>> Marshall McBean: They've got the beneficiary levels information. I'll talk during this presentation about the enrollment or denominator information, and then we'll move on to the PDE information and the characteristics files, which are the numerator data.

So acronyms, some of which I've already gone on -- talked about before. Some new ones, the enrollment database, which is where enrollment information comes from. Research Triangle Institute -- RTI for those of you who don't know, RTI CAHPS the Consumer Assessment of Healthcare Providers and System, which is a survey that CMS has run to try to understand provider satisfaction. PDE of course, and then some new ones here. QMB, SLMB and QI. And I was talking to some friends during break. You may just want to throw me out by the time I get done talking about QMBs, SLMBs and QIs, but there are some people here from Medicaid and they will come to my rescue. Just groups of people who are in what are called Medicare Savings Programs, so they're -- have higher either resources or incomes than people who are in traditional Medicaid. And these are programs that are funded by the federal government, but administered by state Medicaid programs. So Medicare Savings Programs. Here they are. QMB, SLMB, QI and also QDWI and again, they're funded by the federal government but administered by Medicaid. And for those of you who have heard us talk before, they are all called State Buy-In Programs or State Buy-In Beneficiaries. I'll talk more about that. PACE Program, which is the Program for All-inclusive Care in the Elderly, then LIS, PDP, and MAPD.

Okay, so I know some of you have taken the Medicare CMS 101 course, but not that long ago that you've maybe not heard about the Denominator with as much enthusiasm or gusto. But, at the beginning there was one Denominator file. Okay, there was the old, loveable denominator. It was 80 characters -- or 80 columns long and as it says here, we suggested everyone get it. They just needed it, not because I'm an epidemiologist, but you really needed it to understand your populations and to get the correct beneficiaries. Now, over the last let's say three to four years, thing have changed. First of all, the Denominator file is no longer available. So you have to start learning about the Beneficiary Summary File when you start ordering your data. For a certain time period, there was a separate Part D Denominator file, but that's gone away and what has now happened is it has been combined -- both the old denominator and the new denominator information into the Bene Summary File, which then got eaten up by the Master Beneficiary Summary File.

So if you look at the next slide, it says here -- now this Master Beneficiary Summary File has four segments. So one is the Bene Summary File. Okay, so this is A, B, C, and D enrollment information. So this is when you're thinking denominator with a small D, this is the file you're interested in. The Master Beneficiary Summary File has these other segments. One is the chronic conditions, which we don't talk about here. But those are the 26 chronic conditions that CMS has identified, including BPH, hypertension, cardiovascular diseases. And if you want to get to finding a cohort quickly and you're happy with the algorithms that were used to select these chronic conditions, this is a good way to get information on those beneficiaries. There's a cost and utilization segment which

summarizes at the beneficiary level the -- notice they use the word costs, it should say expenditures, Medicare payments and utilization by such categories as Carrier file, Inpatient file, SNF. So my major groups could be interesting for some studies. And then finally, to note that the National Death Index information for all dead people is available through 2008. There may be more recent data coming, but that hasn't been negotiated yet.

But this presentation again is about the Beneficiary Summary File segment, and if you go to the website, sometimes you'll see it has part AB or part ABD, just so you don't get confused. Right. Questions about the death indicator, and I won't talk much about it, but there is -- there are two variables. There's the date of the death and date of -- validated date of death switch. Anybody want me to talk about that? It's pretty well-explained in the documentation, but for every person who dies, there will be a date of death. Okay? Then there's a variable called the date of death indicator, which is a yes-no switch. No, it's a V-missing switch, sorry. If it has a V, it means the date of death has been validated. It doesn't mean the ones that don't have a V aren't valid and they're bad. It means that the date is wrong. Okay? If it's a validated -- if it has a V, then the date is precise. If it doesn't have a V, the date of death is set to the last day of the month. Okay? So if you're doing [inaudible] studies, or, you know, life tables, things like that, things may get in the way. Okay, you've got some imprecise information. You might want to throw these people out. Or you might set the date to the middle of the month, you know, rather than taking the last of month, take the first -- I mean, take the 15th, and that might be a better approximation. But that's the long answer to your question, Jared [assumed spelling]. So you know who died, for sure. And for 96% of the people, the date of death is validated and it is there and it is good. For right now, you can get the -- you can get ISD ten based, underlying cause of death. And as far as I know, it's only underlying cause of death, it's not multiple cause of death data. Those of you who are date of death freaks know there's underlying cause of death or multiple cause of death data. I believe it's only underlying cause. Which isn't good enough for government work. Okay, anything else about death dates before we get back on track here?

