Biosocial research: biological data in social surveys

So, this is the second video looking at how we might combine biological and social data in order to understand people's health and society better. In this part of the video we're going to look at what biological data look like, and some of the things that we need to take into account when we are analyzing them. And I'm going to give you some information about the kinds of studies that include these data and where you might go to for further advice.

So first of all, quite a few studies in the UK have added biomarker data to the rich social information that they collect from their respondents. So, a number of the longitudinal studies have done this. For example, understanding Society which I'm involved in. That's collected biological data across the whole adult age span. The English longitudinal study of Aging, that collects data every couple of waves from people aged over 50. The health survey for England and those for Scotland. They are not longitudinal data, they just collect data from new people each time, but they also include some biological measures. And then the cohort studies. The 1958 cohort in the 1970 cohort also has biological data. And all of these data both the social data and the biological data can be downloaded from the UK data archive. There are other studies, for example some of the ones funded by the MRC, which also combined both biological and social data, but you have to go to the survey teams or study teams themselves to access these data. So perhaps the biggest one is the Avon longitudinal study of parents and children based at Bristol. But there are also others. The 1946 cohort, the Southampton women's study. Closer, which is a consortium of lots of longitudinal studies in the UK has some guides what these studies contain in terms of the different kinds of data.

So to give you an example of the sort of data that we might collect, in understanding society, the way we did it was we sent a nurse to everybody's house and the nurse was there for about an hour and took a range of measures. And that's really traditionally how many of the studies have collected these kinds of data. Some bring people into a clinic, and one of the things we're all experimenting with a bit more now is thinking about new technologies. Could you use your smartphone to measure your pulse and tell us a rather than has have to send a nurse to visit you and measure your blood pressure. But traditionally what we've done is this a nurse interview. So the nurse takes a range of measures, so we measured people's height and weight and their waist circumference and body fat. All

as a way of kind of understanding their body shape and levels of BMI and obesity. And obesity is a kind of problem for a whole range of health issues, but we also see it's associated with other things for example it's associated with employment opportunities. Another thing that we do is make people blow into machines to measure how strong their lungs are, and this tells us obviously about their respiratory health, but what we see from these studies is that your respiratory function is really strongly correlated with mortality, and so it's a really good indicator of people's overall fitness. We measure blood pressure and pulse which are risk factors for heart disease and stroke and diabetes. Grip strength which I talked a little bit about in the first video this is the measure of your upper body strength and it tells us also about how quickly you might become frail in old age. Often these nurse visits collect blood, and I'll say a little bit about that in a second, but alongside all these kinds of measures that we take we need to ask people questions about what they've been doing that day. For example, whether they've had a lot of exercise with they've been smoking or drinking, whether they've taken medications, because these might things might affect the measures that we want to use. In terms of blood, in these big studies we try to kind of identify analytes that are relatively common in the population, so you'll be able to see people who have these conditions. If you're wanting to look at social things and health, they need to be things we expect to be affected by the environment, they need to be kind of true for the whole population not just subgroups. So again, just to use understanding society as an example, the sorts of things we've collected and lipids, so that's your cholesterol levels, that's about fat in your blood and it's an indicator of heart disease. We've talked in the first video a little bit about diabetes, and the measure of that in your blood called hba1c and so often studies collect that. Inflammatory markers which tell us where someone's got an infection or information, but it is also a kind of chronic pathway to stress. We might think about your immune system, so this is something which has a lot of wear and tear as you get older and might therefore kind of not work as well, and we collect measures of that. We look at things like anaemia, how much iron you've got in your blood. And that's really an indication both of your nutrition and your general state of health. And at different life stages and for men and women that might have different implications. These studies often look at people's liver function and kidney function. So, your liver function tells us a lot about the extent to which you might have been harmed by drugs or alcohol or very fatty diet, and those things are really important to understand because the liver might be being damaged by them. Poor Kidney function is is becoming an increasing problem in developing countries at the moment so it's quite important. And then when we want to

