

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: Ready to go, ok. So that is one way that you could make it shorter and then there is another idea that appears in there twice. It is not repeated language, but there is an idea that appears twice.

UNKNOWN SPEAKER: (Inaudible)...

PAUL Z. SIEGEL: Yes, test retest study, ok. If you take this out you have the description of the test retest study. So you don't have to tell the reader that it is a test retest study because you have already described a test retest study. Then there is one more thing that you can do which is this between six and thirty days. You could just show it this way, six to thirty days. So that original of thirty four words comes down to twenty three words, so there you are. You would be under your two hundred word limit. So again, very often this idea of critically evaluating things is really driven by the practicality of the context. I'm not saying that you should do everything in your power to make your abstract as short as possible; I'm not saying that, I'm saying that you are working in a

practical context. You are writing for a journal and the journal is telling you that your maximum word limit length is X and if you write a draft that is over the limit your goal then becomes to make it shorter without removing critical information.

So now let's talk about results and before we go over to group three a couple of comments that I would like to say from the beginning is that I would say that the results portion of the abstract is organized into three subsections. There is this first part which is about sociodemographic variables, chronic conditions and risk factors, it may look a little familiar to you because I used some of that language for the exercise in the very beginning. Then there is a portion on cancer screening practices and a portion on environmental tobacco smoke. So there are three elements, three components to the results portion, and the question that I would like you to think about is what is the readers ability to go back and forth between these three sections and say oh yeah, I can see that the reliability of the questions in this group up here is higher or lower than the reliability of the questions in this group down here. Say for example, to compare this group of variables to the reliability to this group of variables to the reliability of this group of variables, what is the reader's ability to do that? High, low, medium. It's low. Perhaps zero. What is the main reason why that is true? What is the main reason, let me say it another way, what is the main thing that the authors could do to make it more possible to make the readers make those comparisons? What could you do, how could you modify the abstract to make that happen?

UNKNOWN SPEAKER: (Inaudible)...

PAUL Z. SIEGEL: Ok, that's great. So two very important points, one is you could add more CAPA values, it would only be a few words because you wouldn't have to give the CAPA value for each variable, you could give ranges. You could group them together in ranges. Then present it using parallel language. Let me show you a little bit about how you could do that. Now in this first part here, how could you make that shorter? Just that first portion right there, how could you make that portion shorter?

UNKNOWN SPEAKER: (Inaudible)...

PAUL Z. SIEGEL: Ok, exactly. This is almost exactly the example that I showed you at the very beginning. So reliability of the information on chronic conditions, this fraction is also high so you can roll those together. Notice that you can roll some of the words together in a single sentence, but also look at this, this range here is point eight five to one point zero. Then here we have another range, point eight two to one point zero. They are similar, right, they are almost identical, and so the point eight two to one point zero actually includes the point eight five to one point zero. So if we put all of those variables into one sentence then we can use the point eight two to one point zero, we don't need the point eight five to one point zero anymore. So that could be combined, that is an effective way of making the first part shorter.

So now I think let's move onto the second part of cancer screening practices. Here you have regarding cancer screening practices reliability was lower for prostate specific antigen tests and for women cancer screening practices now what would happen if you removed cancer screening practices, what would happen if you began relying on prostate specific antigen tests then for women's cancer screening practices? What affect would that have?

UNKNOWN SPEAKER: (Inaudible)...

PAUL Z. SIEGEL: Very little, ok.

UNKNOWN SPEAKER: (Inaudible)...

PAUL Z. SIEGEL: Right, ok, so you lose a little or maybe even no information if you just take out cancer screening practices because the sentence is about cancer screening practices. Now I would say that for many people familiar with public health, people who are reading public health literature, most people will know that mammogram and Pap smear are the major women's cancer screening practices. So I would say that you probably don't need both. You don't need to say women's cancer screening practices

and then reinforce that with mammogram and Pap smear. One or the other is probably just fine. Then questions on attitude on environmental tobacco smoke, there as I mentioned earlier, you could add some information about the CAPA values , now it becomes possible for the reader to go back and forth and make those comparisons. Did I say everything that you were going to say or did you have anything to add?