Okay, so the Beneficiary Summary File is in many respects just like the Denominator file, or you might think of it as Denominator File Plus. It has information on all beneficiaries who were ever enrolled in that calendar year. Okay? So people who came in in August are in there. People who died in February. It's everybody, okay? Its annual. They're calendar year files. Sometimes you may therefore need maybe two years of data if you're, you know, doing some things that are -- yeah, I don't know why. Forget that. Its annual. And it's what CMS knows on December 31st of 20xx. So if someone's moved, let's say three times, it's the residence information on December 31st -- that third place. And the file is then created around the summer of the next year. It used to be -- when it was just the Denominator file, it was created in March of the following year, but now it's taking a little bit longer.

For the Part D denominator information, which I haven't told you much of

anything about yet except that it exists, that information currently is added a year later -- or six or nine months later. So let me go through this. So, you've got end of the year data. So, 12/31/2010. By the next summer, you've got all the traditional denominator information. Then by the next end of winter, you have the new Part D extra information. Okay? So for some of you who are working very quickly, you might get -- if you asked for the Part -- sorry, for the Beneficiary Summary File for 2012, you might get it in two different segments. They'll send you the first part when it's ready and then they'll send you the second part later. If you ask for older denominator data for which both sections have already been compiled, then you get the whole thing. But it does enter into some decisions we'll talk about later as to how you might identify a low-income people. Recently, we've begun to hear a rumor that things may get speeded up and that the 2013 data, both the first and the second segments, may be available during the time line of that first segment. So I can explain why, but don't worry. They're just getting through the data quickly -- more quicker. So right now though, there is that slight difference in timing.

Okay, so real quickly, because most of you know all this, what are the traditional demographic enrollment data or information in this Bene Summary File segment? So there's the unique identifier. Everybody cool on that? You no longer get the health insurance claim number. You get a Unique Identifier.
^M00:10:00
Encrypted for your study, your study only. Don't cheat and go down the hall and try to merge with somebody else, because it will go to Hell. You know, it's unique to your study -- your DUA.

Then you've got the demographics. Then you've got the enrollment information in terms of Part A, B, and Medicare Advantage, the death information we've just talked about, and then whether the beneficiary is in the 5% random sample or in the continuous 5% enhanced random sample. Questions on that? Things to -- the random sample is based on the last two digits of the Social Security number. Anybody know what the numbers are? 05 -- no, you don't know. You don't care. Okay. And the question was always because people can change their SSN if -- I'm sorry. People don't change their SSN. The SSN under which they are enrolled may change. Going into too much detail. I apologize, but when my mother did -- my father died, she had been previously under my father's SSN and he died. My mother got her benefits for Social Security based on her work history, so her number changed. And she could have gone from being in to being out. And the enhanced 5% sample is if you're in ever, you're always in. So if you're doing longitudinal studies, you want the enhanced 5% sample, and there's a way to know that and you should indicate that always in working up your data request.

Okay, so repeating some of the stuff in more detail, the encrypted ID, date of birth and age as of 12-31. Gender is there. Race-ethnicity is a one-column field. We'll talk about that variable in a few minutes, and then place of residence at the state, county, and zip code level. If this is too fast, tell me to slow down, but I think you all know this. So then,

the comment about the agreement between demographic information and claims data. Now, the question is do we have any income, education or other such information? We do not. We do not. We have some proxies, like Medicaid Buy-In is a variable, and the new state-reported dual-eligible status variable, but we do not know income or education.

