Transcript: 2018 Census technical seminar

Video, 1:34:28

See video in 2018 Census: How we combined administrative data and census forms data to create the census dataset.

Visual: Kathy Connolly, Christine Bycroft, and Adele Quinn presenting the 2018 Census technical seminar. The presentation includes a number of text slides, graphs, and flow diagrams.

Audio:

Hi. Welcome to our 2018 Census Technical Seminar. It's great to have you along today.

I'm Kathy Connolly, I'm the GM census here at Stats NZ - I am relatively new to the census, but I am a career statistician. So not new to Statistics New Zealand, or the delivery of statistics.

They won't be on screen, but will be shortly. I'd like to introduce you to my co presenters. The first is Christine Bycroft, who is one of our Principal Statisticians and a true census expert.

Christine has been the technical lead on our census transformation, and her leadership has been integral in developing the methods that we're going to talk you through today. Christine's expertise is also well recognised internationally.

The other person in our presentation team today is Adele Quinn, who is our Manager of our Data Analytics Team. Adele has a wealth of experience, having worked in censuses in various roles since 1996 - I did get permission to say 1996 - as well as stints on census transformation in our population statistics area, and supporting the electoral representation commission twice.

So as I see it, it is great to finally be talking to you all. I know that some of you will have been frustrated to not have heard from us before now. But we have been very busy, working through the methods to fill the gaps left by the low response to the 2018 census.

Hopefully, most of you will have caught up with our media release on Monday, when our Government Statistician announced that we've now created a data set that meets Statistics New Zealand's quality criteria for population structure information.

Over the next hour and a half or so, we will run through some of the more detailed information about the methods that we've developed.

So what are you going to see? I'm going to do a really quick update, including where we're at with the first release. And then I'm going to hand it over to Christine, who will take the bulk of the seminar, and we'll run through the methods, and why you can believe that what we've done here has improved the census data set.

Adele will take us through the quality work that we're doing, and then I'll finish off with what's happening with 2023 and where to from there. We are happy to take questions at the end, and we'll try and leave a good portion of time for that. So you can just send your questions through as you go.

If you are needing clarification during the presentation, feel free to submit a question and our team will attempt to answer it if they can.

So, where are we at? Before we get into the detail of our new methods, I just want to acknowledge that 2018 didn't go to plan. While some things worked well and as planned, we know that other things didn't.

We didn't make it easy enough for everyone to respond. Mail out didn't work for everyone and the process of getting paper forms could be frustrating. And not seeing face to face, people, did not sit well with some New Zealanders. While we're being careful not to pre-empt the finding of an independent review of the census operation, we know that some themes are emerging in feedback and comments about the 2018 operation. Certainly for some areas not being contacted early enough, either in getting people what they needed to complete the census, or making contact with them, and follow up, is something that we're very aware of. We do understand that the length of time after people we made contact with after census day, did have a big impact on the number of people that responded. And I'm sure the independent review will identify other areas that can be improved for the next census.

Over the past nine months or so, we've been very busy developing and implementing the methods that we're going to talk you through today. This has been made possible by the work that we've been doing over the last four or so years on our longer-term goal to move to an administrative based census supported by surveying.

We've also worked with our international colleagues at National Statistical Offices around the World, leveraging of methods that they have developed, and their expertise, to provide input and review to our own work.

Our key goal today is that you, I think, walk away, with an understanding of the methods that we're using to fill the gaps and what we know about the impact. We want to leave you confident that we will be transparent about the sources of data used for each variable, and the quality of the variables. Our work is, however, still underway, so much of the detail of the information about the variables we're sharing today is not finalised, so please do treat it as indicative.

We know that some of you will have specific interests in particular topics, but I'm afraid we won't be able to answer detailed questions about specific variables, as we are still evaluating. I mentioned that independent review earlier, and this is a really important thing for us. One of the things that we do know is that we need to get a very clear understanding of what worked and what didn't work for 2018.

Liz MacPherson, our Government Statistician, is seeking robust independent advice, and she's asked Murray Jack, who's a management consultant, and Connie Graziadei, who is ex-Deputy Head at Stats Canada, and has run a number of censuses, to consider the design, implementation and operation of the census. And this review is on track to be released by July.

Another very important mechanism that we've got to help us is our external data quality panel. So we've pulled together this panel of experts to provide both advice on methodology as we've

developed it, and to provide an independent view of the data quality on the census data set, which will be released at the same time that we release the data.

So, the 23rd of September is when you can expect the data. And this will include usually resident population counts, down to SA2 level, so that's the sort of new equivalent of an area unit. It'll also include the number of general Māori electorates, and the general Māori electoral population counts. We're planning on having customised data available shortly after the first release, and we'll update customers with a more definitive date once we know. We are certainly very happy to take early requests. As is usual, subsequent releases for the census will flow through to June 2020. And our population estimates, both the census coverage estimate from the post enumeration survey, and the estimated resident population, will be released by March 2020, and then demographic projections will flow through to mid-2020.

I'm now going to hand over to Christine Bycroft, who will talk you through our methods.

Thank you, Kathy. So I'm going to start with a brief introduction about some concepts. Most of the focus of today's on the methods about how we added administrative records to the census file, it'll probably take around 40 minutes, 45 minutes, to go through that. And then we'll talk about where the information about the variables, the characteristics, comes from, and it will take perhaps another 30 minutes.

So I just want to start with some context from a rights based perspective. So we have a right to privacy, as laid out in the Principles and the Privacy Act. And we have a right to live in an informed society. And official statistics plays a crucial role in providing the public with information that they can trust. And the census is especially important in providing a picture of society, a key underpinning of our democracy.

There is a natural tension between privacy, and providing information that is collected directly from individuals. This tension has already been carefully considered when we developed the census content and questionnaire, as it is for all Stats New Zealand surveys. The use of administrative data respects that same balance between census content and the public good that comes from living in an informed society.

Stats New Zealand's role is supported by legislation through the Statistics Act, which is compatible with the Privacy Act. The primary reason for this is that the Statistics Act requires that the same level of confidentiality is applied to statistics or research, as the Privacy Act requires. Many of the principles of the Privacy Act, such as those related to the purpose of collection, contain exemptions for when information is used for statistical research purposes.

The Treaty of Waitangi provides a unique statement of human rights which includes universal human rights for all New Zealanders such as privacy. The treaty also specifically acknowledges indigenous rights. So here I want to also acknowledge that we're still learning and evolving our understanding of Māori data sovereignty and its application, and what privacy means from a Māori perspective. Part of this evolution is developing better governance from a Te Ao Māori perspective, and better processes in perspective tikanga Māori, to support our decisions about data at an operational level.