UNKNOWN SPEAKER: Pretty much, I think we said reliability for male and female cancer screening practices and the range is point five nine and point eight seven with the exception of the PSA test. Similar to what you were...

PAUL Z. SIEGEL: Got it. Yeah, that sounds perfectly reasonable. So if we do those edits we come up with something that looks like this which is quite a bit shorter than the original.

UNKNOWN SPEAKER: (Inaudible) more detail.

PAUL Z. SIEGEL: And it has more information, critical point being just adding those ranges for CAPA values. So now notice the passage has three sentences and all three sentences have the word reliability in it. That is just one of those things that people react to; three consecutive sentences that all have the same substance of words, people start to experience that as boring and monotonous. So how about that, do you see a way

around that? You mentioned parallel structure, can you maybe be a little more specific because I think one solution to this is related to that concept of parallel structure.

UNKNOWN SPEAKER: This maybe too much, but reliability is high for (inaudible) demographic areas and CAPA equals (Inaudible). Chronic conditions, risk factors CAPA equals the new range and male and female cancer screening practices with a range with an exception for PSA and the CAPA score.

PAUL Z. SIEGEL: Ok.

UNKNOWN SPEAKER: There was more reliability (Inaudible) CAPA equals range. Then actions to reduce the exposure to tobacco smoke range.

PAUL Z. SIEGEL: Right, very good. So you have evoked the concept of hierarchy. Higher and lower, and that is really important. If an example of presenting the results in some logical framework that the reader will recognize, that will be meaningful to the reader. So it is not the only way of doing it, but the general principle is trying to present results in some sort of a framework that readers will respond to, will recognize the framework and high, low, lower or high, lower, lowest is a very good way to do that. So it could become something like this. Reliability was high for one group, lower for another group, and lowest for another group. Then the final sentence switches to another

grouping so you don't have to put all of the data into that framework, you can put a portion of the data into that high, lower, lowest framework and that then frees you up to present the rest of the data, perhaps in a different framework.

Just one final comment here. The earlier version said that reliability was high for sociodemographic variables, this is just a prompt to remind me to emphasize that when you present data, it is important to be neutral. There is a place to interpret the data which is generally in the discussion section, but when you are presenting the data in the results section try to avoid anything that people might view as interpretation. So in this case to say that reliability is high, somebody might disagree with you, they might challenge you on that. In this revision now it is highest, lower, and lowest. So it is completely neutral. There is no interpretation whatsoever. That is your safest bet. When you are presenting data try and just present the numbers without interpretation, reserve your interpretation for the discussion section of the paper. Original eighty seven words, the revision is down to sixty five and really has more information than the original version.

Now let's move onto the conclusion that says that these findings demonstrate the overall flexibility and utility of behavior risk factors system, and let's ask our group in the back of the room to comment.

UNKNOWN SPEAKER: We think that (inaudible) these findings (inaudible) the overall flexibility and the (inaudible). It is not actually the overall system (inaudible) conduct the research on. (Inaudible)...

PAUL Z. SIEGEL: Right. I think what you are saying is a good way to conclude this abstract is to give a summary of the results in the context of reliability because the abstract is about reliability. That is what you are saying?

UNKNOWN SPEAKER: Exactly.

PAUL Z. SIEGEL: I think that is a good comment. Now let me go one step further with this to say that when you write conclusions for papers there is a spectrum of categories of conclusions. One type of conclusion is to summarize the results. Another type of conclusion is to make some sort of recommendation. Now if you make a recommendation it should always be supported by the data in your study, but that is very different, making a recommendation is very different than presenting a summary of the results. I'd like you to think for a moment, in this paper are there any action oriented recommendations that are justified by the data presented in this study? Can you think of any recommendations that you might want to make and that you could justify based on the results of this study?

UNKNOWN SPEAKER: Well one item needs to be revised here, it is minor, and I don't know if I would include that.

