National Cancer Institute
Tobacco Use Supplement to the Current Population Survey (TUS-CPS)
2021 Data User Webinar – County Level Analyses on TUS-CPS Using Small Area Estimation Techniques

Webinar Transcript

MS. NALINI CORCY: Hello, everyone, and thank you for your patience. Thank you so much for joining us today. We're going to go ahead and get started. Today, we will be doing the fourth webinar in our Data User Webinar series for the Tobacco Use Supplement to the Current Population Survey. And today's webinar is focused on conducting small-area estimation and using that to obtain county-level analyses on the TUS-CPS data.

Next slide, please. So, just a couple of housekeeping items before we get started. So first, all participants are on mute. If anytime you have requests for any technical assistance or you have questions for the presenters regarding the content, please type it into the chat box. Any requests for technical assistance, we'll try to get to those as soon as possible and assist you. Regarding questions, we'll be saving them for designated question-and-answer periods throughout the webinar, so thank you for your patience with that. We also have a closed captioning service available today. And I'm dropping a link to that in the chat box right now. So if you need that service, feel free to click on that link. The webinar today is being recorded. All of the materials will be posted online in approximately three to four weeks' time. Once they are available, we will send you an email letting you know how you can access them.

So I would like to introduce our speakers for today. First we have Ms. MS. HARTMAN. She is a biostatistician and program director with the Tobacco Control Research Branch at NCI. Also joining us today is Dr. DR. LIU. She is a survey statistician and a program director with the Statistical Research and Applications branch at NCI. So Anne and Benmei, whenever you're ready, please go ahead. Anne, I believe you're on mute.

MS. ANNE HARTMAN: Thank you very much, Nalini. And that was just a disclaimer that these are the presenters' {views]. They do not necessarily represent HHS and all of the other government agencies. So thanks.

We're going to do a little bit of a tag team, Benmei and I. So, I will start off. Since not everybody has attended all of the webinars, we assume, we will start out with some basic information that the TUS-CPS features. Then we'll go on to talk about why small area estimates are needed. And our general goals of doing this project and rationale for specifically two unique variables that we have done the small area estimates, two policy items. That's very unique to the Tobacco Use Supplement to be able to provide that. Then Benmei will continue with the overview of small area estimation techniques and how they're applied to the two policy items, the results from the two policy items, some dissemination information, and website information where you can get the small area estimates, and some future plans for future waves of the Tobacco Use Supplement. And then finally, we'll have some discussion of statistical and policy implications and some broader context, and then the Q&A. Thank you.

So, just very briefly – sorry if some of you have been on for and you probably can actually present the TUS-CPS features, but anyway, it's a key source of U.S. national, state, and even

some sub-state-level data on tobacco use and tobacco control policy and social norms. The Tobacco Use Supplement is actually a supplement to the Bureau of Labor Statistics and Census Bureau's Current Population Survey. That survey is actually conducted by the Census Bureau. It consists of the national complex probability address-based household sample. CPS is conducted monthly using a panel design. And it provides data on demography, labor force, unemployment, and many other key information that's important also to economics and other things other than tobacco use. And there are other supplements, other ones on other topics. So, sometimes you can actually link those things up, and it's very useful. NCI has sponsored this survey since 1972. And recently since 2014, FDA's Center for Tobacco Products and NCI are co-sponsoring the supplement. It's usually been fielded about every two to four years. It provides information on 150,000 and more self-reports from the civilian population 18 and over. Typically, interviews are conducted 35 percent in person with an in-person visit and 65 percent, about 65 percent by phone by the same field reps that do in-person interviews. And because it's a panel, this is more efficient so that sometimes they go to the household, especially the first time and the fifth time, which is a year later. For reports, it's important to be in person. And many of the other times they contact the household by telephone, so it's not like a "you have to dial in," kind of thing. It's also translated into Spanish, and on the occasion that they have other languages that the field reps do not speak, they will get someone from the community. So, it's not just an English language survey. So, that's another feature.

So, why small area estimation for the TUS-CPS? Its design allows reliable national and state-level estimates. But, as you know, you're interested and you're here, policymakers can't control planners, and other researchers often need tobacco-related county-level data to evaluate tobacco control programs, monitor progress, and conduct other research. So, the TUS standard direct estimation methods cannot provide reliable county-level estimates due to small or zero sample. That does not mean that some counties are large enough or some principal cities are large enough and can also be identified on a couple of these files, but most of the counties require restricted data use, and will be an opportunity to talk a little bit more about that.