So, something I generally say back in the CMS 101 lecture, that as the claims come through -- okay, so a person has some services, claims are submitted, they come through and are processed in the CMS processing system -- the demographic information in the Beneficiary Summary File at the moment that claim is processed is what's put on that claim. Okay? So again, if I'm living in Texas now in June -- May, okay and I get some treatment, it will have Texas as my residence. I move to Minnesota come fall and I get services, it'll have Minnesota as my place of residence. Okay, so there is -- in the claims, okay? The denominator -- sorry, the Beneficiary Summary File will have where I'm residing at the end of the year. Okay? People generally haven't worried about that kind of stuff, okay, so that if you're doing a denominator let's say and you're looking to get all the Minnesota residents and then you want all the claims for Minnesota residents, right, you're going to have some claims for services they could have received out of state. Well, first of all, they could have gone to Iowa -- but they could have had while they were Texas residents. And there's no way to get around that unless you, you know -- nah, it would be too hard, because you know, you don't know whether the person went to Texas to get services or was a resident in Texas before. I think you're just stuck with saying, this is my denominator estimate for the state of Minnesota or for whatever geographic area I'm studying, again, based on the 12-31 location. So, then this is all a lot of words just to repeat that. And maybe I shouldn't bother to do it.

So, the last point though is something that I do forget about and Barb has to remind me to keep it in here, is that the gender and date of birth variables in the PDE file come from a different source, and therefore we suggest that when you're working with the Part D data that you go back to the Beneficiary Summary File to grab all the demographic information. Otherwise, you might have some inconsistencies and you're going to start worrying about people changing gender or that the date of birth which one might I use, go with what's in the Beneficiary Summary File.

Okay, residency, state, county, zip as I mentioned before. We all know the caveat that this is the residence which is the mailing address for official correspondence with CMS and SSA, so you know, those things could be sent to a custodian, or a cousin, or a daughter, or son. And so you do get some error there. And again, as I said, it's based on the end of the year information.

So, here's the traditional Medicare enrollment information and someone was asking about the 12 months demification. Michelle, that was you. So that first of all we have variables, and these are -- the variables on this slide for each month indicating something. Okay? We tend to work with those as opposed to the variables that give the count, but there are cases, and your example was a good one, where you're going right to the information,

how many months was this person enrolled? If they've got to be 12, then you don't care if it's only 12. I mean, if it were 10, you might worry is it one to 10, or two to 11, or three to 12. But, we like the precision here -- or the redundancy I guess. So you can know whether they were enrolled in Medicare Part A, Medicare Part B, and now here we go.

There's the state buy-in variable. So the definition of that is did a state pay Medicare for the beneficiary's Part B coverage in the traditional Medicaid program? So, Medicaid, big letters, or in one of these Medicare savings programs that I mentioned earlier, QI, QMB, or SLMB. So there's the state buy-in variable by month and then finally there's the information on whether or not they're enrolled in a Medicare Advantage plan also by month. So here we go. These three key ones, and then the QDWIs. So QMB, SLMB, and QI. Now, these all relate to this state buy-in variable. And again, many of you have already worked through this in CMS 101. So the state buy-in variable has two clusters of variables after the not-entitled. These are Part A only, Part B only, Part A and B, and then with the caveat down here for these values of state buy-in. So, it allows you to figure out who has state buy-in and what coverage they have. And if you look at the detail, you can see that the great majority of people, of course, are in both Part A and Part B, okay? 75%. There are some people, and this is a good number, about 6% nationally, only have Part A, and very, very few people will only have Part B, whether it's buying it on their own or whether the state has purchased it. So, the first thing to recognize again is, most people have Part A and Part B. And what I'm going to say several times and then have to recant is that when you do a study, we at ResDAC strongly advocate that when you are selecting your study populations, you only work with the state buy-in variables with the values of three and C -- those people who have both Part A and B coverage.

