**KTDRR and Campbell Collaboration Research Evidence Training:**

**Basic Steps and Procedures for a Campbell Systematic Review**

*Presenter:*

*Terri Pigott, PhD*

A webcast aired on March 20, 2019, sponsored by AIR’s Center on

Knowledge Translation for Disability and Rehabilitation Research (KTDRR)

<https://ktdrr.org/training/webcasts/webcast66/index.html>

Transcript for audio/video file on YouTube:

<https://youtu.be/WQjzAgxEPz4>

ARIANA HAMMERSMITH: Hello and welcome to today's webcast. Can everybody hear me all right? This webcast is brought to you by the Center on Knowledge Translation for Disability & Rehabilitation Research or KTDRR at the American Institutes for Research, otherwise known as AIR, in coordination with the Campbell Collaboration.

The Center on KTDRR is funded by the National Institute on Disability, Independent Living and Rehabilitation Research, known as NIDILRR, in the US Department of Health and Human Services, Administration for Independent Community Living. The Campbell Collaboration is an international organization that promotes positive change through the production and use of systematic reviews and other research syntheses for evidence-based policy and practice. KTDRR partners with the Campbell Disability Coordinating Group or DCG, to help increase the number of Campbell reviews in the disability field.

My name is Arianna Hammersmith, I am from the DC office at AIR and I will be the moderator today. I would like to thank my colleagues, Joann Starks and Shoshana Rabinovsky, for helping with today's logistics.

KTDRR and the Campbell Collaboration are working together to offer a five-part training course that focuses on high-quality methods for synthesis of evidence including the procedures and methods for conducting systematic reviews and research syntheses, as well as software, tools and strategies for analyzing and reporting data. Last month, we kicked off the series with an overview of systematic reviews and research syntheses. Dr. Pigott explained the benefits of evidence from systematic reviews, identified different types of reviews and discussed the differences between systematic reviews and research syntheses.

In this second session, Basic Steps and Procedures for a Campbell Systematic Review, Dr. Pigott describes the steps in conducting reviews that follow the standards of the Campbell Collaboration, including the literature search, finding the studies, coding the studies and a look at some options for data analysis and reporting.

And now I’d like to introduce today's speaker. Dr. Terri Pigott is Associate Provost of Research and a Professor of Research Methodology at Loyola University Chicago. She is the former Dean of the School of Education. She is currently the Senior Methods Editor at the Campbell Collaboration Editorial Board and the co‑editor of the Campbell Collaboration Methods Group. Dr. Pigott received the *Frederick Mosteller Award for Distinctive Contributions to Systematic Reviewing* from the Campbell Collaboration in 2016 and she is an author on six completed Campbell reviews. She has served on a number of editorial boards and is Founding Chair of the American Educational Research Association Special Interests Group on Systematic Review and Meta-Analysis and is a member of the Society for Research Synthesis Methodology. And finally, I will now hand things over to Dr. Pigott, so we can get things started.

TERRI PIGOTT: Thank you very much. Good afternoon, everyone. Today I am going to provide you with an overview of the stages of a systemic review and meta-analysis, and an overview of the standards for producing a Campbell systematic review using the MECCIR standards. The Campbell Collaboration MECCIR Standards, or the Methodological Expectations of Campbell Collaboration Intervention Reviews, provides authors and users of Campbell reviews of intervention effects with clear and transparent expectations of review conduct and reporting. These are based on the original work of the Cochrane MECIR project team, which I have the reference there, Higgins et al. 2015. This presentation will provide highlights from these standards as I offer the conduct of a systematic review as I go through the stages today. There is a link on the slides to directly access Campbell's MECCIR standards.

The slides are a little out of order, so I will start here and talk about systematic review itself. Systematic reviews are just like primary studies: A form of research and they follow the same basic steps as any research study, but with primary studies as the unit of analysis. So, for example in a primary study, we all know what the basic steps are, but the same kinds of steps used in a systematic review where we start with developing our problem, collecting the data, evaluating that data and then doing data analysis and interpretation. Systematic reviews, though, use primary studies as the unit of analysis rather than individuals.