PAUL Z. SIEGEL: Well, you could, it is something to think about. You are talking about the Prostate Specific Antigen question, which obviously has a very low reliability. So your point is yes, that is true, but you might not conclude with that because it is in some ways minor, well you can. I talked about the two ends of the spectrum; you can do both of those in the same abstract. You are not limited to only summarizing the results or only making a recommendation, you can do both. So you could give something like this we conclude that the reliability of most of the questions of the BRFS is high, but new more reliable questions need to be measured, need to measure and knowledge of Prostate Specific Antigen testing needs to be developed. So you can do both, you can give a summary of the results and then also make a recommendation. The general guideline is a practical one. Some studies generate results that support action, some studies don't. If you do a study and it generates results that you think justify an action step, then I think it is better to include that action step than the abstract. If you do a study that generates some results that you think don't really justify or support a specific action step, then in that situation you are going to be better to just give a summary of the results in the conclusion.

So now to summarize all of this a little bit, the abstract is telling us that the authors are concerned about reliability and validity. When we go to the methods we find out that the

study is only about reliability. When we read the results we find that the results are about reliability just as the methods described. Then when we get to the conclusion, the authors tell us about flexibility and utility, which makes us confused. There is another thing to think about, which is to compare the conclusion to the purpose of the study. Obviously there is a break in the logic between the conclusions and the purpose. So part of the reason that I show this to you is to recommend you write abstracts, to actually do this. To go through it one step at a time and ask yourself, "Are the methods with the purpose? Are the results consistent with the methods? Did I draw a conclusion that is consistent with the results, and then do this additional step of comparing your conclusion statement with the purpose statement. Ask yourself, "Did I maintain a consistent flow throughout the entire abstract?" In the case of this abstract that did not happen. So here and maybe someone could just tell us which page is this on, so everybody could look at it. So this is a revised abstract that includes pretty much all of the comments and the revisions that we talked about. What page is that one on?

UNKNOWN SPEAKER: (Inaudible)...

PAUL Z. SIEGEL: Excuse me, nineteen. Ok so that is on page nineteen.

UNKNOWN SPEAKER: Page eighteen.

PAUL Z. SIEGEL: Page eighteen; we'll get to page nineteen in a minute. So here you have a brief description of the behavioral risk factor surveillance system, statement that is used by nearly all state health agencies, which is more specific than saying that it is widely used. Here you have the denominator information so the reader can understand what the response rate was. Then you have the hierarchical presentation of the results, high, lower, lowest. Additional CAPA values, which makes it possible for the reader to make all these various comparisons, then a conclusion, which includes both a summary of the results and a recommendation, a practical recommendation. I think by this point, you may be thinking he picked this because it is a terrible paper and it has all of these things that are wrong with it, and it is really easy to pull it apart. That is actually not my full intent. I've selected this because I think it is an example of what I think is a good study, but a poor abstract. Let me take you into the historical context to explain why I think it is a good paper and a bad abstract.

This study was done in 1992, I think conducted in 1993. Most of you were too young to remember that time clearly, but I remember in 1993 Prostate Specific Antigen was a forefront public health issue, the test was available, and public health had no idea of what to do with it. In fact, I think I am telling you accurately that congress at that time even allocated money for CDC to do surveillance work with Prostate Specific Antigen and the Cancer Division turned it down because they didn't know what they would do with it. There was just so much uncertainty about the test at that time. The same is true for environmental tobacco smoke, now we are all familiar with indoor air, legislation and states, and in the workplace. At this time those things did not exist. This was a very

forward thinking study. The context is the authors took two cutting edge public health issues and applied those issues to a system, behavior risk factor surveillance system, which is a federal system and as most of you know federal systems can be rigid and inflexible. So the goal, the objective of the study was to see, could Missouri use this federal system, which is potentially flexible and rigid, in a flexible way. Would it be useful for them to do that? So they put these environmental tobacco smoke questions, Prostate Specific Antigen screening questions into their state behavioral risk factor surveillance system and tested how it would work out. They found that it worked. They were able to conduct the survey, they were able to get answers from people and they were able to get people, who looked like they were able to integrate those new questions into the larger questionnaire. Although, they did not use the full questionnaire, I don't know if that became clear to you from the study, they did not use the entire BRFS questionnaire, but they incorporated these special questions into some of the other basic elements of the behavioral risk factor surveillance system. I think it is accurate to say that they did test the flexibility and the utility of the system and I think that they found that it is both flexible and that it was useful, at the same time they tested the reliability of questions, both of these new questions as well as some of the standard questions that are in the BRFS. So I think what happened is that they did a study that had two distinct and separate purposes, one was to test reliability and the questions. The other was to test the flexibility and the utility of the behavioral risk factor surveillance system to accept new forefront issue questions. When they wrote the abstract, when they described the purpose in the abstract, they described one purpose, and when they got to the conclusion, they addressed the other purpose. That I think is