And if you want to know why, it's in this slide here which is an example of utilization. So hospitalization rates by type of coverage, either A only or A plus B. And remember, hospitalization is an A service, right? So if you have A, you ought to be hospitalized just like people who have A and B. well, it's not true, okay? So, if you only have A coverage, your rates of hospitalization within each of these age groups is much, much lower than if you have A and B. And so, we've always recommended that when you're doing studies of people where the outcome is particularly an A service, that you actually require that they have both A and B. Sometimes, if you're just looking at a B service like immunizations -- eh, you might get away with just taking people with B coverage. Things like that. But by and large, we're strongly recommending that you select people who, in the state buy-in variable categories, have these threes and Cs. And again, in about 20 minutes,
^M00:20:00
I'm going to change this. Okay. It's a little embarrassing. I'll walk you through it, but only yesterday about 5:00, we realized that I've been making a mistake for about 15 years. Okay? But I slept very well last night. I didn't lose any sleep over it. But I'm going to try to explain. But the ResDAC position up until 10 minutes ago was, you know, just take the B -- the Cs and the threes. Okay, the threes and the Cs. Now, any questions about that? I mean, do you care? You care in the sense that

as -- I'm going to keep on the three and C business -- that there really is a difference in utilization if you were to include the A people. And we stumbled across this actually in work we were doing in Puerto Rico, where 25% of the population has A only, because a lot of them work at the base [inaudible]. And so -- or have naval experience and a -- so they have Part A, enough earnings, but they have Social -- a program down there called Triple S that covers them. And so they only have Part A. They don't sign up for Part B. And we found that in the District of Columbia, for example, 10% only have Part A. And that's because of the federal employee health benefits. But in other states, it's about 5 to 6% as the number here shows. It's not huge, but it is a percentage that again is there in your study and you might want to -- or we think you should get rid of them. Why does this say here -- this is June 2008. How come 5% almost were not entitled -- or not in the data system if you will -- not indicated as being...
>> [Inaudible response].
>> Marshall McBean: No, it's just that it's June. Okay? Okay? So, some of the people were dead and some of the people hadn't enrolled yet.

So what's in this Beneficiary Summary File that's new because of Part D? And the first is this RTI race variable. And then there's information about enrollment for different things like the low-income subsidy and things like that. And these -- this information is also presented in two different ways. Some of them -- some of it is presented with 12 variables -- one for each month. And then some again, have this summary variable of zero to 12. So the race variable -- some of you are interested in that. A brief history -- originally it was coded as White, Black, Other, and Unknown. Then it was expanded to include Asian, Hispanics, and Native Americans. And the problem was that when his was done, Medicare -- one problem was that Medicare kept the variable to a one-column variable. So when -- if you look at a lot of federal data systems, there will be two race-ethnicity variables, one called Hispanic yes-no, and then you choose another one that chooses race. CMS -- I was there, we failed to convince anybody, but it was an 80 column card. They only had one position. When you think about it now, you expand records like crazy. Back then, they said no, we only have one column. It's got to all fit into one column. So the Medicare data now is hamstrung if you will, and has the race variable as this one column variable.

But there was another problem too is that when you look at the variable and the epidemiologists will love this -- this is looking at the sensitivity and specificity in positive predictive value, where people have done studies looking at the traditional race variable. So this is information from the enrollment database. You can be thinking denominator file information -- so information held by CMS and then some variable called self-race. That was the self-reported race from the surveys called the CAHPS surveys that worked on back in 2000 and 2002. And you can see that the sensitivity for Hispanics and Asians was really low. Okay. You also can see that the specificity for Whites was pretty darn low. So if you're quick at it, you know this was a misclassification. Okay, all of the Hispanics who were not identified and the Asians were being called White primarily. But it wasn't very good. So if you were doing studies on race, and I was doing them at the time, and this didn't make you feel very good,