Our first stage of our systematic review is problem formulation. I am going back a couple of slides to this. So, in the problem formulation part of my presentation we’re going to talk about three questions: What are the kinds of questions that can be addressed in a systematic review?; sort of, what is the scope of those systematic review questions?; and then how does the review question that you develop shape the inclusion and exclusion criteria for studies included in the review?

Let's start with one example of a systemic review question, the most common one I would say, and that is to use systematic reviews to examine the effectiveness of an intervention. I have an example here from a recently published review in *Review of Educational Research*. And this study is a systematic review of the effectiveness of academic interventions of elementary and middle school students who have low socio-economic status. This kind of review question about the effectiveness of interventions takes the form of: what are the effects of some Intervention X on some set of outcomes Y for some population? Or you might see variations on that theme. I am looking at differences in effects of interventions, two different interventions, but in our example, I just gave you, here again for Dietrichson, the research question is about the effects of academic interventions for low-SES students.

Now we also might have systematic review questions that are really interested in how two different groups compare. So for example, they might take the form of how does one Group X differ from some Group Y on some characteristic. One example of this kind of research question was a systematic review published in *Psychological Bulletin* looking at gender differences on a range of outcomes such as leadership, math performance, hours worked in the home. We might also have a systematic review question that looks at the differences between men and women or boys and girls. Another example of that would be the second example here on the slide which is another recently published systematic review on, actually, differences between first and second‑generation immigrant students on academic achievement. Here's another example of using a systematic review question to actually look at two different groups.

You might also have a systematic review question that is actually interested in the associations among constructs. This is a systematic review that looks at the correlation between having been victimized, having been exposed to cyber victimization and your academic outcomes in ages of students 12‑17. For example, we might have a systematic review question that might ask about how does some construct X1 relate to some construct X2 for some set of population, or some variations on that theme. The Gardella et al. systematic review is really trying to estimate the strength and magnitude of the correlation between peer cyber victimization and educational outcomes. This is another example of a systematic review question.

So in summary: There are a number of different systematic review questions you might you pose. One might be, maybe you are interested in the effectiveness of an intervention, so does some Program A have larger effects than a control condition or treatment as usual? We might have a question about the comparison between two different groups. Do men and women differ on some outcome, Group A versus Group B? We might have a question about associations between two constructs, looking at the strength and direction of the correlation between two constructs.

We also might have a whole other range of questions that I am not going to cover here. There are research synthesis and meta-analysis methods for diagnostic and prognostic tests. In a set of studies can we look at what test is more accurate at diagnosing some condition or what test is better at predicting some outcome or also about prevalence: What is the prevalence of some phenomenon that’s estimated across a set of studies?

Systematic review questions can also vary in their scope. We might have very specific, narrow questions that are useful for testing the effects of specific treatments or we might have broad, global questions useful for generating new knowledge. For example, identifying common elements of effective programs or building better intervention theories to guide program development and evaluation design.

I am going to stop here for a second because I see a question in the chat box that might be relevant. The question is asking if there is a difference in research questions between a systematic review and a scoping review. I would say there really isn't any difference. It sort of again, depends on what your aim is. A scoping review could be used to look at what the literature is, what literature exists for a very narrow question; or could be used to exist for a very broad question. So, I would say there isn't really any difference between those kinds of reviews in research questions.

Going back, in terms of the scope of a systematic review question, for example, an example of a very narrow question in a systematic review might be in the case of an intervention. We might be interested in just understanding the effectiveness of one particular kind of intervention such as for example in this example, second step which is an intervention for developing socio-emotional skills in elementary schoolchildren. So, we are just interested in one particular kind of intervention.

But we might also do a systematic review is that's much broader in its scope, where we might ask the question, What are the potential impacts of all interventions that are focused on developing socio-emotional skills in elementary schoolchildren? That would be a very broad kind of question.

The example we first started with, on the effectiveness of academic inventions for children of low socio-economic status would be considered a very broad review. The reason that is important is that the scope of the question will also guide many aspects of the review. And in particular, the systematic review question guides decisions in the review about what kinds of studies, what are the critical characteristics of studies, that will be included in the review. These would be things about what are the participants or patients or clients that we are most interested in. If we are interested in the effectiveness of an intervention, what are the types of interventions that we will be interested in? What are the study designs that are appropriate for the systematic review question? What are the outcomes that are appropriate?