So first of all, what is a census? The tagline is that census aims to count everyone once, only once, and in the right place. By doing this, census is able to produce very detailed statistics for small areas and small population groups. And the census provides demographic counts, social and economic characteristics for people, for dwellings, and for households and families. Today we're going to concentrate on the people aspect. We won't be including any information about dwellings, and we won't be providing information about households and families.

One of the interesting things about census taking is how it's been evolving over the last decade or so across the world. What used to be the traditional census, where everybody was asked to fill in a form, is now called a full field enumeration census. And this is what we set out to do in 2018.

In some countries, particularly by the Nordic countries, they introduced what they call a register based census, where they only use administrative or register sources that they have available. In between, some countries have combined those two approaches. So their census is a mixture of administrative sources and field collection. And in fact this is what we've ended up doing in New Zealand in 2018. Wasn't what we planned to do, but what we're producing is actually a combined census.

So, a technical term now, coverage. Which is about how many people the census should count. The census aims to count the people who are in New Zealand, on census night. So people who live in New Zealand, but they're temporarily overseas on census night, are not part of the census population. And what census actually counts is in that census count. And the difference we call the net census undercount. This is estimated by the post enumeration survey, or our, we call it PES, and clearly we haven't run that yet, so for 2018 we don't know what the actual population should be, but we have that for previous censuses. Our census count has been made up of responses from census forms, and also, we count some non respondents to the census. What we've done in 2013 and previous censuses is have a census data set that looks rather like this.

So, most of the people counts are made up of real people that we counted through these census forms, but we also have some what we call unit imputation. It's a statistical process that counts people where we have enough information, that they should have been counted, but they didn't fill in the form. And the characteristics of people, if they filled in census forms, then we had census responses. There's a little bit of item imputation, so a statistical estimate of what a value should have been, but most variables are missing in 2013 and previous census data sets, if a person didn't answer the form. For an imputed record, there's a little bit of item imputation for age and sex, and usual residence, and otherwise the variables are missing. So that's what people have been used to seeing, in their census data set.

2018 is going to be different in a couple of key ways. We still have the census forms, and that's how most people are counted. But there is no unit imputation. We will be using administrative sources to count people that we don't have census forms for. So that's the big difference. We also have a different approach to filling in missing information. And we had planned to do this, because, we wanted to improve the census outputs. So, we will be using historic census data, so looking back at the 2013 census, using administrative sources, and a wider extent of item imputation for the characteristics that are missing. There’s label's missing from this graph, I'm sorry.

So, we don't know the true population in 2018, but when we've been doing our work and developing our methods, we needed to have a reasonable idea about what the population actually was. So we have four different benchmark populations. The top one on the left-hand side is our 2013 estimated resident population that's been brought forward. That's what people have been using for their official statistics on population. The next one, to that with the red circle, is a new estimate that we've been developing based on a new migration measure. So rather than using people's' intentions to say how long they have, how long they intend to stay or leave the country, the new migration methods are using actual measures of length of stay in New Zealand, and you can see that there's been some reduction in the population size due to that. The other circled label there is called DSE, which is short for Dual System Estimation. That's a methodology that's used to estimate who's missing from a population, when you combine two sources. It's a methodology that's used for the PES, when we match that to the census.

But here, rather than matching the PES coverage survey to the census, we're matching our administrative population. So that DSE is a new and innovative way of estimating the population combining census responses and administrative data set, which we will talk to you about in a minute. The good thing is that while the DSE and the new 2013 based ERP are completely independent methods for producing a population count, they've actually come out very close. So these are the numbers that we're going to show you today. They relate back to the release that a government statistician provided last week. So at the top is what we aimed to count, our best estimate given the information that we have to date, that we expect there is around 4,670,000 people in New Zealand. From the census forms, we are counting 4,175,000 people, which are made up of nearly four million people who provided individual forms, and another 200,000 people who enlisted on a household form, but we don't have the full individual information for them. And, over 500,000 people are being counted from administrative sources. So our census file is made up of 89% people from census forms, and 11% of people from these administrative enumerations. So the administrative data is playing a really important role in our 2018 census data set.

Now the difference between what we've achieved in that census count, so 4,700,000, and what we think is about what we expect, is 58,000 or 1.2%. So that's an indicative coverage gap for 2018, and we can compare that to the undercount in 2013, which was at 2.4% net census undercount, plus or minus half a percent. So, I'm going to step quite slowly through how we've added administrative records to the census file. I'm going to start by explaining the administrative data sources themselves, then talk about the methods, the framework that we've used to combine census forms and administrative data, and then give you some rather interesting patterns that we've used to assess where we've got to. So where does the admin data come from? Basically it comes from Stats New Zealand's integrated data infrastructure, or the IDI. This is a research database where data from across many different sources, across government, have been linked together. It also includes Stats New Zealand's household surveys, and the 2013 census. It's a structured data set. So, those blue data sources around the outside are all linked to what we call the IDI spine. The IDI spine's a big long list of people, there's around 10 million records in the IDI spine. It's trying to represent people who've ever lived in New Zealand. And we've taken a demographic approach to build that spine. So you can see that the circle on the left hand side. So, people into the New Zealand population, because they were born here. So we have birth registrations from 1920. People can also enter the population on a visa, if migrants come to live here, and we have electronic records for visa

applications since 1997. So there's still some gaps from those two data sources. And we use tax registrations for people that won't be included in either of those ways. And we get what appears to be a very complete record, or a list of people who've ever been living in New Zealand, and that means that when we link other data sources to them, most of those people in the data sources are in that spine. But the spine is not very helpful if you want to know who's living in New Zealand at a given point in time. And to do that, we've been developing in our census transformation work what we've called an IDI, estimated resident population. So, it's an administrative population, using the linked admin data and the IDI. So we start with people in the IDI spine. And then, we include people who have had some activity in administrative data sources. So, have you paid tax? Have they been to the doctor? Enrolled in education, have an ACC claim? If that's happened within the last two years, then you're likely to be living in New Zealand. Of course we want a population at a particular point in time, so we take out people who may have died before that date, or those who have migrated overseas before that reference date. So that's in general how we do this. That admin population is available in the IDI for June the 30th each year.