and CMS went through lots of work to try to enhance the information they had and it didn't work. And so finally, what RTI did -- Research Triangle Institute, is they developed an algorithm looking at the language preference for the materials that people got from CMS and Social Security. They looked at the information from the Indian Health Service. They looked at the state of residence, particularly Puerto Rico and Hawaii. And then they took the list of surnames -- there are surnames for Hispanics and for Asians that are held by the US Census Bureau. So they took all this information, created an algorithm, and then identified people and again, did a comparison between their new race, so I call it new race -- or maybe I don't here. RTI race -- I changed it to RTI race and then in blue this is the old information that CMS had, compared with again the CAHPS survey information. And you see this marked improvement, so that the sensitivity for Hispanics and Asians has gone up dramatically as has the positive predictive value so that now we would recommend that you use the RTI race variable rather than the traditional CMS variable if you're studying race-ethnicity. Okay. Another trick is, use the variable for the most recent data set that you have, or year that you have. Okay, because sometimes these will change over time. And so if you've got a five-year study, take the race in the last year and overwrite all of the other years, so they're -- you have the most current information.

This set of pie charts shows the impact where you can see that the biggest loss is in Whites. The decrease is in the White beneficiaries. A slight decrease in the percentage of beneficiaries who are indicated as Black, but then a major increase, up to 7.8% for Hispanics and almost 2.5% for Asian-Pacific Islanders. So a lot more accuracy in race identification if you use the RTI race code. Comments, questions?

Okay, now for the fun stuff. Now, Part D enrollment information. Part D enrollment-related information. First of all, Encrypted Plan Contract ID. Okay? This is very important. For some reason, I always stumble over the name, so if I stumble in the next four or five minutes, it's for whatever block I have here. But it's the Encrypted Plan Contract ID. Very important. Occurs 12 times. Obviously unique. Shouldn't say obviously, but it's unique to every plan and it will tell you what type of plan someone is in. So again, MA-PD versus PDP. Then there's the cost share group variable, which is the one that's all about low-income subsidy, then another one that we'll recommend that a lot of people use if you're studying dual-eligibles is the state-reported dual-eligible status code. Gives you lots of information on dual-eligibility. And then there [cough] -- excuse me [cough], there's a summary statistic for that. And then finally, two other variables that as far as we know have been of little use -- whether or not the person has creditable coverage and whether or not they're in a retiree drug subsidy program. So here's this Encrypted Plan Contract ID. And the, in a sense, funny thing about it is that it's a five-column variable, because it's an ID for the whole plan. But the one that's most important at the denominator level, if you will, is that it identifies whether you are in an MA-PD plan, or a PDP plan, or something else. So the PDP -- sorry, MA-PD plans are designated with an H [clears throat] -- excuse me, or an R. Okay? And as it says here, we have PDE data. Then the Ss are all the fee for service or prescription drug plan plans. And

again, we also have PDE data. If they're not in Part D, obviously we have no data. And then there are these employee-sponsored plans, and our experience is that while you have the data, you don't have any plan characteristics data, and so we haven't used it, and Kyoungrae is much more an expert in this -- in the rationale than I am, but generally if you're studying people with A and B and then D, it's very simple. Go get the Ss.
^M00:30:00
If you want to include the MA-PD people in your study, grab the Hs and the Rs. Okay? But this is a crucial variable for identifying what kind of plan someone is in. And so here for 2008, the month of June, the distribution and it pretty much fits those pie charts that I showed earlier.