So problem formulation, again, guides the study inclusion criteria and we usually use this acronym called ‘PICOS’ to help us use the systematic review question to identify the kinds of studies that are going to be eligible for a systematic review. In the PICOS framework, the P stands for population or participants or it could also stand for problems or conditions, if we are interested in a different kind of review. The I stands for interventions, if we are interested in the effectiveness of a particular intervention. The C stands for comparison group. That is applicable if, again, we are interested in the effectiveness of an intervention and we are looking at studies that compare patients or participants that were exposed to an intervention versus a control group. The O stands for outcomes, what are the primary and secondary outcomes or acceptable outcome measures relevant for our research question and the S stands for study design. What are the kinds of study designs that are going to be most appropriate for our question or in some cases you might also see the S stands for settings: What are the settings of the studies that we will be interested in? For example, if we think back to our first example on academic interventions for low‑income students, the settings were schools.

Again, our first example, the Dietrichson systematic review, the P is low‑income students in elementary and middle school; I are interventions explicitly aimed to improve educational achievement that could be implemented by schools. The C was a typical nontreatment control condition. The O, standardized achievement or academic tests. S were randomized controlled trials or controlled quasi-experimental study. This PICOS comes directly from the research question in that systematic review.

In summary, for problem formulation, as is true in any primary research study, the systematic review research question guides all other decisions made about the methods in the study. I have given you a range of kinds of systematic review questions that would be appropriate, things about effectiveness of interventions or differences among groups or associations among constructs. Systematic review questions can vary in their scope. They can be narrow questions about a particular intervention and the effectiveness of that intervention, or very broad questions designed to understand the literature on a topic. These broad questions are if you are familiar with the Institute of Education Sciences Goal 1, the goal framework for their funding opportunities, this would be, broader questions tend to be eligible for the IES Goal 1 funding. Obviously, systematic reviews are most productive when there is a large body of literature, although both the Campbell Collaboration and the Cochrane Collaboration have published reviews that have no studies in them. Sometimes you do a review and find out none of the studies conducted are eligible, given your criteria. But, again, the research question leads to those PICOS, the inclusion criteria for studies eligible for the systematic review.

Let's move on to our second stage, which is data collection. So, in a systematic review, data collection means researching the literature to collect all of the primary studies that we can that examine the research question and that meet the inclusion criteria. Primary studies is the unit for the systematic review. In a primary study we want to get some sample of, in this case, studies that will generalize to some hypothetical population of studies on a given research question. We want to be able to make the case that the studies in our systematic review are representative in some way of our; of all the population of studies that have been conducted.

We want to use a search strategy and also document that search strategy, so we can support the argument that we have found a representative sample of studies. The next couple of slides are based on the source that is the bottom of the slide here, Kugley et al. 2017, this is the *Campbell Information Guide* *for Information Retrieval*. I will be relying on that source for the next few slides.

Some things to consider when you are developing your search strategy to identify a set of studies that are potentially eligible for your systematic review. You need both sensitive, which are very broad, and specific, which are very focused searches. If you ever use Google Scholar you will note that, if you use it, for example, as a search strategy, those are very sensitive and very broad searches. You get a lot of hits; but they are not very specific. Databases like ERIC or PsychINFO are more specific. You would need both to identify a set of studies for your review.

You should think about the use of keyword searches, of multiple database; and it's important to know those keywords; to have knowledge of those keywords used in the database. Every database is usually tied to a discipline and each discipline uses different key words to represent the same phenomenon. It is also important to search multiple sources of literature, unpublished websites, dissertations, special registers, studies, and so on.

The reason that is important is because we know it is documented in the scientific literature that publication bias exists; so that published literature tends to have larger effect sizes. So, we do need to make sure that we have a representation of the gray literature as well, so that we have an unbiased estimate of whatever that effect is that we are interested in. And also, people use hand‑searching of selected journals to find relevant studies.

