Transcript of Launch Webinar for the State Health Practice Database for Research (SHDPR)
November 18, 2016

Daniel Weinberg:	Good afternoon and thank you for joining the SHPDR Launch Webinar. My name is Daniel Weinberg and I am with IMPAQ International. I serve as the principal investigator for the SHPDR development project.
[00:00:30]	Before we begin, there are a few procedural items to discuss. For everyone's benefit, please mute your line. Also, please enter your audio PIN which is available as you sign on and under the audio tab throughout the presentation. You can enter the PIN at any time. Please submit questions via the chat function. We will answer questions at the end of the presentation. Also, this presentation will be recorded.
	The first item on the agenda is a message from NIH from Dr. Sarah Duffy.
Sarah Duffy: [00:01:00]	Hello, on behalf of the National Institutes of Health Common Fund Health Economics Program, I would like to welcome you to this webinar on the State Health Practices Database for Research. I'm Sarah Duffy from the National Institute on Drug Abuse, and I'm one of the co-coordinators along with John Haaga of the National Institute on Aging, of the program which is chaired by Dr. Richard Hodes, Director of the National Institute on Aging.
[00:01:30]	The Common Fund Health Economics Program is one of a number of Common Fund programs overseen by the talented staff at the Office of Strategic Coordination. One of the six offices of the Division of Program Coordination, Planning, Strategic Initiatives, in the office of the NIH director. The NIH Common Fund was enacted into law by Congress. The 2006 NIH Reform Act to support cross-cutting, trans-NIH programs that require participation by at least two NIH Institutes or Centers, or would otherwise benefit from strategic planning and coordination.
[00:02:00]	Among the goals of the effort are to provide a strategic and nimble approach to addressing key roadblocks in biomedical research that impede basic scientific discovery and its translation into improved human health. The Health Economics Program, which is in its last year, these are short-term programs, aimed to support theoretical and applied research to understand how innovations in treatments, diagnoses, and prevention can be most effectively deployed to improve human health and well-being.
[00:02:30] [00:03:00]	As Daniel will discuss further, among the roadblocks identified early in the program was the lack of a database such as the State Health Practices Database for Research. The data infrastructure team consisting of Partha Bhattacharyya of National Institute on Aging, Greg Bloss of the National Institute on Alcohol Abuse and Alcoholism, Nancy Breen of the National Cancer Institute, and the wonderful Agnes Rupp of National Institute on Mental Health, developed the project. Agnes and David Chambers, now of NCI, and the staff from the contracting office developed the contracts, and the team led by IMPAQ International and overseen again by the wonderful Agnes as the contracting officer's representative, have created the database and developed this webinar.
	Again, I want to especially thank Agnes for her tireless devotion to this project, as well as

	our colleagues at DPCPSI, Leslie Derr and Rachel Britt, and also the folks at the Office of Acquisition for providing outstanding support for this project. We also convened several technical expert panels that helped with this work and I want to thank them, and then of course the staff at IMPAQ and its subcontractors for their fine work.
[00:03:30]	I will now turn it over to Daniel.
Daniel Weinberg: [00:04:00] [00:04:30]	Thank you Sarah. This portion of the presentation gives background on the SHPDR, its development and its intended use. The need for SHPDR was identified in 2010 as a result of a meeting of the NIH's Common Fund Health Economics Program. The group identified a gap in information on state level laws that help economists' need to conduct research. Often researchers gather state level data on their own which may be duplicative with other databases that already exist. SHPDR is a state level database intended to fill this data gap. SHPDR will reduce data collection burdens and help researchers to conduct clinical and technology related studies. SHPDR contains variables plus a lot of other legal content and researchers are encouraged to use all of the information, including the text of the statutes and regulations, to fully understand the legal context and to ensure that SHPDR variables serve the research goals of the study.
[00:05:00]	The next few slides describe at a high level some components of SHPDR development. SHPDR is an NIH funded project that is a collaboration among economists, attorneys, and technologists. The SHPDR team includes the prime contractor, IMPAQ International, plus George Washington University and Edaptive Systems.
[00:05:30]	An important step in the development of SHPDR was the Scientific Expert Panel in spring 2014. During this meeting we discussed priorities for the database in terms of content and the specificity or level of granularity that the database would contain. We also discussed database structure that would facilitate quantitative research. An important outcome of the meeting was the identification of the need for extensive documentation and researcher guidance.
[00:06:00]	Following the Scientific Expert Panel meeting and decisions on content for the database, the team continued planning for implementation. This involved data collection pilots, examination of other databases to identify best practices, engaging an implementation expert panel, and selecting Lexis-Nexis as the main data source. Lexis-Nexis contains statutes and regulations for each state and allows searching and extraction of that information in a uniform format.
[00:06:30]	This section describes how the SHPDR should be used. As described earlier, SHPDR is a state level data set. It currently has data for 2010 and 2012, and data collection for 2014 is under way. Based on the team's research and input from the Scientific Expert Panel, SHPDR contains data for topics expected to affect health outcomes and the use and diffusion of medical technologies. One important use of the database is the formulation of research hypotheses. Researchers can browse the database to identify areas in which states exhibit cross-sectional and longitudinal heterogeneity.
	Also, as discussed, the database can reduce research burden for researchers who need to