what happened. So a major reason why I think this is helpful to you is to emphasize the point that when you write papers, try to keep your papers focused on a clearly circumscribed issue. Describe that purpose in the beginning of the abstract and follow through with the methods that are related to that purpose, results that are related to those methods, and a conclusion that follows through consistently and relates directly back to the purpose. If you do a study that has two purposes, you don't have to put it into one paper, you can write two papers. You probably heard about the Salami Science people who try and take one study and publish the maximum number of papers. Well that is abusive, but I think that this is an example of a study where there were two very distinct purposes. It would be completely legitimate to write two separate papers, and it would be much easier to write a good abstract.

I did go ahead, and this is in your handout, I'm not going to spend more than a few seconds on this. I tried to see if I could write an abstract that would address both purposes at the beginning of the abstract, and talk about the results and conclusions that are related to both of those purposes. There it is, you are welcome to look at it. If you look at it with a microscope, the same way I looked at the published abstract, the original abstract with a microscope, you will find lots of stuff that is wrong with it. So it is there as an example of how it may be possible to write a good abstract that addresses two purposes and draws down to the conclusions related to those two purposes, but in general I don't think that it is a good idea. It is very difficult to do. Before we leave abstracts, just a couple of final comments, on page twenty I want to make sure that I talk to you about structured abstract. Now most journals require a structured abstract,

but not all journal, there are still some journals that use unstructured abstract. If you look at the abstract on the top of page twenty, you will see the headings for that structured abstract that one comes from JAMA. JAMA actually gives authors flexibility in the headings that you could use for your abstract. If you go to the instructions for authors you will see the guidance for authors about some of the options that you have when you structure your abstract. The abstract on the bottom of the page is from New England Journal, and they still use this traditional format of background methods and conclusions. The only way to know how to structure your abstract or whether to structure your abstract is to go to the instructions for authors, which means that you have to chose your target journal. We will talk later, tomorrow afternoon about a systematic process about selecting a target journal.

So that finishes up for abstracts, we are going to move on to introduction, but I think we will take a stretch break for a minute. So if you want to stand up and...So purpose of the introduction, one way to think about it is to convince the reader that you conducted a study that is going to generate some sort of new knowledge or knowhow that will be useful. So I mentioned this earlier in the morning, the new and useful, that is what people are looking for, that is what readers are interested in, and it is certainly what the editor's are interested in. Editors, their job is to publish papers that contain new things, but not just anything that is new, it is supposed to be useful, that is what they are looking for. So anything that you can do to emphasize the newness of your study is going to raise your paper in the eyes of the editors.

Now components of the introduction, we are on page twenty three now. Identify a gap in knowledge or know how. There is something out there that we don't know yet, or that we don't know how to do, and your paper in some way help us learn something that we don't know or learn something that we don't know how to do. People will call that the study problem. When you write about this gap, provide key background information and typically it is about the scope, the nature of the magnitude of the gap, so it could be a disease rate in a special population. It could be a prevalence of a risk factor. Try to be fairly specific when you describe the scope or magnitude of that gap. Be clear that filling the gap will be useful. I will show you a couple examples of that. Describe the relevant limitations of previous studies. Again this is related to the gap, you have done something that is new, try to be explicit about what is new and usually you are like the old story of the midget climbing on the shoulders of a giant. There is the background literature out there if you have done something new. It doesn't have to be monumental; it just has to be new in some meaningful way. So you have this first part of the introduction, which is about a gap that you identified then in the next part of the introduction you present your approach to filling the gap which is more or less the same thing what some people call the study purpose. Be clear that your approach is new. Whatever is new could be different population that has not been studied before; it could be a survey method that enables you to get a sample of that hard to reach population that is more representative of that population. There is lots of ways in which your study could be new and doesn't have to be earthshaking, just has to be different from what has been done previously, that in some way is an improvement from what has been

done previously. Emphasize that your approach addresses the limitations of previous studies in some logical and compelling way. So again, new and useful, very often this can be done in just three paragraphs. It sometimes can be done in two paragraphs. You could literally have one paragraph talks about the gap with those pieces of information related to the gap and another paragraph that talks about your approach to filling the gap.