Okay. Now this extra help, or LIS subsidy. So, there is this benefit for people who have low income where they can get help paying the Medicare drug plan's monthly premium, any deductible, coinsurance, or co-payments. Also within the LIS subsidy is there's no coverage gap liability and no late enrollment penalty. And as I've said a million times, lots of effort to get people enrolled. And the final point here, that this information is known for all beneficiaries, not just for Part D beneficiaries. Blah. Blah. Blah. Blah. Blah. I've said all this before. Yeah. Here and in the back of your Tab C is the detail of the information that allows you to qualify, so a beneficiary group, and then on the right-hand side, the benefits that you get. For people who are what they call here full low-income subsidy or partial low-income subsidy. Another term that's used is deemed and non-deemed. And you might want to write this down, because people who meet these criteria, okay, they're in -- people who have both Medicare and Medicaid, all right? And [inaudible] they reside in a long-term care facility. Then other people who have full Medicare and Medicaid, or are in one of the Medicare Savings Programs -- so the QMB, SLMB, and QI. And then there's people who largely are receiving SSI -- Supplemental Social Insurance. Not SSDI, which is Disability, but SSI. These are really -- these are poor people. This is a minimal subsidy for really -- or income payment for really poor people. These people automatically are enrolled in the LIS, okay? And so they're deemed LIS. And they get pretty good benefits. I mean, they have no premium. They have no annual deductible. And for those who are in nursing homes, there's no co-pays. And the co-pays for people who are in these other income and resource categories -- they're pretty low. So they're in the one to six dollar range depending on the type of drug. So these are people who are in LIS, who are automatically enrolled. There's another group for which the person must enroll. And so these people have higher income and asset amounts. They don't get quite as much. They actually have to pay a deductible of $66 and their co-payments are about the same as the high group up here. But it's still a pretty good situation. But these people must apply. And so, again that's the LIS story in kind of a nutshell.

This is the distribution of LIS beneficiaries by state, going from about 30% in some of the cross-hatched, upper-mid -- or the diagonal hatch to upper-Midwest states, and about 60% in the Tennessee Valley area. So there is variation in LIS enrollment by state. And here's the slide I showed earlier just to remind you that if you're looking at PDP beneficiaries, almost half of them are in LIS. The question is, where do the people fall?

[Background talking] I would say about -- let me -- I'm trying to think and talk at the same time, which is always hard. This is everybody, so we haven't excluded the end-stage renal, the [inaudible] sclerosis, or the disabled. Obviously, in some studies you would do some of those exclusions, particularly end-stage renal, or you might not want to study the disabled. I haven't looked at this -- let's say the two big categories, disabled versus 65 and older. My guess is you're going to have a bigger segment -- so more than 40% are going to be in PDP. There will be a bigger segment of that 40% that's LIS, and there aren't going to be that many who are employed, so these, you know, other coverage are going to go down. Okay, that's my guess. But it's a good question. Yeah, good question. Anyone else want to comment on that?

Okay, so now here's this variable. Now I've talked about it, you know, a number of times. And so, if you look at the values for the variable, it has -- [clears throat] excuse me -- these eight categories, okay? And here's this word deemed again. And then here's these other LIS categories. Lots of words, kind of going from income level most low to income and assets most high, if you will. And if you actually look at the distribution, then you can see what it is, that people who are not in LIS -- let me get this right. So this is -- this is total Medicare beneficiaries. So that if you're in LIS but not in Part D, about 34% of all beneficiaries, and then here are all the LIS different categories. And this includes both MA-PD and PDP. So most of the people who are in LIS -- if you do the summaries of these here compared to these, most of the people in LIS are at the low end, and it's best seen in this pie chart perhaps where you can see that, starting with the deemed -- this is again the lowest income group, those in nursing homes, this group, this group, and this group, and the lower three all the deemed make up about 85% of the LIS. And not that I expect you guys to be thinking the way I've always been thinking is that, I always thought that with this new information, I could then figure out, you know, who are, you know, the really poor, and then the little less poor, and the less poor. And the numbers here are kind of small, you know. And so, maybe you do deemed versus other LIS and then non-LIS. I know I'm jumping around here a little bit, and you guys haven't thought about this as much as I do -- or have, but we were kind of hoping that -- or I was hoping that because of this new information and this new variable that seemed really precise, that we would be able to tease out gradations of low income. But again, it's mainly the deemed people that are in this LIS program. So, more than most of you came today to hear about, but did it make sense everybody? And if you're interested in low-income, this is the place to get more precision, basically.