For our 2018 census application, we're using what's known as the September 2018 refresh in the IDI. And I just want to talk through a few points about why that's a good administrative population for us to use. First of all, the data sources that we are using from the spine mean that we're very sure that we have real people being included. So, birth registrations, tax registrations, and visa applications, they're all high integrity data sources that we can trust. The deaths are able to be removed because we have death registrations from DIA. And we are able to link to the external migration records to take out people who've left the country. We've also linked the 2018 census to the IDI spine. And the circled data sources in red are particularly important for the activity indicators, but also for obtaining address ID, which we use for our geographic location, and the ethnic variables. We know a lot about the IDI ERP population, we've been looking at it for a while. And it is a good approximation of the New Zealand resident population. We've done detailed examination of time series, from 2006 to 2016, comparing the administrative population against the official statistics. Looking at details by age and sex, geographies, and ethnicity. There's a link there if you'd like to see that, there's papers in the actual data series to explore. We've just put up the graph there that shows the comparison between the IDI ERP and 2013, and the official resident population, and you can see that they map pretty closely. There's little gaps for the males in the younger adult ages, which we'll see again. So those are the strengths of the IDI ERP. But we're also well aware of the limitations of our admin population. We know that it doesn't meet all the accuracy requirements for producing official statistics. And most of our methodology that we're using, to use the administrative data, are about managing the limitations that we're aware of. So we know that the admin population does not include everybody. Some people are missed. And it will include some people who shouldn't be there. There are some marked differences in the age sex structure for younger adults, especially males that you can see in the previous picture. Geographic location is very important for census. And from the administrative data, the address information is good for larger geographies, such as territorial authority and Auckland local boards. But the accuracy does decrease at smaller geographies. And when we try to place administrative people into a household, this is where we have real problems. Around half of the administrative households that we've created, and when we compared them with the 2013 census, about half of them had the same household membership as the census. So, that's

the administrative data sources, what did we do with them? So, some pictures now. Some colour coded pictures.

So in the 2013 census, our census data set was made up of individual forms, those dark green ones there. That was 95% of the file. There was 1% of people listed on the household, and 4% from unit imputation. We had age, sex, and place for everybody in the census data set, and some missing data that's represented by those white squares and spaces. And the 2013 census, again we have census forms, individual forms or people listed on a household form. Again, the patterns of missing data of people who didn't answer questions, or if we don't have a full questionnaire. What else we have in 2018 census is this administrative data population. So, the admin population is bigger than the census, quite a lot bigger. It does have some variables, with some systematic gaps in some cases, because not everybody has the variables available, and then, a number of variables where there is no information from more entrenched sources. So we've two sets of data, each with their own unique patterns of missing data. And we want to combine those. To combine them, we match them together, and what we're looking to do is to use people in that blue sliver on the slide. So, some people in the administrative population who haven't been included in the census. And when we've done that, our census data set looks like this. So we have people in the census forms. Some of the information is being filled in from administrative data or the previous census, that's the bright blue and dark green boxes that have been filled in. I think we have a problem with the slide here. Oh no, there we go. The administrative records have been used to add people to the census file, and they are now bright blue. For the 2013 census, administrative data filled the variable gaps. But people that we don't have a census form for, whether that's because they're on the household listing, or because it's an administrative record, then the 2013 census and administrative sources are the main sources of variables. So that's the picture of your census data set. So now how did we do that, this is starting to get on to the methodological framework now. So first of all, we want to add those admin people to the census file, so enumerate people who should be counted as part of the census, but who don't have a response for. The second step is that we put people in private dwellings where we can. As you saw before, the administrative households, quality is not so great. And we will put people in private dwellings where we have good evidence that we're improving households. Otherwise, these administrative records will be placed in a meshblock. We do that when we are sure the person should be counted, and we have good evidence for improving small area information. Now that's another big change in the 2018 census. In previous censuses, all the census records were in dwellings. Here we have people who are placed not in a dwelling, but given a meshblock, where they usually live. So, a little bit more detail about each of those steps.

First of all, what's the eligible administrative population? So we're starting with our IDI ERP, that's people who live in New Zealand. But we have to take out people who are resident. Residents who are temporarily overseas on census night. And we can do that because we have links to the border movements data. Then we link the 2018 census to the IDI spine, so that we know who has already responded to the census. Don't want to count them twice. So we have a match rate of nearly 98%, so that gives us a very good basis for knowing who's already responded to the census, and who hasn't. We've also estimated the linkage error in that, it's a problem list at linking, so there always is a little bit of link error. So we've missed an estimate 1.4% of links, and there's less than one percent incorrect links. So that gives us a very good basis for identifying the people in that blue sliver. Then we want to put them into households, where we are improving households. The census dwelling

frame here gives us a good list of all dwellings, and we know from the census field work, which dwellings, which private dwellings have not responded to the census, where we don't have any census forms for them. So for those non responding dwellings, we use a statistical model, it's a logistic regression model, or two in fact, to predict which non responding households we can create good households for. There's a tradeoff here, between getting strictly the right people into the household, and getting the right types of households. We've looked at the patterns of numbers of adults and children. The picture there shows you the two models that we're using. There's one that predicts whether the person is likely to be at the right address in the admin data, that's on the horizontal axis. On the vertical axis, we have another model which predicts how likely the household is to include the right membership. So the probabilities are on the axes. And we need to set our red line somewhere, and those above the red line will be included in dwellings. They'll be administrative dwellings. Administrative households, sorry. And the others will not be included in the households. So we made a judgement call, and we've set that threshold at the point where the households have a 50% chance or better of having exactly the same people, or the same household type. The next step is what we do with the rest of the people in our administrative population who aren't already in the census. So we are going to put them in meshblocks. We now have to do a little bit more work to make sure that the person should be counted. So, there are two steps there. We need to remove over coverage in the admin population. So we actually take out about 120,000 people, because they are likely to be incorrectly included. And we also make a small adjustment from missing linkages, that 1.4%, so that we're not counting people twice. Again, we use a statistical model for the remaining people to predict which people are more likely to have a correct meshblock. So again, this is the... It's a logistic regression model, it's actually the same as predicting whether the person is at the correct address, but applied for a meshblock. And that's the pattern of distribution that you can see there.

Now there's a trade-off here, it's a little bit different from the previous one. We now have administrative records who we know should be part of the New Zealand resident population, and they haven't answered the census. So every record that we include will improve the national demographic distributions. However, we also know that the meshblock may not be right. So we have a tradeoff between improving national demographic distributions, and trying to provide good, even coverage patterns for small geographies. We want to avoid putting people, too many people, in a small geography, but we also want to make sure that we don't have small geographies which are really still under counted. So again we've made another judgement call, and we've set a threshold of 50% chance or better of being in the correct meshblock. So, that's as far as I'm going to go into to the statistical methodology. We will be producing a detailed documentation with all the full description of the models and so on. Now, we're going to have a look at what happened after we did all that. So who have we added from the administrative sources? So census file, we had that 4.175 million census form respondents, and we've added 162,000 people from administrative sources and been able to place them directly into dwellings. And another 357,000 people who've been added to meshblocks. And we had another 68,000 people leftover that we didn't include, because we didn't think that their meshblock was likely to be good enough. Now this is a slightly complicated picture. It's showing the age distribution by source group for males. On the right-hand side are the colour coding for the different groups. So in blue we have the age distribution of census respondents. The red is people who are added to dwellings. And you can see that we have more children, and, so

probably their parents are being added to dwellings, because typically, families with children are easier, have better household information, in the administrative data. One of the interesting things here is the green line, the people that have been added to meshblocks, and we can see that we're getting far more sort of 25 to 35 odd young people, young adults, also. Added to meshblocks. But the purple line, the people that have been left out, are overwhelmingly in the age groups from 18 to 24, and that's driven by our model formulation, because those groups, people who've recently left home, typically have poor administrative address information. That's the males. The female distribution is very similar, not quite such high peaks, but the same patterns.