One thing I just want to tell you, maybe repeated throughout, is that if you've looked at the technical documentation, Appendix 11 is essential because it talks about all the different geographic-type variables and how you can combine things to get smaller than state estimates. Therefore, we need model-based methods that combine information from multiple related sources to increase precision. So, our research goals in doing the small area estimation project was to produce model-based county-level estimates in key measures. Currently, we have it for 2014 and '15 wave, and of course, that's on adults 18+. That's always because we have to get the data and have it available before we can do small area estimates.

So, the most current TUS wave, in general, is the 2018-'19. So, anyway, what did we do? We looked at the percent of population currently smoking, the percent that ever smoked, the percent that quit for 24 or more hours among those who smoked within the past 12 months, and then, two of the unique policy variables that we can get from the Tobacco Use Supplement. That is, the percent working population reporting a smoke-free – and we used "SF" on the rest of the slides – workplace policy and percent population reporting a smoke-free home rule. That actually is defined where no one is allowed to smoke anywhere inside the home at any time, and the previous – the workplace policy, people are never allowed to smoke in any work areas or any public common areas at the workplace. This was conducted through a collaboration between NCI and the Census Bureau,

specifically, Isaac Dompreh. And this talk will mainly focus on the two smoke-free policy outcomes. Thank you.

What's important about smoke-free home and workplace policy? Hopefully, because you're interested in tobacco control, that's pretty obvious. But, just to make sure everybody's on the same playing field, we're interested in a Tobacco Use Supplement that's designed to consider policy as well as use because, secondhand smoke exposure is causally linked to many chronic diseases in adults who do not smoke, plus Smokers as well, as well as serious illness in children. So, what are the major settings to secondhand smoke? Well, private workplaces and public places such as bars, restaurants, and recreational settings, and homes. So, salient sources of workplace, for adults, that's -- I mean, for adults, salient secondhand smoke source, it's the workplace for adults in general. Home is for children, the unemployed, and retired persons. Thus, workplace and homes are important settings for implementing-evidence-based strategies to reduce secondhand smoke. Next slide.

So, workplace policy also has benefits. Of course, it protects workers from secondhand smoke adverse health effects, but also very important, it reduces active smoking behaviors, both prevalence of cigarette smoking and intensity, and it yields of course a safer, more efficient working environment. How have these been established? Typically, they've been established by state or local legislation, or they've been adopted voluntarily by employers or building owners. Recently, actually, over time, historically, there's been an increase in local jurisdictions enacting smoke-free environments, mainly work areas and public places, especially important in states without strong state smoke-free laws. This is how we've come to this over the last 20 years.

Now, the benefits of home rules, besides actually protecting individuals from secondhand smoke harm, it's shown to be very strong for preventing youth, young adult smoking initiation and even stronger for increasing cessation among adults. Even some people have reduced reports, while most of TUS is cross-sectional, there have been a few instances where we've done a short prospective follow-up, and there's been some papers published showing that not only cross-sectionally that home rules are very important, but also prospectively with follow-up. So, what is the situation with rules? Well, for the most part, home smoke-free secondhand-smoke restrictions are rarely covered by smoke-free legislation as opposed to the workplace policies. They're generally established by adult home residents. When it comes to multiunit housing, it's typically been imposed by voluntary action, by landlords, building owners, or, again, individual tenants. There have been over the last several years some, several, a handful of California localities that have limited legislation covering some types of multiunit housing.

Recently, in 2017, Department of Housing and Urban Development, HUD, required all public multiunit housing being 100 percent smoke-free by the end of July 2018. While this is very advantageous, it's an invaluable benefit to those residents impacted, and so we applaud those efforts, it is important to realize that only about 200 – sorry, only about 2 million – not 200 million – 2 million of greater than 300 million U.S. residents are impacted. So, it's a great step, but it does not take care of everything.