[00:07:00]	assemble state level information on statutes and regulations. Even if SHPDR does not contain the exact data element you need, it will reduce your search space because it can indicate which states have statutes and regulations on a specific topic.
[00:07:30]	We see the SHPDR also as a starting point for researchers. The database contains binary variables that can be used to develop analytic files to be imported into various statistical packages. In addition, as you will see during the rest of this presentation, researchers should review the data and accompanying legal text to ensure that the variables fit the research aims.
[00:08:00]	As discussed previously, SHPDR contains a lot of legal data, in addition to the coded variables. This includes short synopses of the legal text referred to as justifications. Full legal text is also available allowing researchers to further analyze state specific legal context. The user guide contains additional methodological information relevant for variable interpretation. We also note that the SHPDR is evolving in response to user feedback, and continued review by NIH and the project team.
	Next we will move into a demonstration of SHPDR's features. Slava?
Slava Katz: [00:08:30]	Okay, good afternoon everyone. My name is Slava Katz. I'm a vice president at IMPAQ and I have been providing the technical direction for the SHPDR project. Today I will provide an overview of the SHPDR website and talk about how to access it, how to navigate the website and some of the features and viewing and browsing and downloading the data set.
[00:09:00] [00:09:30]	You should be able to see my screen here. You can access shpdr.org through any modern browser. Here I'm using Edge. You can also use Chrome, Firefox. Shpdr.org is the address, and the home page here has the user's guide and a link to about SHPDR. On top there is a menu bar that shows the different navigation of the SHPDR website, and down below you see the footer and it has some information about the contract that we have with NIH and how this project came about.
[00:10:00] [00:10:30]	Clicking on the about SHPDR, a tab here, and Daniel talked about the history this page provides, but a little bit more detail about the project. I'll skip over this page now and go to the user's guide. The user guide is really the first place that we recommend for you to start learning about this data resource. The user guide is structured into six sections. The sixth one being a number of different appendices that we'll take a look at. Again, here we have an overview, provides a little bit more detail and background about the project. We here talk about the purpose and the structure of the user guide itself, which I will go into here in more detail.
[00:11:00] [00:11:30]	Section three is where we start talking about how to use SHPDR, where to go, how to access the data, how to download it. We'll see that presentation here in a minute. Section 3.2, SHPDR data elements, SHPDR elements here. This is really where we recommend you to review and to start. This list is what we have in the SHPDR data set. Here's a list of data elements that we have identified to be included in a SHPDR data set, so year of the legal research that was completed. In the database we now have 2010, 2012. We'll have 2014 soon, and then 2016 as well once that research is completed. State is another variable that