Now I'm going to look at the ethnic distribution. These are a little bit complicated to follow. But again it's the same source groups. If ethnic groups were missing evenly from the census, we would see even bars for each source they come from. That's clearly not the case. And we can see here that Māori and Pacific were not well enumerated through our census forms. The proportion in blue, the blue bar is the proportion of Māori or Pacific who are in the census forms, the census respondents, and that's a lot lower than the proportion of the people that we're bringing into our administrative data. But the good thing is that we are bringing a lot of those people into the administrative data. So the administrative data is really making a big difference for the people who we know are typically difficult to count in a census. The next slides are the ethnic distributions for the remaining two level one categories, so European and Asian. And for the European you can see, there's more Europeans in the census respondents and fewer being brought through, fewer proportionally being brought through in the administrative data. And similarly for the Asian ethnic group. This is a picture of the age distribution, so it's single year of age along the bottom. The solid line is our dual system estimation, estimate, of the population distribution for males by age. The yellow bars are our 2018 census file. So it's a census form as well as the administrative records included. And you can see that the, for the children and for most of the adult ages, the census is following very closely to what we think is the actual population that the census should be counting. So that's great news. There is still a gap for young adults, and that's what we've seen in the previous slides as well, that we're not able to include the younger males in particular, but also younger women. So, and we also, I would compare that with 2013. 2013 here is the blue bars, and it's the relative difference between the 2013 census counts, and the official 2013 estimated resident population. We have adjusted here for the residents temporarily overseas. And again, the 2018 yellow bars are our comparison, approximate comparison against the 2018 dual system estimate. So, the yellow bars are a lot shorter, than the blue bars, basically. That's saying that the census in 2018 has counted, included, a lot more people across all age groups than the 2013 census was able to do. It's particularly good for children, in comparison to 2013's census. You can see, again, those ages, when young adults tend to leave home, there's some improvement on 2018, but there's still the difficult area. There's a little bit of over coverage for the oldest age groups, that may be an issue with the DSE estimate itself or some other reason, we're not sure. But the numbers are very small, and these are percentages. And finally, just look at the results comparing 2018 and 2013 across the territorial authorities and Auckland local boards. The point here is not to look at any particular territorial authority, but to note that the blue bars, the 2018 census, is quite a variable difference in coverage. Some districts were up to 8% lower than the official estimated resident population.

Some other districts were, there was almost no difference, and some were over counted. You can see that the yellow bars, for 2018, they're uniformly smaller, meaning, that the under count across

all territorial authorities in the 2018 census is smaller than we are able to achieve in the 2013 census. And there's little, very little over coverage in 2018 as well. We also did some analysis of what all of this meant for the electoral calculations. In particular we wanted to understand that meshblock threshold, which is the main way that people are included or excluded from the 2018 census file. How does that affect our New Zealand electorates? We had a contractor come and do a sensitivity analysis, and their conclusion was that the 2018 census is robust, for the purposes of determining electoral boundaries and representation. So that also gave us confidence that the 2018 census data set is a really good one for our core demographics. We had a number of other core, really important uses being examined by other groups at the moment, but those investigations are still underway. So in summary, the 2018 census population counts, the core demographics, we have a coherent statistical methodology for adding administrative records to the census file when we don't have a census form. It's based on statistical methodologies which reflect our understanding of the limitations of the administrative data. The administrative enumerations replaced unit imputation. And that is a significant quality improvement. We are adding real people, not statistical artefacts. And they do have some characteristics that come from alternative sources. And what we've seen as well, is that the administrative data does include people who are hard to count in a census field enumeration. We are including people who did not respond to the 2013 census as well as the 2018 census. So, we intend to continue with the use of administrative records in the census file in future censuses. So Stats New Zealand is now confident that we have compiled a census data set that will provide census usually resident population counts, and electoral counts of acceptable quality. I've been talking through how we've been adding administrative records to our census file, so adding people into the census file.

Now I'm going to talk through where the characteristics of people come from, using alternative sources that we hadn't used in previous censuses. So I'll start with some background, and then look at our quality assessment of the alternative sources, and then which variables use what sources. After that, I'll be handing over to Adele, to talk about our quality assessment process. So I think I mentioned already, that we had already planned to improve census quality by filling in for missing answers to questions where possible using historical census data, linking through to the 2013 census responses, and administrative sources where that was possible, and also making greater use of imputation in the 2018 census. And that approach would be known to naturally extend to the administrative enumerations. The same alternative sources are available for any census record from the administrative data because it's all linked into the IDI. So just a reminder that the 2018 census data set is made up of those census forms and the administrative sources, and now we're going to look at the different sources that we're using. So first of all, we thought we could use the 2013 census, go back to the previous answers that people gave, for variables that didn't change a lot over time. We didn't have this information when we decided to do this, but since then, we've been able to look at the consistency of responses for a person who responded in the 2013 census, and the person's response in the 2018 census. These are the variables that we're using 2013 census data for. You can see at the top that very factual variables, country of birth and Māori descent that we use for electoral, which is either a yes or no response. There's 99% agreement between those two censuses. So people are very good at providing the same answer to very effect base questions. There's another group of variables where the consistency is greater than 90%, which is still quite high. Some of that difference will reflect real world changes that we're not able to pick up in our census 2018 data set.

There's been particular concern around the smoking variables and that's a valid concern. However, some of that variability is simply that people don't always respond the same way. If you ask me to fill in the census form next week, I'll probably give you some different answers than the ones I gave before. Religious affiliation is a little lower because there's been some changes to the census question, which may have affected how people answered it. And you might think that using the secondary school or qualifications variables is a bit odd, however, for anyone who's gained a higher qualification since 2013, we have that information from the education data. So we're able to update the 2013 value if they got a new qualification. So that is our 2013 census data, and what we're using here. And then, we also are able to use some administrative data with some variables. This has been some work that we did, have been doing for our census transformation programme, that we've been looking at. Can we use the administrative data for census? We needed to understand which variables we were likely to have some good information for. So this is a summary diagram of how to present our findings from that work. We're using quality framework for this, so along the bottom is what we've called measurement area. So that's really consistency with the concept that we're trying to measure. So on the left hand side, there isn't much consistency with what we want, we can call it poor. And it progressively gets closer to what we're trying to measure.