So, now, why small area estimates? Maybe they're available from other sources. First, there's been few publications providing detailed geospatial variation, indoor smoke-free workplace policies or home-rule coverage. Babb et al. studied variation in smoke-free workplace policies across states, but not at lower geographic levels. There are comprehensive ordinance lists compiled by the

American Nonsmokers' Rights Foundation, or ANRF, that provide information on the presence or absence of an ordinance, but not as implementation or enforcement. There are tables that provide the percent of population covered by three categories of laws, mainly nonhospitality workplaces, restaurants, and freestanding bars.

Now I turn the time over to Benmei.

DR. BENMEI LIU: Thanks, Anne, for a very nice introduction to our talk. So, I'll cover some of the technical details on how the estimates are produced. I'll start with a brief overview of the model-based small-area estimation techniques. Some of you may already be familiar with this field or some may be new to this field. So, I'll cover some basic ideas. The key idea of model-based small-area techniques is to borrow strength from relevant sources, for example, Census' American Community Survey, or administrative records and other areas with similar characteristics to increase the precision of small-area estimates. Choosing a good small-area model is the key because the inferences are mainly reliant on the assumed model. They also need a good statistical methodology to make inferences. Mixed models, including both fixed effects and random effects, have been popularly used in producing small-area estimates because of its flexibility in combining the information from different sources and taking account of different sources of error. Models can be built at both the error level or individual level, and different inferences, methodology are involved, depending on the model that you choose. For our research, we focus on error-level models. Among the many models developed in the small-area estimation literature, a commonly used well-known mixed model is the Fay-Herriot model, which was originally developed to estimate per-capita income for U.S. areas with populations of less than 1,000.

Next slide, please. So, the fundamental Fay-Herriot model consists of two levels of models, the sampling model and the linking model. The sampling model assumes the direct estimates at the error level follows a normal distribution with the error mean, theta-I under sample variance φ_i. The linking model assumes that error mean, theta-I relates to a set of covariants obtained from external sources or so appear. The V_i is the random effect. I is the index for the areas to be estimated. The direct estimates in the sampling model, D_i, can be computed using the standard server software like SUDAAN SAS Proxy Survey Package or STATA. The sampling variances, χ_i, needs to be estimated and smoothed before we can run the small-area model. This is just a very basic area-level model. There are much more complicated models developed in the literature, but I didn't touch those here, just throw some – like, a basic idea on the model. Next slide, please.

So, in theory, the final estimates are combinations of the direct estimates linked to the synthetic estimates through modeling. When there is sufficient survey data for the small area, the combined estimates depend largely on the direct estimates computed for that area. When there is little or no data available for a small area, then the combined estimates increasingly depend on the assumed model to produce estimates for areas with similar characteristics. For the direct model I just presented, a close form of the final estimate can be derived, but for some more complicated models, there's no close form, so we have to use other approaches, like Bayesian approach or some asymptotic estimates, to estimate the final estimates. The commonly used approach are fully Bayesian approach or empirical best prediction approach, which involves some analytic formulas. Again, mostly are asymptotic to derive the final estimates. Next slide, please.

After some comparison of several potential models through simulation, we eventually chose the Faye-Herriot class of model with the arcsin square root transformation for the sampling model

because the direct estimate is proportions. We found that arcsin square root transformation to the direct proportion works better for our data. The pool of auxiliary variables included 30-plus county-level demographic and socioeconomic variables obtained from the American Community Survey five-year average from 2011 through 2015 to match with the TUS-CPS 2014-15 data period. We also used data from Census 2000 and 2010 and several variables obtained from other administrative resources. We also included five state-level tobacco policy data, including cigarette taxes, clean air laws, tobacco control funding, et cetera.

Before running our small-area model, we run classical model selection procedures to reduce the number of auxiliary variables for each outcome because we used [00:24:52 inaudible] approach. If we included too many covariants, sometimes it causes a lot of problems in addition to increasing the running time greatly. So, we choose the most predictable covariants to be included in the final model. So, hierarchical Bayesian approach through Markov Chain Monte Carlo methods were used to estimate the parameters of the statistical models. We did extensive model selection and a model diagnosis to select the final models. We also assessed the goodness of fit for each model. We would use several manners [00:25:33 inaudible], like the famous [00:25:38 inaudible] book. They also evaluated the model by comparing the model-based estimates to available direct estimates. The ratio of the two types of estimates is expected to converge to 1 as the sample size is larger. Next slide, please.