In thinking about the databases to search it is also important to remember--and this is true in many fields, but particularly in disability research--that many researchers across many disciplines are interested in the issues that arise in disability research. You will need to think through -- there are researchers in education, medicine, psychology and so forth that are interested in these questions. You will need to search multiple databases to make sure you have found all said on the topics. *Campbell's Information Retrieval Guide* Appendix A contains literature to consider.

The search strategy in general tends to have three different sets of terms. So, if you are interested in an intervention, if your research question, systematic review is about the effectiveness of an intervention, the three sets of terms tends to be something like: The condition of interest for the population you are interested in; the interventions or components of the interventions that are important; and the outcomes. And, again, each database may use different terms for the same phenomenon. It is important to understand the concepts in each discipline. The formulation of the search terms will include controlled vocabulary, which is the vocabulary that is specific to each of the databases.

The use of key words, Boolean operators and the use of limiters. Again, Campbell's *Information Retrieval Guide* gives you some examples of this. My next slide, which would be hard for you to see, is an example of one piece of a search strategy that we used in a review that is available on the Campbell Collaboration website on teen dating violence prevention programs for high school students. You can see in a little piece of that search string that it uses “prevention” or “intervention” as one search term; then the actual kind of intervention that we are looking at: sexual violence, sexual coercion, peer support; then the type of study design.

So, this is just one example of that search strategy. If you look up the reference for this review at the end of the slide deck, you can look up the completed review to see the examples of the search strategy. Always, my go-to is, I am not a specialist in search strategy, and I think the best advice I can give you about searching literature is find a librarian or search strategist, a search expert who can help you navigate the multiple databases that you will need to use to do a systematic review in disability research.

Another issue to consider is that it's very important to document your process from start to finish. One of the MECCIR standards is to be able to provide documentation of your search strategy so it can be replicated. It is important to document all of the information sources searched; the dates covered in that search and the date of the search and your full search strategy, including the limits and keywords for each of the databases you search; for the replication of that search.

It is also important to remember that searching is an iterative process. While you are trying to define your key concepts and discovering appropriate search terms you may go back and forth. I see there is a question here as you begin your search you may come across better search terms. (Question) Can you revise your search terms? Yes. In fact, you will always revise your search terms. That is what happens in the search process. It is an iterative process. It is difficult to know ahead of time what search terms are going to be most appropriate in each of the databases.

Another really important thing to think about is to use some kind of software to manage the whole search process. So, there are links here to Endnote and RefWorks. These are two programs that you have to pay for. But for some of you attached to an educational institution, you may have access to those particular reference managers. There is a link there also to Zotero, a free reference database manager which I have some familiarity.

So, how these reference managers work: There is a screenshot; so, for example if I happen to be searching and I see an article that I think is important, you can see here I have circled in the top right‑hand corner there is a little icon that looks like a book. If I click on that, when I have Zotero open, that reference will be uploaded right into my Reference Manager. I don't have to do anything else and it will go into a library I have set up; and it can also download the pdf or a link to the pdf that I can go back to.

Another way that Zotero works is that, while you are in your search, if you save a bunch of items after you have say searched ERIC you can put the results of your search into a folder and download the items in that folder usually using some kind of link in your browser that directly exports them into as you can see a bunch of different citation managers including EndNote and Zotero. The use of these reference managers are really important and help you organize the number of studies that are identified in your searches.

To summarize, I can't stress the importance enough, the importance of the information retrieval search process. This is your, this is the stage at which you are trying to gather all of the studies or some representative sample of studies on your topic. It is not a one‑shot deal. It definitely can, it can require several iterations of searches. You need to make sure you have a way of documenting those. It requires some expertise and planning of some of the searches and that is why it is so important to have a library or search specialist. It is also important to use some kind of management software to store, manage and organize those results. And then also to document that search so that others can replicate it.

Let's go on to our third stage which is data evaluation. So, once we have identified all of the studies that might be important, then the next steps are to identify the studies that meet the criteria and to code information to use in the systematic review and meta-analysis. Joan asked, “Is there a minimum number of studies in order to have reliable data?” No. This is the part of systematic reviews that I would say is an art.