On the vertical axis we have coverage, so that's how many people do we have this information for. So at the bottom, very few, right through to almost everybody at the top. One of the main things from this is just the variety of places that different variables end up on this graph. What we're using for the 2018 census are variables where the measurement area is small. So the variables are close to what we're trying to measure. We don't need to worry about the coverage, even if we're only able to fill in a small group of missing data with good information, that's helpful. So, this is a bit of an abstract diagram, lots of lovely colours for those who like these things. I am going back to the colours that we used earlier. It's really just to say that it's complicated. We have census form information in the bright green for all variables, except one actually. Some variables we have admin sources for, some we have 2013 census, some both, some neither. Some variables we have item imputation for. And other variables we have to leave missing 'cause we have no other sources. This is a picture of questions, or variables from the individual form, with a bracket down by what source they've got. So each bar is a particular variable, and, the green is the proportion of that subject population which has been available from the individual forms. It's ordered by percent missing, at the bottom, and then the percentage of imputation. So at the top, we have our core demographic variables. So age, sex, usual residence location, and ethnicity, where there's no missing data, there's very little imputation, and most of the data comes from either the census forms, administrative data, and, or, this 2018 census. So, that's our core demographic counts. As we go down from there, there's increasing amounts of imputation being used, because we don't have so much available from other sources. And at the bottom, we see the increasing amounts of missing data. Again, there's no imputation available here and decreasing amounts of information available from other sources. Second from the bottom, you'll see a dark green line, that's usual residence five years ago, where we have planned not to ask that question. There's a lot of address questions being asked in the census. So we did not ask that one, because we can link back to the person's 2013 census response, and then we'll know where they were five years ago. So that is the variables where we have some other sources or imputation being used for them. There's another group of variables where we don't have other ways of supplementing the information from the census forms. A lot of these are actually the

activity limitations from the new disability questions. And you can see that most of them have just under 85%, between 80 and 85% of information available from the census individual forms, and the rest is missing. And at the bottom, we see that Iwi affiliation is worse than the other variables. There's a lot of missing data for Iwi, and the government statistician has made some decisions about that that Adele will talk about later, but that's why. So now I'm going to hand over to Adele, to talk about the impact of those patterns of information for variables.

- Thanks Christine. So I'm going to finish the characteristics section and summarise the methods, and then talk to you a little bit about quality assurance and assessment. We'll talk about our quality management strategy shortly. But it's important to mention now that a key aspect of our quality management is that all variables are given a priority level. This is so that we can determine the relative importance of each variable, and ensure that we focus our effort across the census programme in the right places. Variables are rated as either priority one, two, or three. The priority one variables are listed on the right side of your screen. They include our population and dwelling counts, core demographics, so age, sex, ethnicity, Māori descent, and location. The method slides that we've shown you to date show that we have good coverage and quality for our priority one variables. Producing these variables to a high quality achieves a lot of the value of the purpose of running a census. The only priority one variable that we wanted to mention to you today was ethnicity. Ethnicity is a hierarchical classification, and most of our standard outputs are produced at either level one, or level four of that classification. Level one includes the broad ethnic groups. European, Māori, Pacific, Asian, and MELAA. Middle Eastern, Latin American, and African. That's used in a lot of our core crosstabulations. More detailed investigations sometimes occur at level four, where the specific ethnicities are. On the graph that Christine has just shown you, ethnicity is the fourth row on this graph. You can see that the majority of the information comes from census responses. Then we have some information that comes from our 2013 census data, and then in the blue is a segment that comes from administrative responses. We'll make it to the stage of using ethnicity data from administrative sources. It comes from three different sources. Some of that information is available at level four of the ethnicity classification. But some is only available at level two and three of the classifications, which means that the very specific ethnicities are not there in some instances.

There's a lot of variety in terms of the information at level two and three across the ethnicity classification. So if you have an interest in this area, it's really important that you look at the classification and understand it, and therefore understand that there may be some impact on some categories. Now I would like to talk a little bit about missing data. Before we mentioned that when there are no high quality alternative sources, and no feasible approaches for imputation, that there's a high proportion of missing valuable, values, than for other variables or for previous instances. This can be a little problematic for our 2018 variables that are new. If there's no alternative sources, and no ability to impute, obviously there's no information from 2013, that we can draw through as well. So this will impact some of our variables such as usual residence one year ago, disability and activity limitations, and our new housing quality variables, so damp, mould, and access to amenities. Now I would like to talk about our Māori variables. As we've discussed, there was no mitigation for non response for most variables in the 2013 census. And we know that the characteristics of those who respond are not necessarily the same as the characteristics of those who do not respond. The impact is therefore greater on some data, and some parts of the population. For example, the post

enumeration survey from the 2013 census showed that the net undercount for the Māori ethnic group was 6.1%. This compared with 1.9% for the European ethnic group, or 2.4% for the total population. No mitigation for Māori descent or ethnicity in previous censuses is therefore going to have led to an undercount of those populations. But introducing mitigation, and alternative sources in the 2018 census, we now have better coverage for the Māori descent and Māori ethnic populations than we did in 2013. Unfortunately the news isn't so good for Iwi. The lower than anticipated levels of participation in the 2018 census have resulted in a significant proportion of Iwi showing a decline in affiliation that was not consistent with our expectations. Our ability to fill the gaps in this missing data is limited, and this is for a couple of reasons. Firstly, there's a lack of coverage and quality of Iwi administrative data, and secondly, there have been significant classification changes. Between the 2013 and 2018 censuses, there was a major review of our Iwi statistical standard, and classification. This resulted in 35 additional Iwi being added to the classification that was used for the 2018 census. These changes mean that it's very complex to try and use the 2013 data to help repair or mitigate the 2018 census responses. This is because the additional Iwi that were added to the classification were not part of the information that we tried to collect, they weren't on our help notes or information to guide people as they responded in 2013, so we can't assume that those counts are representative for the Iwis in 2018. As a consequence of these factors, the government statistician has decided that Iwi counts will not be released as official statistics from the 2018 census. And she's noted this as a significant loss. The decision was supported by our external data quality panel. We will be working in partnership with Iwi, and Māori, to find solutions to Māori data needs. This includes exploring options for the provision of non official Iwi data. That is, we asked all likely to make Iwi information available, it just may not be as an official statistic.