So, this plot shows the ratio of the direct over modeled estimates for the proportions of workers covered by smoke-free workplace policies against sample size. It's under a locked scale. The X axis shows the sample size, which is under a locked scale, and the Y axis shows the ratio. As I mentioned, when the sample size goes larger, we expect that the two estimates go closely to each other. So, the ratio would be approaching to 1. So, the funnel shape shown here is what is expected. We did search validation for all the outcomes that we investigated. So, this is just one of the plots, and the patterns are similar.

Next slide, please. So, some of the map, show you some final resource. The map on the left side shows the model-based estimates for percent of population governed by a smoke-free workplace policy, among ages 18+ based on the TUS-CPS '14-'15 data have circle. The lighter color on this map showing higher coverage. The darker color means lower coverage. So, we already used this color scale for our small area estimates website. I'll talk about the website later. Darker color shows less coverage. And so we can see from this map, some of the western and southern states have clusters of low-coverage counties. The map on the right shows that there are states that have 100 percent smoke-free air laws posted by the American Nonsmokers' Rights Foundation, covered earlier. This map was from April 2, 2015.

I think they update this map every year in spring. This was obtained from 2015, which is very close to the data period of the TUS-CPS data that we are showing here. And for their map, the darker blue means states with 100 percent smoke-free laws in workplaces, restaurants, and bars. The lighter blue means 100 percent smoke-free laws in one or two of the three categories but not all three. And the green means no state smoke-free laws, and the small triangle or dots means local, city, or county with 100 percent smoke-free laws in places. So, the two maps are not exactly comparable, because the right-side map is a state map and the left side, our map is a county-level map. But we can see some general consistency between the self-reported model estimates based on the small error estimation technique, with the state smoke-free laws on the right side, the areas where there are no

state laws, it seems they have lower self-reporting the smoke-free policy coverage. And our results show the patterns for the states without state 100 percent smoke-free laws. For example, in Texas, you can see some clustering within the state. But Texas didn't have a smoke-free, 100 percent smoke-free laws at the time, back in 2015.

Next slide, please. So, this map shows the county-level model-based estimates for percent of smoke-free homes. Again, the darker color means lower coverage. The lighter color means higher coverage. And we can see some clear patterns by looking at this map. Like, the darker color, more condensed in the east side and the central side. I think I will discuss the implications of these results after I finish my slides. Next slide, please.

So, for data accessibility, after we produced our small error estimates, so far, we have produced estimates using the 2010-2011 data and 2014-2015 data. So, we released the results on our small error estimation website, sae.cancer.gov. Users can view or download the results in the format of tables or maps from this website. This website actually contains three small area estimation projects. One is based on data combining the National Health Interview Survey and the BRFSS. And then the second one is estimates based on the Tobacco Use Supplement. And they also have results from a third project, based on our trends data. But the page I'm showing here is for the Tobacco Use Supplement small area estimation results. For those of you who are interested, you can visit our website. It has the introduction, methodology, and some of the publications we posted there.

Next slide, please. So, we just started working on estimation using the 2018-2019 dataset. It has been delayed because we have to build the collaboration paperwork with the Census Bureau. The pandemic delayed us somehow. But now the paperwork is ready, so we are making the wheels going. In addition to the five outcomes we have considered in the past, this year we are adding a couple new outcomes, including menthol use and e-cigarette use. We also considered evaluating the current approach and might consider some improved modeling approach, if needed. So, the work is ongoing. Next slide, please.

Some discussion from the statistical point of view. Generating the county-level model-based estimates for prevalence of five smoke-free-related outcomes for 3,134 U.S. counties and equivalents, there were a few – I think nine county equivalents we couldn’t adapt to the covariates from the census websites. So, we couldn't produce estimates for those few counties. We applied hierarchical Bayesian models allowing – borrowing strength from covariates and other counties with similar profiles. And we did extensive model selection and diagnostics to choose the best model that works for our data. And for all the outcomes, the model estimates showed consistency with direct estimates in aggregate and reduced variance for each county in a general sense. We showed some more data in our paper. We have a citation at the end of this talk.

Our results demonstrated the positive features of [00:35:15] strength for the small area estimation techniques for small areas with large variation throughout relatively stable estimates. Now, I'll turn the speaker to Anne to finish the rest. Thanks a lot for your attention.