Okay, now here's another cool one -- a little bit different. Almost redundant in fact, but this is state-reported dual-eligible status code. Now this one is also pretty exciting, to me anyway and hopefully to some of you, because in the past, when we had the state buy-in variable, that lumped together, as you may remember, the traditional Medicaid plus these three Medicare Savings Programs. SLMB -- or QMB, SLMB, and QI. Now, these are broken apart in this variable so that, although it's a little hard to see here when you look at it for the first time, number eight here is the traditional Medicaid. Okay? They're not at the top. They're kind of

here at the bottom, okay? So people with the value of eight are dual-eligibles, read Medicare and Medicaid, but not the QMB, SLMB, QDWI, and QI. Those are up here in the variable. And so, again for people who are interested in studying the duals, this is the first time that you can accurately identify what are known as the full-duals and the partial-duals. Okay, let me just -- does anybody care? And I don't mean that in a defensive way. How many people in here have worried about that, because the people studying the duals have worried about that incessantly.
^M00:40:00
Who are the full-duals -- the people who are getting all the Medicaid benefits, and then the partial-duals, and is there a difference in that it's -- for that research population? And again, given that so many of these people are deemed and they're in Medicaid, you know, it becomes more of a kind of interesting thing perhaps for people doing Part D studies. One is low-income subsidy. Okay? This just came out of left-field. Okay. It's brand new with Part D and read about it, you know, and study that. Then, go home, take a nap [laughter]. Okay. Well, seriously, because they are so similar that it's easy to get them co-mingled. Then go talk to someone who has been worrying about dual-eligibles, okay, for all kinds of reasons. Okay, mostly for policy reasons I would say. And then say, is it important to know who are full-duals or partial-duals, and they'll talk to you for 20 minutes. Okay? [Laughter] The rest of us will say, huh? Partial, full, you know? But the full-duals get all the Medicaid beneficiaries -- Medicaid benefits allowed by that state, okay? Then the partials don't. Okay? They're different. This is again, for those people who -- and its 18% more or less of the Medicare beneficiaries are duals, so it's not a small population, and we've never known who were the full-duals -- or we the researchers, and who are the partials. And so this is actually very exciting. Because it now, enables you to know this and here are the variables. Or here is the variable and its values. Here's the distribution. Again, most people not being in Medicaid, okay, or not enrolled for that month. So around that 80% I implied. Or it's actually -- and then here's the other. So most of the people who are in Medicaid programs -- I'm sorry, that's not true. Most of them are QMBs, but the full Medicaid duals are an important number.

And let me just summarize that real quickly. Oh, I was going to say. Yeah, so who are the duals? I mean, I kind of lost the flow here in trying to get you into the discussion, but it is again -- what I was trying to draw out is there are some people who are full-dual, some people who are partial, and some people worry about that. And you've heard that a couple of times. So then you have, who are the full in blue or black, who are the partials, and then there's a group that I like to call to be decided, because CMS can't quite figure out what to do with them yet. But if you want to know the partials, here they are. If you want to know the fulls, here they are. You can find them. And then there's this other group which we looked up because we were trying to make this slide as comprehensive as possible. There are things like Section 1099 programs in Vermont, Nebraska, and Illinois. These are very special people that have some drug coverage back when Section 1099 programs, you know, were and still continue. You know, most of us aren't going to care, and you can stick them in the partial-duals if you want. But some people don't and some people do. Again, here's some

percentages that you can look at. And I have summary statistics, so of those who are in the Medicaid category if you will, 74% are full-duals, 21% are partial and there's a few that are undecided. So this variable should be useful, again for people interested in that category and use it. And then -- excuse me, there's the monthly variable for state-reported dual -- I'm sorry, the summary of the monthly state-reported dual-eligible status codes one through nine.

Okay, and my final comments here are about the linking variables, which are very important when you start looking at plan characteristics and other characteristics files, and they are the Encrypted Contract ID, Encrypted Plan Benefit Package ID, and the Encrypted Segment ID, and Kyoungrae will talk more about these.