Next I'd like to talk about families and households. Christine talked earlier about the fact that some admin enumerations have been added to dwellings, and that some have not. You remember that the lowest geographic level that we're able to add those individuals to, is a meshblock. Because families and households are derived at a dwelling level, the administrative enumerations that have only been added to a meshblock will be missing from the derivation of families and households. We saw earlier that this applies to around 357,000 people, and with therefore result in some families and households being either missing, or incomplete. This is a key area of investigation for our evaluations team who are building an understanding of this, and of any other factors that may also impact the quality of our family and household data. So I'd like to summarise the two methods sections that we've had so far. Alternative sources add real value to the census data set, where they're available. The information we've been able to share with you today, particularly the graphs, shows that there's a significant benefit to a range of variables, especially, but not limited to, our priority one variables. At this stage in the New Zealand context, there are variables where the only place we can get the information is from our census forms. The variables in this instance cannot be mitigated or replaced by administrative data. So Iwi was a prime example of that. Our strength, both for the 2018 census and going forward, comes from having a combination of both census forms and administrative data. Reduction and response rates and missing information from within forms, have been a particularly significant issue for the 2018 census. But they've also been impacting our data quality increasingly for some time. The best quality census data going forward will be achieved through a combined approach of field collected and administratively sourced data. Now I'd like to move to talking about

quality assurance and assessment. The presentation that we've given you so far highlights that data quality is not consistent across all variables. It's therefore important that we understand the needs of our customers, that we have a means of prioritising our effort as we develop, collect, process, and disseminate data, and that we have a clear and structured way to communicate data quality for each variable to you, our customers. In order to do this, we have a quality management strategy. The 2018 quality management strategy is available from the Stats New Zealand website for anyone who requires more data than we can present today. Or if you just need some additional bedtime reading. The quality management strategy outlines a number of dimensions that contribute to quality. The three that we are focusing on at this point in our census cycle are accuracy, consistency and relevant, consistency and coherence, and interpretability. These are the things that are intrinsic to our data quality assurance processes that are currently being implemented across the census.

As mentioned earlier, the priority one variables include our counts, and our core demographics. Priority variables include things such as qualifications, Iwi, income, work and labour force status, and families and households. Our priority three variables include topics such as religion, disability or activity limitations, smoking, and our housing quality variables. The quality of census variables can be affected by a number of factors. The first is missing data. When there's no alternative source, and no statistical imputation. This will result in bias, and impact distributions when non responders are different to responders, such as the example we mentioned earlier with Māori descent and Māori ethnicity. It will also result in cancel levels for a variable being too low. The quality of data will be impacted when we use alternative sources as well. It's therefore very important that we make sure that we understand the quality of the 2013 census data we plan to use, of our administrative values, and of any imputation. And finally, variables will be impacted by the quality of received responses. If a variable is answered poorly, if it had design issues, or if there's issues in our processing system, the quality will be impacted. Over time, we've tried to increase and improve the information that we make available regarding the quality of census data. In 2013, this resulted in the development of our first quality rating scale, where we formally and consistently provided a quality rating for each variable. Each variable had a quality rating of either very high, high, moderate, poor, or very poor. In 2013 we had three metrics we used to produce these ratings. They were the level of non response, the consistency with time series in other sources, and data quality. If quality issues were observed in one metric, that brought the overall rating for the variable down. So for example, if the rating for consistency with time series in other data sources was moderate, but the rating for non response or data quality was high, the overall rating for the variable would be moderate. Ratings for variables were published in our information by variable product on the website, and the information is still available for you if you wish go to and have a look at it, and refresh your memories on it. In 2013, most variables received a rating of either very high or high, with some moderate ratings. The only poor ratings were given to variables that had specific output categories with issues associated with them such as civil unions, where the number of civil unions reported by respondents is often higher than the actual number that have occurred in our population. Or for some of the derived household outputs, such as household income and extended family household income.

So moving on to the 2018 quality rating scale. For 2018, we wanted to retain consistency where we could with the 2013 census quality rating scale. But we also obviously had to make adjustments to make sure that it reflected the changes in methodology that we have, and the range of data quality that we're likely to observe. The main change that we've made is that what was on the left, the non

response metric, has been replaced by a new metric called data source and coverage. This results in a weighted score for each variable based on the proportion of census responses, administrative responses, 2013 census responses, any imputation, and how much missing information there is. And we'll show a couple of examples in a minute of how that works. The second and third metrics, consistency and data quality, will be assessed in the same manner as we undertook previously in the 2013 census. Also as was undertaken in 2013, the overall rating will be produced by taking the lowest rating received across the three metrics. For 2018, we will be publishing the ratings, both in terms of the overall rating, but also the individual ratings for each metric, so that you can fully see how we have assessed the quality for each variable. Now for the examples. So both of the examples use individual form responses. The first example is where we have been able to mitigate some non response with additional sources. So this variable is sources from census individual forms, historic 2013 census data, admin data, and imputation. You can see in the second column there that there is a rating for each source. You'll be able to see for example that individual forms, are sourced at, are rated at a one. Historic 2013 census information for this variable has been rated at .95. The 2013 census source data comes from the table that Christine showed you previously. And so, that is where have undertaken a comparison of the 2013 and 18 responses where people have answered both censuses. The rating for administrative data comes from the work that our census transformation programme has undertaken during the assessments of determining which variables are suitable for inclusion. And in terms of imputation, the writing is based on the quality of the source information that is used for the imputation. For this variable, you can see that it has resulted in an overall score of .973. If we go back to the previous slide, you will see that our now data sources and coverage table there on the left, that results in this variable getting a score of high. And our second example here from an individual form, there has been no means of mitigating the variable. So, the table that is presented only shows individual form source data, and missing or non response information. That's resulted in an overall score of .84, which means a rating of poor. What we're trying to make sure that you know from the information we're providing here is that we are going to be totally transparent about the sources of information for each variable, and what impact that has had on the quality for it. And we will be making all of this information available at the point of first release, in September.

So what does all of this mean for output and dissemination of our data? At the moment, our evaluation process is still ongoing now that all the administrative enumerations have been included in the data set. Our focus has been on ensuring we can produce a high quality count that meets the needs of our key customers. Our focus has been on our priority one variables on Iwi, and we've been able to provide a clear signal of the quality for those variables. Decisions to restrict or not output any further variables will be guided by the data evaluation that we've been talking about, and the use of the quality rating scale that we've just presented. Any at risk variables will then undergo further investigations, and go through a thorough risk and assessment process before any decision is made. If any further output variables are to be restricted, we will communicate this as soon as we can, with the intent of doing this by the end of July. What I'm going to do now is hand back over to Kathy to talk a little bit about 2023.