But I think what you are trying to do, what I always tell people: We are trying to make an argument that you have done your due diligence in trying to find all of the studies on a given topic and that argument will be based on…the argument is stronger if you can say I have searched a range of databases, I have used this range of search terms, and these are the number of studies I identified. So, you think about how do I build that argument of, I have done a good job of trying to identify all of the studies that might be relevant.

So, again, data evaluation, there are two stages to that. One is screening to identify the studies that meet the inclusion criteria and the next is to code the information so that we can use it in the systematic review and meta-analysis. We are going to start with screening. So, screening of relevant studies usually takes place in sort of two waves. The initial screening is usually based on reading titles and abstracts. We use this sort of screening, the initial screening to exclude obviously irrelevant articles such as opinion pieces or non-empirical studies.

Even though titles and abstracts are sometimes unreliable and bear no relationship to the contents of the study, the first screening really does tend to exclude those studies that really don't include any relevant research data.

The best practice, which has been demonstrated in empirical studies, is to really use double‑coding by two trained raters working independently. Because, again, this is the stage at which we are trying to be accurate at getting the eligible studies. You want to, again, make the case that you have done everything you can to include all of the studies that are eligible for your systematic review.

And then there are also machine learning strategies that we can use for screening that can help the screening process. I am going to show you a screen shot of one of those databases, which is Abstrackr, developed by the Center for Evidence Synthesis in Health at Brown University. This is free software available that is designed to screen for systematic reviews. What it does is you upload your list of studies, the titles and abstracts for studies into Abstrackr and code the studies. As you code the, sorry, start to screen the studies. That screening is just in or out. Is the study eligible or not? The machine learning algorithm in the background resorts your study by the probability that your study will be eligible for the systematic review. So, it begins to learn as you are coding which studies are most likely to be eligible for your review. This can really help in the screening process. It also helps you organize multiple screeners. There are ways to set up an abstract or project that will help you use multiple screeners to look at the same data and produces reports on the agreement between the two screeners. This is a useful and free tool that can help with the abstracting, with the screening process.

And the reason you need this is in the next slide. As you can see, our first example Dietrichson et al., identified initially almost 12,000 studies in the search. That is an enormous number of studies that have to be looked at by two different people to decide if they are in or out. You can see, I have given you a range of initial numbers of studies that are typically identified in a systematic review.

So, a tool such as Abstrackr is important. I want to point you back to the chat: Joann told us April 17 there will be another session in the series that will actually talk about this kind of software that can really help you manage your systematic review. On this slide I want you to note the number after the first screening. The Dietrichson screened 11,807 to get to the 1,000, past the initial set. There were still quite a few that had to be looked at.

After initially screen the titles and abstracts you move on to screening full texts of potentially eligible study. This where you are including or excluding criteria based on the PICOS. You will use the full text of the potentially eligible studies to look at those elements of PICOS to make sure that each study, and using the full text of the studies, can see how they conform to your PICOS.

Best practice I'd say is blind here. What I really mean is independent, double‑coding from two different coders. It is really important, again, the recruitment and training of the coders as you are trying to keep the process of identifying eligible studies as reliable as possible.

This next slide shows you a PRISMA flowchart which is the way we document the search process of a systematic review. So, the top box shows you the number of studies that were identified in the search and this PRISMA flowchart is the flowchart for the teen dating violence prevention systematic review that I talked about earlier. So, you can see that we started with 1,600 - you can't see but we started with 1,600 potentially relevant studies. 1,500 of those were excluded in that abstract and title screening because they weren't empirical studies. 90 were left for a full text evaluation. After that, 22 more were excluded because they didn't meet the PICOS. 58 then were identified, which I am about to talk about in the coding review.

Once you have that complete set of eligible studies, your next step is to code aspects of the study contents and to get the effect sizes; so normally you would you develop a coding manual that includes codes for recording things like the setting of the study authors, publication type, the methods used in the study and the quality of the methods, which I will talk about in a moment. If you are interested in an intervention, you would be coding things like components of the intervention and what was happening in the comparison or control group, coding things about the participants or sample, outcomes, then the effect size.