Next page you have "Elements of the Introduction," this is really just a repetition of what I just said, but I put these little symbols here because you will see how they come up in the next page. So you have the gap which is key background information to the gap, something about why filling the gap will be useful, and then key limitations of previous studies. Then in the portion about your approach to filling the gap, some background information to the approach, not always necessary, it could be in some cases like we did first some randomized double blind trial study such and such, intervention, and it would be clear that you wouldn't have to explain to people that randomized double blind trial is considered to be the gold standard. So some background information about the approach, what is new about the approach, make sure that you are clear about what is new, and how does this new approach address the limitations of previous studies. So you've got something up here about the limitations, the previous studies, and something down here about how you have done something new, and try to write it in a way that will draw the reader's attention back here. So you are saying that we did such and such, try to be clear that such and such that you did is different from what has been done previously and is in some way superior.

So now please turn over to page twenty six, I will ask each of you individually to read this introduction on page twenty six, look at the questions on the bottom, and just have a brief discussion with a neighbor and try to answer the questions on the bottom of page twenty six. What is the gap?

UNKNOWN SPEAKER: (Inaudible)...

PAUL Z. SIEGEL: Did somebody make a suggestion?

UNKNOWN SPEAKER: The section about where you talk about the association...

(Inaudible)...

PAUL Z. SIEGEL: Let's see, you mean this portion here, the early Med Analysis fail to demonstrate an association between workplace CTS, so they are describing some of the literature.

UNKNOWN SPEAKER: (Inaudible)...

PAUL Z. SIEGEL: Let's see...approach, well here they are describing the literature. It says that "The association has been reported to publish the three most recent Med Analysis." So this to me is a little bit more of the background to the approach talking about the background. They are talking about studies and what they think may be limitations of studies, limitations of previous studies. So I would say this more so as the statement of the gap. The demonstrating and the association of workplace CTS exposure and lung cancer has been more difficult. They are saying that up here they are emphasizing that the association in the workplace has been clearly documented but that documenting the association in the home setting has been done, documented clearly, where as in the work setting the association has not been documented clearly. So to me this is the statement of the gap, and I think it is clearly stated. How about the next question, does the study approach address the limitations of the previous studies in a logical and compelling way?

UNKNOWN SPEAKER: Second sentence?

PAUL Z. SIEGEL: Second sentence?

UNKNOWN SPEAKER: (Inaudible)...they always demonstrate an association between

PAUL Z. SIEGEL: Uh huh. Failed to demonstrate association between workplace CTS exposure and lung cancer risk among nonsmokers, but it says that “Significant association has been reported in three most recent published Med Analysis.” Now this is a summary of the literature, it is describing the results, but it is not really describing the limitations. It is not saying here is a shortcoming.

UNKNOWN SPEAKER: (Inaudible)...

PAUL Z. SIEGEL: Right, I think it is here, “We saw to extend previous Med Analysis by including additional studies and by conducting analysis stratified by level of exposure. So they are saying that they have done two things that are new. I think that this is a good example of an introduction that makes a very clear statement here. Here is what we have done that is new and it hasn’t been done before. So I think this here is the statement about how the study approach addresses the limitations of previous studies and I think in a logical and compelling way. Then how about why filling this gap will be useful.

UNKNOWN SPEAKER: It doesn’t really say much, it looks like they are going to (inaudible) response, but they don’t.

PAUL Z. SIEGEL: Ok, I think you are exactly right. There is nothing in here about why this study will be useful. So that is something to think about. They could say something like, for example, if this association in the workplace is documented it could lead the way toward new public health programs that could have an important impact on reducing tobacco related illness. That would be a perfectly reasonable thing to say, but they don't, in this case they don't. To me, the abstract would be a little stronger if they included that statement about why it would be...why the study is potentially useful. So an example of an introduction section that does not contain any information about why the study would be useful, but I think it would be stronger if it did have that. This is pretty quick, but do you feel that you have a good sense for how you can organize the introduction section often in just a couple of paragraphs. Even just a couple of paragraphs could be enough for you to give adequate literature review and focused very clearly on ...it is a literature review that is focused on and around what is new in this study. It is not a broad sweeping literature review. Just again, any questions before we move onto methods?