MS. CORCY: Anne, I'm sorry, this is Nalini. You're on mute again.

MS. HARTMAN: I didn't want any bells or whistles coming through the computer to interfere with Benmei, so thank you. So, the first study of this – this has been the first study of

smoke-free home rules and workplace policies, county-level estimates from self-reported national survey data. The self-reports measure for workplace policy are actually more an indirect measure than, for instance, something like the ANRF data. These ordinances and laws are included, but also any of those applied voluntarily by employers or building owners are included in this. So, it is a bit more comprehensive to that extent, if it's not just considering laws, but really considering smoke-free work environments. That includes, like I said, all types of workplaces, including hospitality sectors like restaurants, bars, casinos, nonhospitality sector, as well as those in the government.

So, we have a variable that also has, in terms of occupation, work for the state, national, local government, there's also a lot of detailed occupation. So, one could look at that. And all of that actually is combined with the small area estimates. So, in general, more nationally, it's about 80 percent of the U.S. workers 18 and over in 2014-2015, reported smoke-free workplace policy coverage. Yet, no state achieved greater than 90 percent smoke-free workplace coverage, which suggests, even in states with comprehensive statewide smoke-free laws, some workers may remain unprotected. And also, actually, we got a fairly high percentage across the nation, in the U.S., of greater than 85 percent of adults reporting coverage by smoke-free home laws. Next slide, please.

So, a couple of things that I wanted to note. The lowest percent for smoke-free workplace policies and often lowest percent of smoke-free home rules are in states with the highest smoking prevalence, mainly the U.S. south and central parts. So, while Benmei showed – and this is consistent with past results. Next slide.

Just, quickly. So, this is just, again, showing you, the highest smoking prevalence actually is usually the south and eastern central part of the country. Next slide. And in the home rules, a little less so because it's mainly voluntary, so it's not dependent on tobacco – states that grow tobacco, manufacture tobacco, lobby heavily and so there are fewer smoke-free laws there. But since the home is a place that's really voluntary, it's a little – while there's the correlation with smoking prevalence, there's less because we even find that some current smokers do have smoke-free policies in their home. They want to quit. They say they want to quit. So, while it's lower than nonsmokers' homes, it has grown over the years, even in nonsmokers' homes.

Next slide, please. So, the other thing I want to point out is that, in general, states with minimal within-state variation are largely those with strong state-level tobacco control policies, possibly compensating for counties that do not have their own laws. One example I really want to point out is California. It's got a large within-state variation for workplace policy, partly explained by California's early adoption of comprehensive tobacco control statewide law at a time when a lot of exemptions were common, in effect penalizing California.

And I will actually – next slide. So, I'll just show you this, that it's really up to the local – so there's a great deal of variation because, while California had a smoke-free state law, it was rather weak over time compared to many other states that adopted things later. So, luckily, a lot of things in early times, over the last two decades, really began at the local level. And that shows the heterogeneity here.

Next slide, please. But in contrast, California has a high smoke-free home rule percent and minimal variation between counties, likely a result of its program, regardless of funding, always emphasizing social norms. So, one thing we must realize is social norms really have an impact, and as I

said earlier, we've seen impact on increasing cessation and decreasing initiation. So, don't only think that it's the workplace policies. Home is very important.

Next slide. So, that's here, showing you that the home is very – it's rather homogeneous compared to what we saw. And it's also got high – remember, the light colors mean high smoke-free home rule. So, that's just illustrating that.

OK, next slide. Some broader implications of these results. There actually was a recent study that correlated the Tobacco Use Supplement, smoke-free workplace policy, and American Nonsmokers' Rights Foundation, constructing certain scores, because if you remember, their data is a bit different. They also have three categories of laws. And one thing that's really encouraging is that they found consistency between the two measures in terms of observed impact on smoking cessation and the number of cigarettes smoked per day. So, as I previously said that workplace policies can reduce how many cigarettes are smoked and also the prevalence. The other thing that is good about the TUS that's different than, say, ANRF data, is that there are some advantages of self-report while everybody – some people say, oh, self-report, why consider that? Because it includes more than just there being a law. It includes the actual impact of the law for understanding of the workers, their perception as well as knowledge, which could reflect enforcement strategies.