[00:12:00] [00:12:30]	is key in our data set. Then the variable itself, and the variable value, so those are two elements of our data set that are very important. The variable itself, it describes the regulation or the information in the data set and it's grouped, the variables themselves, we've developed a taxonomy to group those variables into topics, subtopics, and the lowest level is the variable itself.
[00:13:00] [00:13:30]	We'll actually take a look at the taxonomy here to give you a sense of what the variables are and what are some of the topics and subtopics. The variable value right now is a yes or no, and again it is a binary value that indicates the presence of yes or no for that particular variable. We also have justification or notes as part of our element, so this actually is a very useful piece of information in that it provides you the context for why the particular variable was coded yes or no. They're fairly succinct explanations of the rationale that our researchers used to code the variables one way or another.
[00:14:00]	We then talk about the statute, or include the statute or regulation. Whether or not that variable was coded from a statute or regulation, and we include the citation, the section title, the chapter title, and the actual title of that statute or regulation. We include information about the regulation or statute and where that variable was abstracted from.
[00:14:30] [00:15:00]	We talked about the taxonomy and it's relation to the variables, so the variable being the lowest level of the taxonomy. We have topics and subtopics as well and they're listed here. We also have another view of it that we'll see here in a minute. Another very useful feature of the website and the data set is that we actually have a link to the statute or regulation that, the text of it, the actual text of it, that the variable is coded from and so what that could be used for is that to further understand the rationale behind a particular value of a variable and for further research as well.
[00:15:30]	We talked about the taxonomy a little bit. The taxonomy's organized into two domains. The provider domain and the insurance domain. From there we have our topics and subtopics and so forth. The user guide explains and outlines the descriptions of those pieces of information. We also talk about some of the limitations of SHPDR. That is another useful resource there just to make sure that when you're doing a research you're going about it the right way and using the data set as it was meant to be.
[00:16:00] [00:16:30]	Looking at some of the appendices now. I will skip over appendix A. We'll show it here in a second. Appendix B here is a representation of the topics and the subtopics and the variable descriptions. This is where you actually can see how our taxonomy's organized and the definitions for each of the levels of the taxonomy. I should note that when you see the download button that means you can download whatever you're seeing on the screen, so in this case if I hit download here, I'm able to see and download on my local desktop the screen that we were just looking at in a PDF form.
[00:17:00]	Switching back here to the user's guide to walk through other appendices and there I was just using the breadcrumbs to navigate back to the user guide. You can also click on the user guide on the menu, but looking at the glossary of common legal and healthcare-related terms, a useful resource here. Many of these terms appear in our justifications or appear in our variables perhaps, and so having a common definition of those variables is

[00:17:30]	useful. If you need to see that, that resource is available under appendix C.
[00:18:00] [00:18:30] [00:19:00] [00:19:30] [00:20:00]	Navigating back to the user guide I will now look at and so appendix E talks about the methodology by which we did our research to identify the relevant laws, statutes and regulations, that aligned to the taxonomy that was developed initially by the Scientific Expert Panel, so as Daniel was saying the Scientific Expert Panel developed the taxonomy including, identified the variables within those taxonomies, and so then we used Lexis.com to identify the relevant statutes and regulations that align to those variables and then we actually performed the legal review. Here we highlight the various keywords that were used in Lexis to identify the relevant regulations and statutes. Switching back, the legal research methodology which I just mentioned, so here we actually talk through the details of the methodology itself. I'm not going to spend too much time on this, other than to say that it was a comprehensive process and we felt like it was important to document our steps in order for the data set to be used appropriately. Appendix F here talks about where we actually got our statutes and regulations from, and talks about some of the definitions of each of those categories.
[00:20:30]	I mentioned I was going to skip over the SHPDR taxonomy and we looked at appendix B which showed us the topics, subtopics and variable descriptions. We can actually look at the taxonomy itself here, and navigate to it through the menu bar up top. Before we do that I again want to point out that we've got the download button here, where again you can download what you're seeing here on the website for local use.
[00:21:00] [00:21:30] [00:22:00] [00:22:30]	Switching to the taxonomy now. The taxonomy we looked at it from one point of view but this is a slightly different point of view without the definitions. Showing you the insurance domain here and the topics and the subtopics and the variables that we've defined here. These are the variables that appear in our data set for each of the 50 states and the variable values then are coded yes or no. Insurance domain is on the first tab and the provider domain is on the second tab. It is color coded here if you notice and what that represents is the fact that we were able to develop a set of sub-variables from the key variables that were identified, and so we've essentially identified those categories of variables that are similar to each other, or that are derivatives of the main variable and we've designated them to be the sub-variables of the variable that was initially identified. That's what the color coding means and you can see here that we've got that sub-variable prefix in the variable title. Again, we can download the taxonomy by clicking on the download button. That can then be saved on your local desktop.
[00:23:00]	Before I actually move to the data tab I just wanted to show that there is a contact page. If you have questions, suggestions, that is the best way to get in touch with the team and we will then respond to you, confirm receipt within 48 hours and assist you or answer any questions that you may have.
[00:23:30]	Now I will navigate to the data tab and the data tab is where we can actually interact with our data set. On the top here you'll see a set of filters, and so the first filter that's selected is by state. Here we're able to click on a state and then see all of the variables that have been identified for that particular state for both 2010 and 2012 years. In this case we're going to select, let's say New York. In selecting New York we see a data table here that has

