I'd like you to look at (Inaudible)... this principle of composition number 15...find the page for that...excuse me?

UNKNOWN SPEAKER: Page 19.

PAUL Z. SIEGLER: Page 19. Put statements in positive form. And so let me explain what I mean by that. If we go back...if we look at this notice it starts by telling us who was excluded. And what's more logical to start with what's excluded or to start with what's included...rhetorical question. It's more logical to start with what's included and then tell people what's excluded. So if you apply that principle here you have something like this...124 patients died of an avoidable cause relatives of 82 are interviewed. Well now we don't have to worry about making people think that the dead people were interviewed because you've told people what did happen. And once you explain what did happen that limits the possibilities about what didn't happen. So as a general rule, try to put your statements in positive form. Begin...if you need to tell people what did happen and what didn't happen, try to start with what did happen part. Then do the did

not happen part. So the remaining 42 were excluded either because they had no known telephone or address, relatives declined to participate or no family member could be contacted within five call attempts. So two things that I think you will find helpful...one is use colons for lists. Think about presenting things as lists. Use the colon if you're going to do that. And then try to present things in positive before you present the negative. So that's it for methods. And let's see how close are we...its 2:00, break is at 2:30 so I think let's just take a two to three minute break and then we'll start off with results. Okay, don't stray too far.

Now we're on page 54. Move from methods into results. To me it's helpful to think about the results this way. It's the place where you describe the results of the data analysis again relevant to the study purpose. And the reason I emphasize this is again just that idea that you've conducted an analysis. Maybe you've spent three months, four months, six months, eight months analyzing data. You've got a mound of data in front of you. But for the purpose of the paper your goal is to focus your presentation of the data that you actually analyzed that are relevant to the purpose of the study. So you have to decide for yourself what is the purpose of this study. And it's not anything magical and it's not rigid. You might actually change. You might start out thinking okay here's the purpose of my study and you might do some data analysis and you might say oh I didn't realize that I had no idea. You might actually modify the purpose of the study. Well if you do that you might include some data that you weren't planning on including in the study or you might exclude some data that you would include in the study. Try to maintain a very close and coherent relationship between the purpose of the study and the data that you

show. Okay this I think is in the list of the five most important take home messages that is when you write the paper, start with the tables and figures and write the text later.

This is one of the most important things because it is very easy to write an entire paper and spent three months, six months writing an entire paper and then find out that you went off course somewhere and you could literally spend tens to dozens to a hundred hours or more and then you find out six months later oh boy I kind of made a left hand turn when I should have made a right hand turn and you could literally be looking at a hundred hours of work that is not productive. But if you start with the tables and figures I think you won't have that problem.

So let's talk about tables and figures. When would you use a table and when would you use a figure? What's the difference? Thoughts anybody?

UNKNOWN SPEAKER: Well you had an example where you showed the game the subject group and how you excluded and came down to the final number for the analysis. It's more of an illustration a figure is. For me, usually and a table is a clustering usually of analysis in a formatted way.

PAUL Z. SIEGLER: Okay well the content of the table is usually what? A group of...? Numbers, right. A table is typically a group of numbers whereas a figure and you mentioned the flow diagram which is a good example. I wasn't thinking about that. I was thinking more of a figure with an X axis and a Y axis. A content of a figure is typically what? A table shows numbers, a figure typically shows...comparisons.

UNKNOWN SPEAKER: She said it first.

PAUL Z. SIEGLER: Okay good. Use tables when your goal is to highlight individual values because that's what's in a table is individual values. If you want to highlight trends and relationship almost always a figure will do that better. So this is an example where this is not trial and error. It doesn't need to be trial and error. You can say to yourself okay what do I want to emphasize? Am I most concerned about an individual value or a group of individual values? Or is my core message a relationship? Trend being included in relationship. Trend or relationship. If you're most concerned about a trend or relationship then almost 100% of the time a figure is going to be better. If your major concern is an individual value or a group of individual values then table will probably be better. So here's an example of a table from (Inaudible)... risk factor surveillance systems. I'm not going to ask you to look at it. The point is that it contains data from 40 or 50 different states over a period of several years and because it has individuals state data there's a readership out there for all of the numbers that are in that table. If you're from Indiana you will want to see the Indiana number. If you're from Wyoming you'll want to see the Wyoming number. There is no meaningful way to summarize the data and present the critical point when you're showing surveillance data from a variety of geographic areas. So surveillance data with geographic specific numbers is a classic example of when a table is a better way to present the data. Now in this case there was five years of data and when we looked over the five years we saw that there were in fact some trends so we did a figure like this which illustrates a couple