- And we're obviously, we've done the seminar three times before now, to real people. Listen, I know you're real people out there, but to a live audience. So we are considerably faster when we don't have questions, so, we will have plenty of time for question time and who knows, you may

even get a bit of time back in your day. So, I just wanted to take the opportunity to talk a little bit about 2023. So although we are still working our way through 2018 and very focused on that, censuses never sleep, and we do need to kick off our work around 2023.

So, every census, we go to government to request funding for the programme. Normally we would have asked for the budget to deliver the next five years worth of the programme, for this year. So, this time we've taken a different approach to seeking the funding for 2023, because we really need to understand what worked and what didn't work and independent review is going to be a really important part of that. So instead this year we applied for one year's worth of funding, to both complete the 2018 and to start 2023. The Prime Minister and Minister Shaw announced last week that the funding will be made available in the 2019 budget. So, what that means is that this year, as in this calendar year, we will need to complete the business case in the budget to be submitted in December 2019.

What we do understand about 2023 is that we need to work differently. And we need to partner if we're going to deliver a successful 2023 census, we're really looking for that to be an inclusive census. Both in terms of the way that we collect information but also just all the way through. So we will extend an invitation to any interested party who are interested in being involved, where it matters to you, where it's relevant, so, we will be running a few workshops on different topics over the next few months, so if you are interested, we'll extend an invitation and you're welcome to join us.

So, when is the next information coming? You'll hear from us this webinar, we will send out a link to all of you who did register. And the link will also be available on Lumio, which is the tool that we use to interact, and we will be using again to interact over the next wee while, so if you're not on Lumio, and you would like to be, please get in touch, you can use our census standard email address, and that will come through to Sophie, who will action all of that, and make sure that you get access to that, to that forum.

In terms of about 2018 census, as Adele has mentioned, we will provide an update to our customers on the quality of variables. This will come via usual mechanisms, the census advisory newsletter, the Lumio engagement, and where we know that variables are critical for particular customers, we will endeavour to have some face to face meetings as well so we can talk it through with you. We will be at the PENZ conference in June. There probably won't be a whole lot of new information there so if you are a PENZ attendee probably don't expect to see a whole lot more than what we've already talked about. And obviously we mentioned the independent review, that will be released by the end of July, so, you'll get more information about what worked and what didn't work then. And as I mentioned, 2023, we will be in touch to provide you with an opportunity, if you are interested to be involved.

Okay, so, that ends our presentation to you. We know we have got a few questions that have come through. Please, so we'll start to answer them, Adele and Christine will join me now. And, if you are, if you are interested in asking some more questions, please send them through. We've got Theresa and Mike waiting eagerly to receive any questions which they'll flip through to us. So we have got a couple.

The first one was around roughly what percentage of the Māori population will be drawn from census data, and what percentage from administrative data, so we've given you that 89 and 11 at that total level. I'm afraid we haven't got the Māori data for either descent or ethnicity to hand, but, this is not the first time we've been asked that question. We understand that that's important for us to be able to release, so we will be providing the information once we've pulled it together. Just takes a little bit of time to work that through. So, where you have data from both census and administrative data, have you compared how often they match, and if so, how does that comparison show? So I'm pleased to say the answer to that is a yes. In some instances, anyway, and I'll pass to Christine to answer that question.

- So the diagram that we had with, we call it the bubble diagram with all the blue circles, that represents the work that we'd done comparing the 2013 census to the 2013 administrative data. So we have all the information about how consistent people were in 2013. Right now we are running the comparisons for people who've responded in 2018 versus the people that we would be using administrative data for in 2018. So that's in progress, clearly we've only just got that information. But it's been based off the 2013 comparisons.

- Great, so hopefully that answers your question. What other countries use a combined census?

- So, I don't have a big list in my head, but I'll mention a couple of countries. Germany. They switched to a census that's largely based on the provincial or land population registers, as well as a housing survey, I think, and a sample survey for characteristics. Holland uses their registers for most of their census but also combines it with their labour force survey and other household surveys for variables that aren't in the registers. And there are a number of other countries that I can't recall off the top of my head.

- Okay, so hopefully that gives you some idea. Given that there were issues with data collection in census 2018, was there any consideration to do an early census to get back into the pre Christchurch earthquake census cycle, so that would be going to a 2021 cycle, and yes we did consider it, but we're, yeah. A census is a big programme of work, it takes five to six years to run a census, and, we did seriously consider whether we thought it would be an option, but we just do not believe that we could run a successful census in 2021, 2023 is, is when we're planning to run the next census, so... Okay, how comparable is the 2018 data set with previous years? Stats NZ usually re releases two earlier censuses on new geography each time, is this happening?

- So that's referring to rebasing, and we will be undertaking rebasing again and at the point that data becomes available, we will have the historical data there for comparative purposes.

- Was there a fistful of other questions?

- Yeah, I can probably answer that, so, I guess you're talking about the time series consistency, and I guess there's a few things that's probably worth us commenting on there. We did change a number of things this census. And we changed some questions as well, which, what joys and picks on the time series. We did make a decision pre the collection period to also introduce imputation for quite a number of variables, it's something that we haven't done in the past. So that again would introduce a discontinuity. And the other thing is, we obviously, given that the participation rate

hasn't been nearly as high as we hoped, and through the use of administrative data, this introduces quite a different approach, so that is obviously going to introduce discontinuities as well, so...

- I guess one of the things we're always weighing up, isn't it Kathy, is that it's that balance of consistency versus quality. And the decision that we made at the point that we decided that we would use administrative data, historical data and imputation for missing responses in the census was because we believed it would bring better quality data to our customers, so it will have an impact on time series. But we're doing that with the intention that it's going to bring better quality.

- We certainly hope so. So I don't think we've got any more questions that have come through.

No. So, if anybody has two minutes, we'll give you two minutes to... So we have to stand here, sorry, we're... We're entertained, okay, so we'll just give a little bit longer, to, if you do have any questions that you want to come through.

- But that's just so our colleagues can watch us sweat for a couple of minutes.

- I think so. I think so. So... I'd have to say this is all our first experience doing a webinar, so...

- So thank you for being our first audience.

- We hope it's worked okay for you, and on that note, if you have got any feedback that you want to provide us, we would love to hear from you. Has this form worked? If it has, it's something that we could look at using again. If not, then...

- It's helpful to know.