So here is an example from the coding manual on the teen dating violence review. We can see, for example, where they are looking at sample and program characteristics. We are asking the coder to say what was the general SES of the students in the sample, what was the dependent measure, outcome measure, teen dating violence knowledge, what was the attitude. How was the outcome measure reported?

There is a question here about managing the coding and I will get to that in just a second. The reason why we code is to develop and provide descriptive details of the studies included in the systematic review. When we are reporting out the systematic review we can understand sort of the landscape of the evidence base and then it helps us to identify gaps or issues in the evidence base; and also, to record information about study effect sizes so we can later examine and explain the differences we might find in effect sizes across‑studies.

An important thing to think about is the coding of study quality. There are a myriad of ways to actually code study methods that are used in study quality, the quality of the study methods used. For example, there is the Cochrane risk of bias framework, the GRADE system. There are many, many method quality checklists; direct coding of methodological characteristics. For example, you might code whether or not you are interested in intervention, for example. In experimental studies you might code whether or not there is evidence that the treatment group and the control group were balanced and looked similar to each other at the beginning of the study. And we use these sorts of ratings and study coding quality in a meta-analysis to examine if the results differ across studies in study quality.

I am going to give you an example of the Cochrane Risk‑of Bias Framework, which is most relevant to randomized control trials. The items here are for each study, you rate the quality of these items for each study, and you can see that these questions here are really focused on the quality of the randomization, the quality of the study itself. For example: Attrition - Was there differential attrition between the control and treatment groups? and so forth. So this is one example of a way to code study quality for each study.

There are also coding strategies for non-randomized studies. The Newcastle-Ottawa Scale focuses on the quality assessment for case‑control and cohort studies and I have given you a link for those studies. Lots of different methods are used to code study quality and it really is up to you to think about how you might want to do that, depending on the kinds of studies that are eligible for your systematic review.

Some other issues to consider. Reviewers definitely develop these coding manuals for reviews. And I have to go on record here saying that coding is the most time-intensive portion of a systematic review, and the best practice is to use software to conduct and organize coding. Some of my colleagues are very good at database software and are able to develop Access or FileMaker coding strategies that are very elegant. Many people use Excel or Google Sheets if you are sharing across to actually code these kinds of things in each study.

Another important thing to consider when you are planning a systematic review: How are you going to manage the coding part; what kind of software are you going to use? Also, in this stage it is important to attend to the training of coders, use double-coding when you can, and to record the reliability of your coding decisions. And continually check on that as you are coding through studies.

Common challenges in this stage. It is impossible to underestimate the time needed to prepare and code studies. This is by far the most important and the time-intensive part. The research strategy, coding protocol; don't assume it is not an iterative process. Developing coding protocol is definitely an iterative process. There is always something that you have missed once you start reading these studies.

It is important to attend to coding drift. All through the process, check that all of your coders are interpreting the items in the same way. And to just organize and supervise that process by making sure that your coders continue to be motivated to do it, and using appropriate software.

We have a question in the chat box: Is there a way to register your systematic analysis so someone else isn’t doing the same thing? There are a couple of registries for protocols. Obviously, the Campbell Collaboration helps you to register and review your protocols so there is one place to look to make sure that someone else is not also doing the same kind of work. It is escaping me right now. Let me get back to you, Lynn, about the other registries. There is *Systematic Review*, it is an open access journal that also registers protocols, and that is a place also to look. Thank you, Joan. PROSPERO is another registry of systematic review to see if other people have registered a systematic review. I touched on this in my first webinar.

OK, so now we have talked about problem formulation, how we search for studies, how we screen and code studies; now we will be talking about data analysis and interpretation. I will try to get through these fast. Once you’ve coded your eligible studies, the analysis consists of two stages: One is to describe the studies you’ve included, usually in narrative and table forms. Good systematic reviews give your reader some sense of what the landscape of the literature looks like by saying, I found this many studies, these are the kinds of people in them, and so forth.