Ok. Let's go ahead to methods on page...let's begin on page thirty, skip a couple of pages ahead. So we are going to skip a couple of pages, we just don't have enough time to do everything. Actually let me make one more point before we go onto methods. I mentioned just a moment ago that this is not a broad sweeping introduction or the literature review is not broad and sweeping and that is a tendency that many people have, to write an introduction that is broad and sweeping. That actually can cause problems. Number one it can consume more words than you have room for, but then

another potential drawback is that it can actually draw the readers focus away from what is new about your study. You can lose the readers focus by giving a broad sweeping introduction. So just as a quick, sort of whimsical attempt to illustrate that point, here you see a jigsaw puzzle, if you can imagine that you are writing and introduction for a paper and the setting is this, that I'm taking my daughter for a hike with some of her friends, and we want to make sure, we have decided that we are going to take bananas for a snack on the hike. We want to make sure that there are at least six bananas; there are the pieces to our puzzle. It sort of looks like there might be six, but there are some pieces missing there. So if you look at it this way you can see one, two, three, four, five, well there might be six. So there is just one piece of the puzzle missing, notice there is just one piece of the puzzle missing. When you fill in the piece, now it is completely clear. There are definitely six bananas, might even be seven, but it is me and her and four of her friends. We want to make sure that each of us have at least one banana, so we are set. So this is a little bit like writing the introduction section to a paper. You give the setting and you try to carefully write the introduction so that there is just one piece of information missing. What you explain in the introduction is how you are going to conduct a study to provide that one key piece of information, when you fill in that blank, now the picture becomes complete, which is different from something like this where you can still see that there are some bananas there. It is not clear whether there is six. You write the introduction this way, perhaps to explain what I've just explained and then you fill in a piece of the puzzle. Now you've done a study, you've added a piece of information, but we still don't know whether there are six bananas or not. This is something to be careful that you don't do, don't write a study or

a paper where you add a legitimate piece of information, but after the new piece of information is added, we already know that we still haven't gotten to the purpose. We still haven't answered the question, haven't filled the gap that the author's told us. So be careful to avoid that scenario. Now I was very careful to pick a puzzle that has six pieces because I've already talked about the gap, the approach to the gap, and that there is one way to approach this is that there are three essential pieces to describing the gap and three essential pieces to describing the approach to the gap. There doesn't have to only be six pieces to the puzzle, there could be more pieces to the puzzle. So I very carefully told you a story about how my daughter and I and four of her friends are going on a hike. That might not be the whole story, I could say things like well we are going on a hike and we had discussion about a variety of different snacks that we wanted to take on the hike. We thought about apples, we thought about sandwiches, we thought about gorp. We spent about thirty minutes discussing which type of snack that we would like to have. Some people wanted something that was healthy; some people wanted something that tasted really good. Those are all pieces of background to the story. So there are many pieces to the story and one of your key challenges when you write papers is to reduce the story to a manageable level. Again this is part of keeping the reader, the train of thought for the reader. If you write an introduction section that is broad and sweeping it becomes more difficult for the reader to maintain a focus on what you want the reader to stay focused on. So the introduction could have twenty pieces rather than six pieces.

Here is another kind of scenario where you've got a puzzle with some pieces, but no matter what piece you fill in, the reader is still not going to know that there are six bananas. Then you have this kind of thing, this is the paper that says bananas are an ancient fruit, they were first cultivated in South America, and then nobody knows how they got transferred from South America to Africa. Many studies have been done to analyze the nutritional content of bananas and in some cultures they are considered a mainstay, where in other cultures they are considered to be a treat. Well, you could write your abstract that way also. That doesn't help you help the reader recognize and stay focused on the fact that you are trying to figure out that there are six bananas. So when you write your introduction, this is my recommendation and my plea, try to resist the temptation to write a broad sweeping introduction. It may make your paper too long, but even worse it can distract the reader from recognizing the focus that you are trying to bring to the paper. So there is the banana. We will skip over this. Now let's move onto methods.