- Yeah. Yeah, it's very helpful to know, so, this question, how big was the difference between the 2013 and the 2018 responses? So... I mean we haven't, until we run our post enumeration, we've collected the data from our post enumeration survey, but to deliver the results around census coverage, we need to process that fully, and that relies on having a completed census data set, so that work is still to happen, and I mentioned that results of that will be published by March 2020. So that's its own event that we will get our true estimate of the coverage rate and the response rate, so I think, you know, we've talked to you about our estimations of our coverage rate, anyway. And we've estimated that based off our best guess, if you like, of the 2018 total, using the revised ERP and that, you know, aligns very well with our estimate using the dual system estimation as well, and that's coming at 1.2%, which compared to the equivalent rate for 2013 of 2.4%, so we've obviously very encouraged by that. We don't know that that's exactly where it's going to land, until we've done the census coverage survey and worked that through, that's when we'll have our official results around that. Okay. How does imputation work? I think there's probably not a quick answer to that, but, do you want to?

- So just briefly, there's some variables that we usually, if we're going to impute a variable like ethnicity, for example, then, we'll have a number of other variables that are related to ethnicity. Some of that's just a dual set, because it's the series by that but also we'll have birth place, or language, or Māori descent or things like that that are related to the variable that you want to impute, and it's effectively a prediction model based on the variables that you have. But we use a

particular software that's been developed by Statistics Canada, to get technical, that uses a nearest neighbour data imputation, so it's a statistical process where you find the most similar person to the person who's got missing data and then borrow the missing values from them. So it's a statistical process that gives us our best prediction. We don't use it unless we think we can get a good prediction of what the true value might have been.

- Okay. I can answer the next one. When will the 2018 census data be published? So, the first data will be released on the 23rd of September later this year, and from there, we will flow with usual range of products, through to mid-2020. Okay, another question. Is it likely that any variables will be released only at broader geographic levels? EG, available for territorial authorities but not SA twos or SA ones. Or if it is released, will it be an all or nothing level? So, that's a really good question and we still are actually working through how we do that. What we can say is that whatever we do release, we will make sure that we have got a lot of metadata available so that you can understand the quality of it, and Adele talked you through some of the tools that we'll be using to provide that. So, I guess it is possible, that we may only release some data at high geographies, but certainly wherever we can, we will be looking to release it at that finer level, so, that'll only be if we think that the quality is really, you know, no good.

- Everyone's got irrelevance.

- And I guess that's because too, we want to make sure that if the quality is good enough at those higher levels, then we would want to make sure that we made that information available which is why it wouldn't be an all or nothing approach, like if the quality is good enough, we will make it available, if it's not, then we will have to go through a process of assessing what our options are.

- Yeah. Yeah.

- But everybody, every person is placed in a meshblock, so, there's the potential to release all information at statistical area one level.

- Yeah. And that's certainly what we'll be aiming to do. Do you have any thoughts on the distribution of admin people records? Is it evenly spread across the country or concentrated in certain areas? I guess the thing that I would just say and then I'll hand over to Christine for probably the more technical answer is, we have, we know, and I'm sure a lot of you who are, you know, active users of census will understand that historically, we have had challenges, and some particular locations, Northland, East Cape, in particular, South Auckland, and again, we have experienced challenges in those areas. This census we also have more of an everywhere missingness on top of the usual challenges that we have. So, I guess that's the answer to that question is almost both yes...

- No, I think that's a good answer Kathy, I don't have the detailed answer more, but, there will be administrative enumerations across the country but there will be more of them where we have lower responses.

- Okay. Here's another question, which might actually look at Mike to answer, I'm not sure. Is there a release date for the census dictionary? We need this to get our custom data order in the queue. Right, for custom, yes, okay.

- So we'll make available at the same point that we start processing customised requests.

- Which we hope to be soon after the 23rd of September.

- Yeah, we do, certainly do. And I guess if that's going to cause a problem for anybody, just get in touch with our customer team and they'll try and help you as much as they can. What advice can you give you creating trim data for special areas such as census unit, I presume it's it, unit, in meshblock level?

- In your unit?

- Oh sorry, area unit in meshblock level. How to compare the smaller 2013 data set to the 2018 data. So there's a number of questions there, so creating trimmed data for special areas, such as census area unit and meshblock level. That might be a question that we don't have the expertise to...

- Well I don't think it's going to work.

- And I think it's going to depend on what variables you're working with, so I think it might be a case of having to work.

- I think it might be quite complicated. Well it will be quite complicated. The first thing perhaps is to say is that we have new geographies now, so, there won't any area unit or meshblock data released. Instead, we have new statistical geographies called SA one and SA two. SA two roughly equivalent to area units. So suburbs or small towns. But they have been created in a way that's good for statistical analysis, so there's a fairly even distribution of size, for example, and we've tried to represent communities of interest better than in the past. The SA one geographies are smaller. They are replacing our meshblock outputs. The SA ones are again, a better statistical unit for producing information. There are no tiny meshblocks, that we'll have to, tiny SA ones, that there's a minimum size, and they do try and represent communities of interest better than the meshblocks as well. So there should be a different, a better geography for statistical analysis. And are we creating previous census data on those geographies?

- Yes, we are.

- So, on the geographies themselves, we'll have a time series of the same areas. In terms of time series for difference variables, I think there will be some interesting and difficult issues perhaps, to work through, because of the changes.

- Okay, when will custom data requests be processed and released? So, Adele just mentioned that, that's going to be as soon after the 23rd of September as we can. The travel to work variable, it's travel in the census week, so how can that be imputed from surveys, et cetera? - It's imputed from the responses that had been provided on other census forms?

- Yeah. So anything that we've imputed is from data that we obtained in 2018. And using the information that we have about the person to find somebody similar.

- What impact did the loss of statistics house following the 2016 Kaikoura earthquake have on the 2018 census just run 15 months later? Look, this is something that the independent review will

be touching on, so, I don't think I can say anything useful right now, but certainly that will be considered as part of, part of that review, so that is, it's an independent review, and it is due out in July, so, and it is on track to be released, so expect us to hear more about that there. Would the quality of data, very good, good, et cetera, for variables, be released for each of the TLAs, regions, or a national level?

- That's a great question and again, one that we've had before. I think that's something that we'll need to be working through with our products and services team. It's a change that we've had this time, so we will have to work with that, and we get the best way of making that information available to you, we don't want to overwhelm you by providing so much metadata that it's actually hard to, hard to use it, but at the same time, we need to make sure that there's enough, and so we're going to have to think about how we present that information at lower geographies.

- Okay. Okay. So, we haven't got any more questions, I think at this point. So, again, we'd really appreciate your thoughts on how the webinar went. So, a link will be sent out for you to provide some feedback to us on that. We certainly hope that it's been a useful session for you, and thanks for attending, and, yeah, we'll be in touch. Mā te wā.

Stats NZ Tatauranga Aotearoa

[End]