Then, when possible you want to do a meta-analysis, especially if you are interested in estimating, say, the treatment effect of some intervention or the magnitude of the correlation between two constructs. And as in any statistical analysis, that analysis includes examining the average effect size and its confidence interval, and also exploring why, how variable, how heterogeneous, those results are across studies. Then if it's not possible to do a meta-analysis because you only have one or two effect sizes, you would think about doing a narrative discussion of study results.

So, the meta-analysis is focused on effect sizes. Effect sizes is our unit of analysis in the meta-analysis. So, for every study, we express the results of each study using a quantitative index of effect size. We can think of effect sizes as a measure of the strength or magnitude of a relationship of interest. That may be the difference between the treatment and control group; so that treatment effect. It might be a correlation. Effect sizes have the advantage of being comparable. They estimate the same thing across all studies and can be summarized in a meta-analysis. We use effect sizes because not every study uses the same measure of a particular phenomenon or outcome. For example, math ability can be measured by a myriad number of different measures; and so, we use an effect size so that we can say something about math ability across all of these studies, even if everyone is not using the same measure. Please note that session 5 of this series will talk more specifically about effect sizes and meta-analysis.

Here's an example of what's called a Forest Plot. Each of the rows of this chart is a different study; and you can see that what is in the middle there are a representation of the effect size. This is the difference between treatment and control groups on their knowledge of teen dating violence after the treatment group has been exposed to an intervention. The dotted line in the middle is the zero line, no difference between treatment and control; then you can see that each study has an estimate of the effect size, the square in the middle and the confidence around it which are the whiskers. The bottom shows the average effect across the studies. So this is an example to show you how we might represent a meta-analysis. It gives you a sense of the effect across the studies and also how variable they are.

As in any statistical analysis--I want to put in a plug for this--the mean and its associated standard error may not be the best representation for distribution of data. Because in education and disability research, we expect that study results will not be exactly the same across every study. So an important part of any meta-analysis is also examining the amount of heterogeneity across‑studies and across the effect sizes in the studies.

In meta-analysis with a sufficient number of studies we can explore potential correlates of heterogeneity. If we only have one categorical predictor, say we want to compare randomized controlled trials to quasi-experiments, we can use a simple one-way ANOVA model. With multiple predictors we can use meta-regression to examine of there are predictors of the variation of effect size across studies. A common predictor of heterogeneity is study quality. We use study quality as an important predictor for examining the variation across studies.

What I have here is an example of sort of an analysis of why studies might vary, based on studies that used random versus nonrandom assignment. This is again the teen dating violence example. If we just look at the first row, the effect sizes for teen dating violence knowledge, studies that used random assignment and those that did not. The studies that used random assignment have a larger effect size 0.36 than studies that did not use random assignment. This is an example of the exploration of heterogeneity using study quality. Again, you can find this source in the Campbell Library, free access, and you can see an example of this.

Some computational challenges that might exist when you are using meta-analysis, when you are trying to compute effect sizes and so forth. Every researcher I know that has used systematic review runs into the problem of missing data in primary studies for computing effect sizes, particularly in older studies. Many of the older studies don’t give you the information you need to calculate an effect size, so that you will run into that problem.

It is also a problem sometimes getting appropriate effect sizes for complex designs in primary studies. There is much research going on about that: How do we get an effect size for a clustered randomized trial?; How do you get an effect size if they only give you results of their regression model? You might run into issues of combining effect sizes from different metrics. What if some studies give you a correlation and other studies give you means and standard deviations?

Another common problem is multiple effect sizes presented within studies. I have never seen a primary study that only gave you one outcome, there are usually multiple outcomes; this leads to a problem with dependent effect sizes. This is a common challenge as well in meta-analysis. You may have to consult with a meta-analysis specialist when you run into some of these issues.

In summary, data analysis and interpretation. I always tell people, this might be the stage that people feel most anxious about but it is certainly not the most time-intensive part of the systematic review. Most of the time, again, is in coding. But exploration, if you are going to use meta-analysis, the exploration of heterogeneity is of greatest importance because we expect variation across effects in studies. So, we need to carefully plan for moderator analysis, for the exploration of heterogeneity both in the coding stage and in the analysis. When coding for studies you should be thinking about the reasons studies may vary in their effects and make sure you code each of your studies for those issues.