So page thirty, this one is an exercise, just as we have done before, I would like each of you to read that little passage at the top of page thirty, look at the three questions at the bottom of the page, talk it over with a neighbor, and then we will discuss it. I think five minutes. We will begin discussion, we actually it is almost twelve o'clock so I will just have you do this reading now then we will do the discussion after lunch. It would be in the methods section in other journals.

In other articles it would be in the results section. So I don't think that there is a consensus about exactly where the response rate goes. So if you encounter a situation like that you don't know what section of the paper to put something in, really a very helpful idea is to look at a recent issue of whatever your target journal is and just see how they handle it. How about question number two, is the description of how studies subjects we identified adequate? Somebody comment why?

UNKNOWN SPEAKER: It doesn't say how they were recruited, if it was random, if it was a paper survey or a phone survey, it doesn't say anything.

PAUL Z. SIEGEL: Right, lots of more details, we could highlight a lot more details, but one that I think is critical is it doesn't say what the sample frame is. That is a critical element of any survey is to describe what the universe or target population is and what is the sample frame. So how did they identify, how did they know those, that there were one hundred and sixty seven.

UNKNOWN SPEAKER: Is there any reason to avoid in saying which state it is?

PAUL Z. SIEGEL: That is a judgment call. Some people might feel that if the results of the study might get a negative impression then that would be a reason not to do it. Basically, not trying to offend people is one reason to be discreet about identifying

information. Then question three, “In what way are the methods described above consistent with the following excerpt of the abstract in the same article?”

UNKNOWN SPEAKER: Randomly chosen versus self selected.

PAUL Z. SIEGEL: Right, big difference there.

UNKNOWN SPEAKER: Healthcare units versus NICU.

PAUL Z. SIEGEL: Right.

UNKNOWN SPEAKER: And upper Midwest U.S. versus Midwest states.

PAUL Z. SIEGEL: Ok that is a good summary. I think that you got all the key points that I wanted to highlight. That is pretty dramatic and this is an excerpt from a published paper. There was literally one version in the abstract and another version in the results section. Part of the reason that I show this is really to just caution you to be careful about what you read, don't just believe anything that you read just because it has been published.

Next if you move over to page thirty one. This next exercise goes from page thirty one to page thirty five. I'm not going to ask you to read all of this, I would just like you to look at the brief study introduction on page thirty one, look at the methods section , which is one pages thirty two to thirty five. First of all, recognize that it is very long, it is the methods section, and it is over one thousand words. I think it is even over fifteen hundred words. Imagine that you could be writing this article where you could be have only three thousand words for the entire article, maybe even twenty five hundred words for the entire article, most of the time you won't have fifteen or seventeen hundred words to devote to the methods section. What I would like for you to do here is not to read it word for word, but look at it in its outline, from the big picture perspective. Think about strategies for making it shorter. So again, not to edit it line by line, but think about what strategies that you would use to cut it, say by fifty percent or more. Just take a couple of minutes to do that. It is intentional. Any thoughts, what could you do to make this much shorter?

UNKNOWN SPEAKER: Refer to this paper with the methodology (inaudible)...

PAUL Z. SIEGEL: Ok, excellent. Now notice that there are two references in there. So is there any other text in there that could be applied to those two references. Well, I think probably yes. It says, "States may chose sampling methods, such as random or stratified sample design. The interviewing instrument consists of three parts, etcetera.

There are several pieces of information that probably are mentioned in one or both of these publications, so some of this text could be completely removed. Ok, now how about over here where we get into the analysis variable and the covariates, any thoughts there?

UNKNOWN SPEAKER: Is it possible... (Inaudible)...

PAUL Z. SIEGEL: Actually, you are raising a point that I'm not sure if the text is put into a table, does it count in the word count. I don't know the answer to that. If you could possibly put it into a table, the font size would be smaller, possibility, possibility. Any other suggestions? I don't know if you have notices that this is the questionnaire verbatim, this is word for word content of the questionnaire. So if you developed a new questionnaire that has never been used before well then you really have a responsibility to do that, but if it is not a new questionnaire, if you are using questions that have been used previously, and that is obviously the case here, you don't need to include the questionnaire word for word. You can provide a reference where the reader can go and get the questionnaire. Now this was done years ago, before the internet. So at that time there could have been a reference, now it would just be a URL, where the questionnaire is available online. So you could describe and summarize the content of the questions, give the highlights of the questions and then refer the reader to a reference.