of different things. These are smoking prevalence's. Maybe I'll just change this from (Inaudible)...maybe I'll just make these numbers more recent so people won't know how old I am. But it's a five year period and you can very easily see that in this low education group that the smoking prevalence was pretty much constant over the five year period but that in this total population group age 18 or older there is a downward trend. Now there is no confidence intervals, there's no inference testing here this is just eyeball stuff. But it's pretty clear that the gap here is bigger here. So that's a take home message from this that the gap between low income smoking prevalence and low income people or low education people compared to the total population was growing from 1986 to 1990. So with a single figure you can show trends over time you can also show relationships between comparison groups and then you can benchmark. So this is at this time...the year 2000 objectives. So again tables highlight individual values, figures highlight trends and relationships.

UNKNOWN SPEAKER: Question. Sometimes when you have the figures and so if you do a year maybe the data (inaudible)...collecting (inaudible) maybe for seven years that data was not collected but you have a trend analysis (Inaudible)...we don't have data...(inaudible)..

PAUL Z. SIEGLER: You can show a break. Some sort of a slash. Yeah that's one way you can do it. Now if you're doing a trend analysis then you have to use some type of statistical technique that accounts for missing data. So that's more complicated but just

to show it on a graph you can use a break point. Some sort of a slash and then probably with a footnote at the bottom of the table saying no data available for a specific year.

UNKNOWN SPEAKER: Okay.

PAUL Z. SIEGLER: Another thing can be careful is to check your math. It's really easy to make calculation errors so be careful. Let's see...let's see this is...yeah this is in your hand out page 55...so if you would take a look at page 55. So this is the point that we were talking about earlier the denominator is obviously shifting but the author is not telling us how the denominator is shifting. So...just be aware of that when you give numbers that are numerators be sure that the reader knows what the denominator is. Both the number and the characteristics. The reader should know what is the number for the denominator, but should also know the characteristics of the study subjects who are in that denominator. Okay so that was an example of sort of a wildly horrendous or wildly horrendous example of something that's not well written. Here you have something that's really much more subtle and the theme here is calculation errors. Can you find...there is a calculation error in this table? Take a look for a couple minutes and see if you can find it. I may as well give you a hint.

UNKNOWN SPEAKER: (Inaudible)... doesn't add up to 743...

PAUL Z. SIEGLER: We are going to talk about that later. That is not...

UNKNOWN SPEAKER: Oh that is the number...

PAUL Z. SIEGLER: Okay...

UNKNOWN SPEAKER: (Inaudible)... the total in the (Inaudible)... they didn't get the female. That should be a positive, so...

UNKNOWN SPEAKER: (Inaudible)....

PAUL Z. SIEGLER: Hey anybody could make that mistake. I'll give you a little hint. It's in the percentages. It's not the numbers...the numbers are correct. There's a miscalculation...

UNKNOWN SPEAKER: (Inaudible)... well there's 11% of male (Inaudible)... 13 so it would be there...

PAUL Z. SIEGLER: Okay so that is the mistake so how can you tell...how do you know it's a mistake?

UNKNOWN SPEAKER: Well if anything it would decrease.

PAUL Z. SIEGLER: Right...okay. The denominator here is 5782 and the denominator here is 7080 so the percentages here have to be lower than the percentages here. So

for example 66 is lower than 70. 18 is lower than 22, but 13 is higher than 11 so that can't be correct. So let me show you how that happened. So that's the mistake or this is correct but this percentage, this 13 is incorrect because it has to be lower than 11 because the total is lower so moving along...let's see 11% would be the correct number here. I don't know if you have a calculator or if you're good at doing division maybe I'll just show you. Here it's 628 divided by 5782 which is 11%. This is 628 should be divided by 7080 and so that's 9%. Okay 628 divided by 7080 is 9%. Now the error is because the previous table used the 4639 as the denominator. So again it's related to shifting denominators. If you divide 628 by 4639 that's when you get 13%. 628 divided by 4639. So it's very easy for these kinds of mistakes to happen. There's no protection against it other than just being careful. Being careful and getting feedback, so be careful but in addition to being careful get feedback. Have people read your papers.