think a bit about how people grow in childhood or how they decline in adulthood, we won't would think that a range of hormones that are implicated in either growth or decline. So, these are the sorts of measures that studies tend to have. And when you analyze these data you know there are lots of things about them are just normal statistical continuous variables, so the sorts of things you think about then are there outliers, what's the distribution like. You need to consider analyzing these biological measures. But there were also a lot of other things you need to think about. So for example with biological measures it's often not just high as bad and low as good, it's often U shaped, so having both low BMI and high BMI might be bad for your health in different ways. Some of these factors are different depending on with your men or women or for different ethnic groups that you need to consider in looking at them and how they relate to people's social lives. So an important thing to consider is clinically feasible ranges. So sometimes we might have measures that we know would be impossible and we can kind of use them to clean the data, but there are also clinical cut-offs when we know if your hba1c level is above a certain level then that's a sign that you have diabetes. We need to consider recent events. If you've had an accident or an operation then your inflammatory markers are likely to be really high, even if you're kind of chronic stress levels the things we normally look at them for are fine, and so you need to take those things into account. The way in which the blood has been taken and processed is really important, but equally the time of day, so there are some of these analytes I'll show you in a minute that are much higher in the morning than the evening and you need to think about that when analyzing them. Sometimes having other conditions might affect those blood levels, not the reason that you think of but because of something else going on in somebody's life. And medications are important, so you need to think if you're interested in whether someone's has cholesterol, do you care about whether or not they're taking statins that will be reducing that, or not, is that important to your research question? If it is then you might want to adjust your measure of cholesterol because they're taking medications, but if you're interested in for example subsequent health, the fact they're on statins is a good thing and you might leave the measures as they are.

So, all these different things are important when analyzing the data. And on the understanding society website we have a glossary which explains them for all the different biomarkers and what you might do, and I'm just going to give you a few examples now. So first fall I want to talk about c-reactive protein. This is a measure of inflammation which might be high just because you've got an infection like a cold, or if you've bumped into

something. But it also is high if you experience chronic stress. And so, we see from studies it's associated with different measures of your social life and with aging, and we know it's a risk factor for a lot of important diseases such as heart disease, cancer, arthritis. When you look at c-reactive protein data, it it ranges from 0 to higher, but measures over ten are because you have a recent infection like a cold, so mainly you want to exclude them from your analyses because that's not the sort of condition that you want to be looking at. What you're interested in is that long-term impact of people's social lives, onto biological stress, onto their health. So, for that you want to be looking at measures under 10. For people in the range 0 to 3, that's a very normal kind of level of CRP so they're healthy and we don't need to really worry about them. It's those people whose CRP are between 3 & 10 therefore that we need to worry about. And those are the people who have heightened risk for example for heart disease. So, this is what CRP looks like in the population. As you can see, most people have a CRP level under 3, but there is this tail between 3 and 10 of a range of people and those are the people with heightened risk scores for heart disease and cancer and arthritis, so in your data you might want to be looking about what sorts of social factors are associated with that that might help you to understand why we see social inequalities in these diseases.

So, the second example I want to give is ferritin, so ferritin is the measure of how much iron there is in your blood. How well you store iron, and iron is really important for a range of things. But this is one of those measures where both low and high are bad for you. So low levels of iron in your blood means you have anaemia. It's much more common in women, and it's associated with feeling fatigued and kind of not having energy to get up and do things. But high measures of iron in your blood also really bad for you. That's much more common in men and it's associated with heart disease and diabetes. So, if we look at ferritin in the population, the dark bars are those where people don't have enough iron, and you can see that's much more common in women and much more common in women of childbearing age, and then it tails off as women move through the menopause and into older age. And there's very few men really that have anaemia like this. But men do have iron overload. You can see it goes up into their 60s, age 60s, and then it comes down a bit. But it's much higher than for women and so those are the men who are at risk of developing heart disease and diabetes because of this. And so we can kind of start to think about needing to treat them differently than perhaps other causes of heart disease.

So, the third example I wanted to give is testosterone. So, testosterone is a hormone. It's a hormone we associate with male behaviour. We think of it being associated with aggression. But it's also the hormone that helps us as kids develop, and in older age very low levels of testosterone can lead to frailty. There have been studies that show that levels of testosterone are associated with self-employment. There were high levels of testosterone in the stock market crash in some studies that kind of looked at that. So, it's associated with perhaps aggressive behaviour in middle-age.

So, these two graphs have a number of points about the way you might think of looking at testosterone. So, you can see for men, the left-hand graph, there's a relatively normal distribution. And between the two parallel vertical lines, that's kind of normal levels of testosterone. So those people to the left of the first line, those men have low levels of testosterone and we might need to worry about that because it might suggest some sort of frailty and later life. And those men to the right of the second line, that's suggesting they have excessive testosterone and we might need to worry about that in terms of aggression. For women there's this one tall spike of data, and that's because most women have testosterone levels that are too low to be detected. And so really these data are only usable for men. But you will see some little bumps as you go along that line for women which shows there are some women with high levels of testosterone, and this currently is study going on in understanding society which is looking at how that relates to infertility in women. So, although for most women these data aren't very useful because their measures are too low to be detected, for those where it is detectable, then there are some health issues that we can look.