People sometimes ask about an appendix, putting it into an appendix, that is a possibility, but an appendix, if it is part of the paper would probably have the same amount of words as putting the questionnaire into the body of the paper, but now with the age of the internet, there is a phenomenon that has developed very robustly over the last ten and especially over the last five years, where more and more journals are publishing longer versions of articles online than what appears in print. So this is a good example of something that might be included, all of this detail might be included in the online version, but might not be included in the print version. Those are some of the options for managing content that is long. The key critical principle is anything that is not original, it is ok for you to mention it briefly and then give the appropriate reference where the reader can find the full content. So I just highlighted one paragraph with three hundred words and it is nothing...it is just the questions from the questionnaire.

Now there is a story that goes with that because this is an example of a real live paper that I worked on. I just showed you a version that has sixteen hundred words in the methods section only. So you could imagine that the paper was at least three thousand, four thousand words long. So a question comes up, "Well how can you make it short, how short can it be, process for editing it." In this case, this paper like I said, I was working on it; I was a coauthor on this paper. The coworker who was the lead author, I remember he called or stopped by my office and he said, "Paul, I'm going on vacation next week, this was like Thursday afternoon, I'm going on vacation next week, and I would like to send our paper to BMJ and I just looked at the instructions for authors and they have a two thousand word limit." So here we are on Thursday with this large paper

that he would like to get sent out before going on vacation. That is a good idea to send things off before you go on vacation. We had this big job in front of us and the reason why we had a big job was because he waited until sending the paper to read the instructions for authors. Again, this is just another example, to encourage you to read the instructions for authors early in the process. No we could ask the question, "Well, how short could you possibly make this?" What you have on page thirty six is the published version of this paper, and I emphasize that it is on page thirty six. So this is one of those examples where we sent out a paper, the journal was interested and said, "Oh, could you make it a brief?" So it just a real life example about what I was talking about his morning that you could...however many words are in the manuscript you write there is always a chance that when you send it out, the journal will come back and say, "Yeah, we are interested, but can you make it shorter?" So I would like everybody to be as prepared as possible to be able to make it shorter, but also psychologically that when you get that letter that your reaction is, "Sure, I can make that shorter." That you don't feel like it is going to be a burden to do that.

So now moving along, page thirty seven, page thirty eight is an excerpted article. This one I'm not even going to ask you to read, I really just want you to focus on the figure on the bottom of page thirty eight, which is up here. This is an example of a flow diagram that illustrates the number count starting from the number of people who were eligible to be in the study all the way down to the number of subjects who were included in the data analysis. This is a very useful tool. Journals more and more are using this format, using flow diagrams and including flow diagrams in papers. It is also good for you as a

data management tool because I know that I have had times that I do an analysis and I think I have one thousand and one hundred subjects and then weeks later I do some run and suddenly there is one thousand, one hundred, and four subjects. Well were did these four subjects come from? I thought that there was one thousand, one hundred. Well, if you do a diagram like this and you account for every subject that is in your database, then you won't have an unpleasant surprise like that. You will have everything, everybody, every single subject will be coded appropriately, and then you can go ahead with your analysis and not have any disruptive or unpleasant surprises like oh, I thought I had this many. So flow diagrams are a really practical tool.

Then on page thirty nine is an excerpt from a paper published in JAMA in 2001 which really talks about a systematic approach to developing flow diagrams, and there are these four elements to the flow diagram. There is the enrollment. The second is allocation. The second step, allocation, applies only to studies where study subjects are divided into groups, compared with one another. So if you were just doing a cohort study, for example, there would be no allocation. You would just be following one group of study subjects. Every study has the enrollment component and I think that every study also has the follow up component and the analysis component. So you can use these flow diagrams to manage your own data and they are also very favorably viewed by the journals. So just to look at it in a more detail again, we are on page thirty nine; there is the enrollment portion...