The other thing is, there's detailed smoke-free county and state-level workplace policies, and home rules can help identify coverage disparities, like Benmei showed in the maps, and differential impact by looking at behaviors. And they also can help to tailor interventions. If you know the nature of the problem and, you know, maybe it's different in different places. So, the other thing is the framework – let me just – OK. The framework is useful for modeling different tobacco control variables. As Benmei mentioned, we're considering menthol use, which is certainly a very hot topic, and e-cigarette use. So, different tobacco-control variables, and it also can be applied to other areas, such as other behaviors, other policy or health topics.

And as promised by Benmei, here are some references here. I want to note especially the first citation, which is the paper that's been published that our presentation's based on. And that's the link and citation and some of the others that Benmei mentioned.

So, with that, I'd like to thank you for your attention. Oh, this is important and not just old information. And again, thank you for your attention, and we're happy to answer questions.

MS. CORCY: Thank you, Anne, and Benmei. So, once again, if you do have a question, please feel free to type into the chat box. The other option is that you can raise your hand. Here are instructions on this slide on how to do that. It's at the bottom right of the participants panel. You'll see a little hand icon. So, if you raise your hand, I can actually unmute you if you'd like to ask your question and discuss with the presenters. But otherwise, feel free to type into the chat box.

I know, Anne and Benmei, that there was one question that had been submitted through the registration. I don't know if you wanted to tackle that one now.

DR. LIU: Yeah. Anne, do you want – you're on mute.

MS. HARTMAN: I am on mute?

DR. LIU: Now it's all good.

MS. HARTMAN: Can you show the question?

MS. CORCY: Yeah, I'll read it out. So, the question was, what is the percentage of TUS-CPS respondents with nonmissing county information? Is county variable contained in the public-use TUS-CPS data?

MS. HARTMAN: Benmei, do you want to start, and then I can add?

DR. LIU: OK, yeah. So, all the TUS-CPS respondents have valid county information is in the restricted-use files. However, in the public use TUS-CPS data, only counties above the population size threshold, which is bigger, equal to 100,000 population size, have valid county information. There's a variable name that JTCO with values between 001 to 840 means that county FIPS code is valid. The rest counties were masked together with assigned JTCO code equal to 000. So, I think it's about one third of the counties have valid county FIPS code, but you have to use the county FIPS code together with the state-based FIPS code to uniquely identify your county.

MS. HARTMAN: I'd like to add for everybody's benefit a critical reference. In addition to the record layout, which identifies variables, there's an attachment number 11 called the Specific Metropolitan Identifiers. And it contains lists in the technical documentation for each month's CPS or TUS-CPS fielding, which includes the names of the geographical areas, counties, and principal cities and coding to be used for some useful geographical variables.

These variables include GTCSA, which stands for FIPS Consolidated Statistical Area; GTCVSA, which is a FIPS Metropolitan Area; and then GTINDBBC, which is for Individual Principal City; and variables such as GTCBSAST, consisting of combinations of counties and/or principal cities. So, these are very useful to start because if you want, really, very small-level, local information, you can for certain purposes be able to use the public datasets and for other things where you really need all of the county's information to access restricted data through the research data centers or possibly through some of our interagency agreements, Census and other NIH entities and FDA and CHS. So, hope that at least gives you a start in terms of reference.

MS. CORCY: OK, thank you. I don't see any other questions in the chat box at this time. But, again, if you have any, please go ahead and type them in or raise your hand and I will unmute you. We'll just give everyone a couple minutes in case there are any final questions.

MS. HARTMAN: Also happy to answer questions after the presentation as well. So, if it takes you a while to think about it, we're happy to help.

MS. CORCY: I don't see any questions coming through, so I think we can go to the next part. So, once again, just thank you to everyone for participating today. Like Anne said, if you do have questions for us afterwards, please feel free to reach out. Our team contact is the email at the very bottom of this slide. You can also visit our website at cancercontrol.cancer.gov/TUS-CPS. We also have an email listserv that you can subscribe to for the latest news and updates on TUS.

And finally, we'd love to get your feedback on today's webinar. I'm going to drop the survey link in the chat box, so if you could please take two minutes out of your day just to give us feedback on what you think of the content and the format of this webinar, that would be very

appreciated. And once again, thank you for joining, and we will be in touch once the materials are available online. Thank you, everyone.

DR. LIU: Thank you.

[Event concluded]