It is also important to be careful when we interpret these kinds of analyses, because we are talking about relationships among characteristics of studies rather than relationships among people. Because there is so much development of statistical methods in this area, you may want to make sure you consult with someone who has done a meta-analysis before.

Finally, we might need to use narrative synthesis when there are few or no studies identified for the review, or if in fact, the review question we are interested in, is really more interested in a conceptual understanding of the literature rather than aggregation or estimating the effect size. There are lots of methods for narrative and qualitative synthesis that also focus on transparency and replicability, and Campbell is beginning to develop standards for these types of reviews. We have come to the end of my talk. Here are some of the references that I have gone through and I just want to thank you for your time.

ARIANA HAMMERSMITH: Thank you, Dr. Pigott, that was a wonderful presentation. Let's see if the audience has some questions for you.

TERRI PIGOTT: Could you talk a little about developing a protocol or plan?

So, I would start at the Campbell Collaboration website. But a protocol for a systematic review is basically just laying out everything that you are going to, everything you anticipate you are going to do. So, going through the MECCIR standards will help you develop a protocol; but you want to, in a protocol document, talk about your research question, give a rationale for why that research question is important, connect with prior reviews that might have been done on that topic, and then go into the methods you are going to use.

Basically, a protocol will say we are going to search these kinds of databases, we are going to possibly use these kinds of search terms; then it goes into details of the meta-analysis. The Campbell Collaboration is a really good resource for looking at examples for protocols, as well as the journal *Systematic Review*.

ARIANA HAMMERSMITH: Wonderful. Any other questions for Dr. Pigott from the audience? So, I am actually curious, how, since our Center is focused on knowledge translation, how authors of systematic reviews can make reviews more accessible to a wider audience; and if you could speak to experience writing plain language or lay summaries?

TERRI PIGOTT: I don't normally do that. You know, my adviser, Larry Hedges, often says that if you want to understand how people understand systematic reviews, you probably don't want to ask the statistician. But I would say that it is really important to me that people use systematic reviews, and when you read a systematic review, part of the problem with the knowledge translation part is because they are trying to give you all the results you need; all of the information you need to replicate that review. But I would say that it's really important, when you are finished with a systematic review, to really think about what are the key messages we want to give to a policymaker; how do we want to express both the average effect that we found and our uncertainty around it. Again, the Campbell Collaboration has a number of examples of plain language summaries that do that, I think, really well.

ARIANA HAMMERSMITH: I see. Thank you very much. It looks like we have one more question in the chat box if you have time?

TERRI PIGOTT: Yes, sure. The question is: In a broad scoping review, a search will bring an enormous number of articles. Is it ever appropriate to use title and abstract only in the initial search? If you are asking about screening, yes, you will always have an enormous number of articles identified, usually. in that initial search. So your first level of screening will just be looking at title and abstract, and that tends to help you eliminate all of those articles that probably don't have any empirical research in them. I hope that was the answer to your question.

ARIANA HAMMERSMITH: Okay, great. Firstly, I want to thank Dr. Pigott for taking the time to prepare and to give us this description of the steps in preparing a systematic review that follow the standards set by the Campbell Collaboration. I would like to thank everybody for participating this afternoon and we hope you will take a few minutes to give us some feedback about the webcast by filling out a brief evaluation. The link is listed in the slides and has been posted in the chat box. We will also be sending an email with the evaluation link to everyone who registered. And finally, I want to thank all of the AIR staff and representatives from the Campbell Collaboration who helped with planning and logistics and of course, NIDLRR for their support for these webcasts and other events.

We want to invite you to attend the third webcast in this series, Management and Analysis Tools for Reviews, that will take place in one month on Wednesday, April 17 at the same time. And we will have several presenters talking about a variety of software tools to help manage a review. As a courtesy, everyone who registered for a previous session will be registered for the other sessions in the series. We invite everyone to visit our website, ktdrr.org. I can post that link in the Chat box in a moment. Thank you and have a nice afternoon.