[00:24:00]	the year column, variable, variable value, and justification. We talked about those data elements when we were talking about the user guide. Here is the variables or the sub-variables that are present in the data set and ultimately their value and the justification for why that particular data set in this case is a yes.
[00:24:30] [00:25:00]	We can increase the number of entries that appear on a page, and so here we would have 100 entries and a really long page here to scroll through and to see the entries and if we do get all the way to the bottom here, we'll actually see the total that are in New York for 2010 and 2012. The detail that's provided in the table does not include all of the elements that are in the data set, so to see additional elements that are not shown on the table here, we can actually click on the record and then a pop-up would appear and give us some more information about that particular entry.
[00:25:30] [00:26:00] [00:26:30]	Year, this is 2010, we know that. State, New York. Domain topic and subtopic for the variable, in this case it's in the provider domain for health IT. The variable here, in this case a sub-variable, is the State Health IT Authority authorized demonstration of projects to implement health IT. The value for that variable is yes. The justification is here as well. It provides some context for why that yes was coded. We have the citation itself, which is a link, and I'll click on that in a minute. It is a statute, in this case from the following section, commissioner, general powers and duties. Chapter, title. We can actually click on this link of the citation and we now see the actual text of that citation from which the variable was coded from. Here researchers can look at the actual law and verify if they like the value and then learn other context about or other information about that particular variable.
[00:27:00] [00:27:30] [00:28:00]	In switching to by year filter, so in this case in the data set we have 2010 and 2012. If we click on 2012 we'll see all of the entries for 2012 for all the states and so proceed accordingly here. You can also download the data set that is currently presented to you, and so the entries that appear here, there's actually 13,637 entries, so once you actually filter down your data set to the variables of interest and you can do that through the search box here. Let's say we put in health IT, by putting in that term, we now see all the hits and now we've narrowed down that data set to 178 entries, and provided that that was the data set that you wanted to download, you can do that by clicking on the download. And at that point you can see the data set only or you can download the data dictionary for the data set that you've selected or both.
[00:28:30] [00:29:00] [00:29:30] [00:30:00]	I'll just download the data set here at this point. You'll actually have a different prompt that asks what format and I'll choose Excel and the system will generate the file, we'll save it here, and then I can open it up and that should be, and this is my data set here, with all the elements rather. You can navigate the data set by State, Domain, two domains, by topics. Again the topics underneath each of the domains and subtopics and by specific variables here. I do [inaudible 00:29:26] You may have noticed that some of the values in our data set, some of the variable values, were we thought not identified and so we actually have, this is a new feature that was recently implemented, where we have [inaudible 00:30:29] to the variable values where we essentially did not find, and through our research find any information or any search results, we did not yield any search results for that particular search term that was related to the variable that we were looking at.

[00:30:30] [00:31:00]	In this case, for example, State restricts provider balance billing managed care enrollees. In Texas our search did not find any hits where that was the case and so we made sure that we provided that cue to the user that, in this particular case, our search didn't yield any results. There could be laws or statutes or regulations using other search terminology that could yield a yes or a no, but in our particular search we did not find any, a positive or a negative. That's what the not-identified value means for the variables that you see here.
[00:31:30] [00:32:00] [00:32:30] [00:33:00]	The final way to work with the data set is through our custom filter, and so where the filters beforehand were just by one particular element, state, year, domain, topic, subtopic, the custom filter actually allows you to select multiple elements to tune the data set of interest. I'm going to choose, let's say provider and then I'll choose care delivery and financing, and so with each option I'm actually narrowing down my data set and then I can actually go to state, multiple states, Alabama and Alaska and maybe Arizona. Then just to say for 2012, so at this point I've narrowed down my data set to 78 entries, and I can go ahead and then download the data set for local analysis. At this point, and here are the prompts to do that, and I've already shown that to you so I'm not going to bore you with that, that concludes the presentation of the website. Again if you have any questions about how to use the website, or if you have any questions about the data set itself, please reach out to us through the contact us tab and we'll get back to you promptly.
Daniel Weinberg: [00:33:30]	This portion of the webinar will use a sample research question to illustrate the basic steps necessary to use SHPDR data in the research context. The sample research question relates to colorectal cancer screening. In particular, what is the association between coverage of colorectal cancer screening and death rate and incidence of colorectal cancer? This research question is linked to the existing literature investigating the relationship between screening and patient outcomes.
[00:34:00]	For this demonstration we will use two data sources. A complete study might use additional information from other sources as we'll describe in the next step section toward the end of this presentation. We use SHPDR data on whether states require coverage of the colorectal cancer screening benefit. The outcomes come from the CDC's U.S. Cancer Statistics Program. We will use basic methods for illustration purposes, including descriptive statistics and linear regression.
[00:34:30]	First let's discuss data from SHPDR. All data in the insurance domain of SHPDR is for the individual and small group markets. This particular variable tells us whether the state requires insurers in those markets to cover colorectal cancer screening. The legal text available in SHPDR may also give insights into other insurance markets, but the coded variables pertain to individual and small group only. At the end of the SHPDR data preparation process, the goal is to have a state-by-year data set that we can merge with our outcome data.
[00:35:00]	As Slava demonstrated, first we must identify the SHPDR variable of interest. There are several ways to identify variables. I use the second option on the slide where I looked at the by variables page, and searched for colorectal. Next I extracted the data to Excel. The spreadsheet I downloaded includes all of the state-by-year data for the chosen variable, including justifications, citations and other information. It also includes the domain, topic