Now...okay...so next point that I want to emphasize and we're going down the list on page 54 now. We're under tables and figures the fourth bullet under tables and figures on page 54 is provide consistent row or column summation. That means that you could choose. Either the columns add up to 100% from top to bottom or the rows add up to 100% from left to right. Don't mix those two in a single table, which is exactly what we have here. So here we have 100% here and if you add these percentages 70 plus 22 plus 4 plus two plus two it equals 100, but when you get down here it says 82. And here when you have 70 plus 47 you have 66 so this table mixes. It shows that in some places the percentages add up across the rows and in other places the percentages add up down the columns so you can fix that by in this case putting 100% at the bottom of each

column. Keep lines to a minimum. Avoid vertical lines. I just want to show you a couple of examples not even to ask you to read the table. Just to show an example of a typical table from a peer review publication. Notice there are two horizontal lines at the top, well not quite at the top, there's a horizontal line underneath the table heading and then there's another horizontal line underneath the column headings and then there's a third horizontal line at the bottom. That's very typical. Many journals organize their tables that way. Another example here this table actually it does have vertical lines but they're just vertical lines along the border. The right-hand border and the left-hand border. There are no vertical lines in the interior of the table. Very few journals almost no journals two will have vertical lines in the table. And the reason for that is that the lines are not information and so if there are a lot of lines in the middle of the table the lines are actually competing with the numbers for the reader's attention. Use footnotes to clarify points of potential ambiguity. So here you have IDU with an asterisk, IDU equals injecting drug user. This is an example of a table that has a lot of footnotes. They are labeled A through G. But notice that above the A through G...I'm sorry is this in your...yeah this is on page 56 of the handout. So notice that there's A through G and they correspond from top to bottom as you go through the table. There is A and B, C, D and so that's the way to organize them generally is from top to bottom. A is the first one that appears and then comes B, C, D and then as you go down the table then you will find E, F and G, I think.

UNKNOWN SPEAKER: (Inaudible)....

PAUL Z. SIEGLER: Oh yeah there is E. Are F and G in there? Oh yeah there is G. And I think the F should be in there somewhere but general principle is you start at the top, you go to the bottom and then if anything...if there are two footnotes that appear on the same line the one on the left comes first before the one on the right. But I also want to point out that there is this other footnote here that it says note...and so be aware of this as a strategy that you can have footnotes that pertain to a specific point in the table and it could be a sequence of them that would be A, B, C, D, E, F, G or they may have the little squiggly dagger different symbols but in addition you can include a footnote that applies broadly across the table. If you have a footnote that applies broadly across the table then you may not give it a letter you may call it something like note. And that alerts the reader to realize that is not a footnote that applies just to a specific point on the table. It's a note that applies broadly across the table.

Check the heading labels of the rows, columns and axis and footnotes.

UNKNOWN SPEAKER : ...that there were two on one line. You did say that the first one should be to the left and go (Inaudible)...

PAUL Z. SIEGLER: Yeah.

UNKNOWN SPEAKER: (Inaudible)... is the F and then over across prevalence is the E so that's balanced isn't it?

PAUL Z. SIEGLER: It might be I would say no because this is really parallel...oh sorry.

UNKNOWN SPEAKER: I see, you're making it across the top.

PAUL Z. SIEGLER: Yeah, see this is really parallel to this and then this is kind of below.

Well it's a little...oh you found the F. So that's the reason why it shows up that way.

Okay so check heading labels of rows, columns, axis and footnotes and I think it's time

for a break, right? 2:30? So let's take our break and when we come back I'll ask all of

you to look at the table on page 57 and we'll start looking at ways to improve this table.

So when you come back please look at this table on page 57 and think about things that

we could do to improve this table.