There's a second issue I wanted to talk with testosterone to illustrate another point about the way you need to think about these data perhaps differently to social data. So, you may remember I said that testosterone declines with age, and you can see that happening really nicely from the age 16 until the mid-40s. But then it flattens, and it goes up. So that's really counter to what we expect with testosterone. We really expect it to be a kind of onward decline into old age. But then you need to start to think about the interview process. So those people who are aged say 16 to their late 40s, they're probably likely to be at work and therefore when the nurse visited them, that's probably going to have been in the evening. For men in their 50s and 60s and older, maybe they're not working so much now, maybe they were available during the day and so they were interviewed perhaps more in the morning. And this shows that if you don't think about

the time of day when somebody's interview took [place], you might have misleading understanding of your data, so this suggests that actually testosterone doesn't decline in old age. But it does, it's just that the older men were interviewed in the morning younger men we're interviewed in the evening after they came back from work.

So, the last specific example I want to talk about is kidney function, and the reason for raising that now is because it's one of those measures where you can't just take the blood result that we have, but that you need to adapt it in different ways to make it meaningful. So, kidney disease is an increasing public health problem. It's much more prevalent in aging populations in developing countries, and we're beginning to see it's really socially distributed. So, we really need to understand how and why it's socially patterned. But we can't just use the measures in blood. What you need to do with kidney disease is have different cut-offs for the blood analyte, creatinine. Whether they're people are men or women, whether they're white or from ethnic groups, and for different ages. So, there's a complicated formula that you need to follow, and you'll find it in textbooks it's a standardized international approach to doing this. But this is what you need to look at if you want to study kidney disease, not use the raw data. And what this does is create stages of kidney disease. So, stage one means somebody doesn't have kidney disease, up to stage five where they're in kidney failure, renal failure. And if we look at that in the population then you do see what you would expect. That basically most of the population doesn't have kidney disease. Stage one, maybe stage two, but they're really kind of free of disease. But increasingly as people age, the proportions with the higher levels of stage 3, 4, & 5 kidney disease become apparent, particularly at the older stages. And that really reflects what we see happening in the wider world and clinical studies.

So the last thing I wanted to talk about is a measure that's really popular in social science called allostatic load. This measure was developed by Bruce McEwen in the States. It's a way of looking at how chronic stress over people's life courses might impact on the different physiological systems in our body to cause damage to them, and how we might measure that. So, when you think about the stress response that goes on in our body there are kind of three stages to it. So first of all, when something stressful happens, you stand up to give a talk, you run for a bus, you don't have enough money to pay a bill. There's a kind of flight-or-fight response. Your adrenaline goes up, your heart rate goes up, and these are perfectly natural things kind of going back to cavemen. They're getting you ready to

either run away from the situation or stand and kind of see it through. So there's natural and it's good for us, it helps us perform in those kind of stressful situations. But if that happens over and over again we have high cortisol because of the stress levels going up constantly in this fight or flight situation, or our blood pressure and heart rate keep going up, that can start to cause damage. And that causes damage in a set of secondary kind of aspects of your body. Different physiological systems, so your immune system, your metabolic system, your cardiovascular and respiratory systems. All of those might start to get damage as kind of this this constant fight-or-flight raising of different hormones and heart rate. And then finally in the third system, once those things have happened. You know your blood pressures constantly raised, you're increasing your weight, your respiratory system is is being reduced because of this constant stress then, that leads to what's called this tertiary outcome which is you actually manifest into disease like heart disease. So allostatic load is a measure that combines indicators from each of these three systems or stages to create a cumulative burden score on people's health. How people do that really varies by studies. So, lots of studies find it really hard to measure these primary responses to stress. Cortisol is really difficult to measure or measuring heart rate reactivity. You know you have to kind of put someone's hand in cold water or something to make them react. So they're not things that are traditionally done in normal surveys, so also studies don't have that and they research has just combined the second two sets of variables. And then how they combine them, it really varies as well. Some people just add up basic kind of, if you're unhealthy on this thing you get a one, and if you're healthy you get a zero, and add it up. Other people weight it. How being unhealthy is defined varies. So, there's lots of debate about the best way to measure allostatic load. But what we find is that however these people might measure it in their study, there's a really consistent pattern of increasing allostatic load as you look at increasing ways of measuring disadvantage over time in people's lives. So, while there may be some debate about how we measure it, it seems to be a really good way of capturing this physiological burden or wear and tear on the body that's associated with social stress.

So those were just some examples of the ways in which you need to think about biological data differently to perhaps income or other measures that you might include in your research. Not to say that they're more difficult but just you need to consider different things. On the screen now, a number of websites where you could go to get more help. So, the study websites themselves. Closer, which has brought together a range of ways of looking

at biomarker data in these different studies. The data archive where you can download the data. I haven't talked about genetics but Metadac is where many of these studies that have genetics data share it, so the Metadac is a committee across a range of studies for sharing genetics data. And then the NCRM website provides training information about how you might use these data in different ways.

Thank you.