	and subtopic under which the variable falls.
[00:35:30] [00:36:00]	The next step is data preparation. There are 30 states in the data set I downloaded with variable values other than not identified. If each state had one record per year, one for 2010 and one for 2012, we would have 60 records, but we actually have 129 records. This is because there are some states with multiple records for each year. Within a state-year combination, the difference between records is the citations and justifications. In most cases the records are consistent in terms of their coding. Either the benefit is covered in the state in that year, or it is not. But in two cases there are differences, discussed on the next slide.
[00:36:30] [00:37:00]	In 2010 and 2012, Illinois has one record indicating that screening must be covered, and another record indicating that it need not be covered. First I reviewed the justifications and found that, in general, it appears that screening must be covered. There also appears to be a pilot program related to colorectal cancer screening. To investigate further I reviewed the legal text. To see the full text of the statute or regulation, click on the record of interest. This results in the pop-up in the screenshot to the right. Then click on the citation, which gives a pop-up containing the legal text, shown to the left. After reading the legal text for the yes and no records, we coded Illinois as a yes. Screening is covered because the record with the no refers to a specific government pilot program, while the record with the yes refers more generally to the state's rules. After these types of questions are resolved, the SHPDR data are prepared and ready for combining with the other data set.
[00:37:30]	As discussed previously the outcome variables are death rates from colorectal cancer and incidence rates of colorectal cancer. These data come from CDC's U.S. Cancer Statistics Program and are expressed per 100,000 population. The data are available by state, year, gender, and race cohorts. We downloaded the various data files and appended them for the relevant years and cohorts so that each record was uniquely identified by a combination of state, year, race, and gender.
[00:38:00]	Now that we have the data we are ready to combine the files. This involves importing the data into a statistical package of your choice and preparing the data if needed. Additional data preparation could include, for example, variable formatting. Lastly the data are merged by state and year. I used Stata for this example but SAS and other programs could have been used as well.
[00:38:30]	We then use the analytic file to produce descriptive statistics. Stratified by the SHPDR coverage variable and year. Changes in incidents over time moves similarly for states in which screening was covered, and those in which it was not. There doesn't appear to be a clear pattern in the incidence levels across states with and without coverage requirements. We did the same analysis for death rates and found similar results.
[00:39:00]	Lastly we ran ordinary least squares regression. The top panel uses incidence rate as the outcome variable. The bottom panel uses death rate. In addition to the SHPDR variable on screening coverage, both regressions control for gender and race ethnicity. We see that females tend to have lower incidence and death rates than do males. Hispanics have lower incidence and death rates compared to non-Hispanics. Blacks have higher incidence and

	death rates compared to whites. Coverage of colorectal cancer screening is not statistically significantly related to incidence or death.
[00:39:30] [00:40:00]	For this demonstration, we used SHPDR data to identify states in which colorectal cancer screening was required. We prepared the SHPDR data and then used Stata to merge SHPDR information with colorectal cancer incidence and death rates from CDC. Then we examined the association between mandated coverage and the two outcomes. We found no statistically significant association. As I mentioned previously, this research example is for illustration purposes only, and there are many ways to refine the study. It would be nice, for example, to identify the causal impact of required screening on colorectal cancer outcomes. Doing so requires additional data and methodological rigor. One could use more granular data such as all-payer claims, and you would also want to control for additional covariates which would likely require additional data sources.
[00:40:30]	It's also important to match the cohort for which we have outcome data to the cohort to which the screening laws apply. There also may be scope within state contemporaneous heterogeneity, depending on which insurers are required to cover screening. It's also likely that state's decisions to require screening are endogenous, requiring instrumental variables or some other quasi-experimental method. There may be heterogeneity in the screening requirements which could provide interesting variation across states. For example, whether the state requires screening in general, or guideline concordant screening.
[00:41:00]	Thank you very much for your attention and interest in the SHPDR. We don't see any questions in the chat function. Unless we receive some now, we'll adjourn. Are there any additional questions? Seeing none, we'll adjourn. Thank you very much.