

Welcome to the series of presentations on some of the methodological considerations researchers should keep in mind when they're doing biosocial research. My name is Tarani Chandola I'm from the University of Manchester and part of the National Centre for Research Methods (NCRM) and I also like to acknowledge the Understanding Society Biomarker team who contributed a lot of the materials and slides to this set of presentations. First of all I just like to acknowledge that the methodological considerations we should be keeping in mind when we're doing biosocial research is it's not special, it's nothing new, it's just the same sort of issues that should be kept in mind when we're doing our standard biological or social science research. It's just that some of these methodological issues get really highlighted and emphasized in particular because of the nature of biosocial data and I'll be going through some of the examples in which these issues really get emphasized. So this series of talks is divided into three components. First of all I'll be talking about the need for a biosocial research framework, secondly I'll be going into some of the data quality considerations researchers should keep in mind when they're doing biosocial research and finally I'll be talking about some of the missing data implications in biosocial research.

So the very first part is about why we need to have a biosocial research framework and for those of you who are completely new to the area and want to know what is biosocial research I can recommend a really excellent talk by Professor Michaela Benzeval on the topic 'what is biosocial research' and I've got the YouTube link over here and this is part of a series of talks that she has done on biosocial research and this is a talk that she did for the NCRM. Why do researchers want to combine biological and social data and that's perhaps the most important question in determining the kind of research framework that we have in approaching biosocial research. There are a number of reasons why people are interested in combining biological and social data for example people might be interested in using bio markers as an objective measure of physical functioning of health and illness, they might be interested in using biomarkers to look at the pathways between social factors and health or they might be interested in using biomarkers to understand how biological factors are active distal causes to influence social outcomes. In line with that biomarkers can also be used to understand gene and environment interactions and I'll be going through some example slides of how each of these biological bio social frameworks can help in our understanding the associations between biological and social data. So biomarkers have often been used as a sort of a better more objective measure of health and that's

because self-reported health that is collected in standard surveys often has a lot of bias, it depends on people's own perceptions of their health so it's very subjective, it depends on people being aware that they have particular disease or illness conditions, so you can imagine that a self-report of health might be considerably affected by all of these biases, by persons mood on the day that they were sampled, whereas the biomarker which is you know based on a clinical objective measure may seem to be free from such biases but actually there are advantages and disadvantages for using biomarkers in place of self-reported health and I'll go through a few examples of that. In the labour force survey for example they measure stress or work-related stress and that's my own field of research and they measure stress by asking people have you suffered from any illness or disability physical or mental problem that was caused or made worse by your job and if people say yes to that they then get a list of conditions that they ask, you know how would you describe this illness and one of those sets of descriptions is stress, depression or anxiety. So if people say that they their work that their health was made worse by work because of stress, depression or anxiety that's an indication that they have got work-related stress. Now you can imagine that there are quite a few biases associated with this measure, individuals are asked to self-report any work-related illness that they believe to have suffered over the whole the previous 12 months, so it really depends on their ability and willingness to self-diagnose these links and they have to be in sense epidemiology or medical doctors because they have to ascribe the cause of the illness to work and maybe there is a link, maybe there isn't a link and people may actually fail to recognize a link with the working conditions when there is one. So there all kinds of reasons why you know self-reported data on stress or work-related stress might be hard to measure from a questionnaire or interview. So in contrast you hear a lot of people like using biomarkers of stress, so there are well-known physiological systems where which produce stress hormones which activate the body stress symptoms and this just gets really you know this gets a number of hormones such as cortisol, adrenaline, noradrenaline going in your system and this has been measured in quite a few biosocial studies and it's also important to remember it's not just the activation of these stress hormones during an acute stress response period, that is important, but it's the recovery. So this is some of the disadvantages of measuring biological data as opposed to self-reported data because you've got to be measuring these stress reactions as physiological stress reaction over a considerable period of time and it costs a lot of money to not only measure these stress response systems with biological stress response systems but also if you if you're

thinking about measuring over a long period of time then that adds considerably to the cost. So you can imagine asking people a question by questionnaire might be a lot less costly than trying to measure their biological stress responses. Biomarkers have also been used to measure how people grow and develop and change over the life course, and there are there number of biomarkers that are useful for measuring the life course of developmental processes, growth hormone for example. High levels of growth hormone or testosterone in early life might be indicated for one kind of process that may mean something very different when you have got high levels of the same hormones in later life. So it's important to keep in mind the life course processes that the stages of the life course when these particular biomarkers are measured because they have different implications and different meanings. The gene environment interactions are shown some of the possible gene environment interactions are shown in this slide. The G denotes genetic factors, the P denotes phenotypes. Sometimes these are biological phenotypes, sometimes these are psychological or personality phenotypes. The E denotes environment. In gene environment analysis a lot of the times the environment is the social environment and the Y denotes a sort of distal outcome usually it's health but it can be another kind of outcome people are interested in. So in the top right hand corner we've got a causal process that shows that the genetic factors influence the distal outcome, say its health, independently of the environmental the social environmental factors. The genetic factors are influencing the phenotype usually as I said it's a biological phenotype. On the top right hand corner we've got effect modification going on here so the genetic factors which influenced the biological phenotype or personality of psychological phenotype, they modify the association between social factors and the distal lie or the health outcome. In the bottom left-hand corner we've got a process where the genetic factors are actually causing the environmental response, so a lot of the times people have looked at the way how genetic factors are correlated with educational attainment for example, or intelligence, and so there's a whole process by which genetic factors can influence the phenotypes which result in a low educational attainment for example and so there's been a whole series of studies that have tried to disentangle the way in which biology is the independent variable and that affects social environmental outcomes. And in the bottom right hand corner we've got an example where genetic factors are influenced by the phenotype which is actually causing both the environmental factor to occur as well as the distal, Y the distal health outcomes. So the association between the social environment and the distal Y or the health outcome is confounded by the genetic factors.

So in each of these examples of biomarker research we really need to be very careful in trying to find out what is the association between the social and biological data and that really needs careful consideration within the relevant theoretical framework. If we don't have that theoretical framework we are in danger of making multiple comparisons resulting in trying to find associations with the lowest p-values which is basically unscientific and non-reproducible and that's because with bio social data we have hundreds and maybe thousands of variables that are both relate to biology of somebody as well as their survey responses and the temptation for a lot of researchers is just to do a lot of correlations and do key hacking for example, which would be nonsensical because we need to be do analyzing these associations within the context of a particular theoretical, a particular biosocial theoretical framework, and finally I just like to emphasize the need for interdisciplinary research teams because this kind of bio social research framework does require expertise of both in biological sciences as well in social sciences and unless somebody is trained in both it is quite hard for one person to come up with a relevant theoretical framework and relevant biosocial theoretical framework which is why some of the best bio social research does rely on through multidisciplinary interdisciplinary research teams.