Epidemiology and Risk

Epidemiology is the study of patterns of disease in populations. And this can often be one of the first steps in development of a risk prediction model, is just to show the numbers, to show the number of people within a population and the number of people who get a disease or become unwell. And one of the things about epidemiology is that it can tell us what's going on in within a population, but one of the first steps to do that and also to develop a risk prediction model is to actually have good data, to be able to explain potentially what things lead to an increased risk in a population.

So this is just an example of a patient’s or an individual's pathway. We use things like real world data or routinely collected data interchangeably. And this is just information that we collect on people when they have contact with health care services, and we can look at information that's collected repeatedly, for example, like BMI, things that might change over time, we can also look at diagnosis of conditions, and if we have all that information on individuals, we can build that up to population level data.

The only way a risk prediction model can discriminate is if it knows who does get ill and crucially who doesn't, it enables us to look at risks. For example, the impact of smoking on an individual’s risk. You can look at hundreds of thousands or millions of individuals especially when we’re thinking of cancer; there’s lots of heterogeneity, whereas in a randomized control trial, for example, you might have a very narrow age range, or you might only look at those people who don't have other comorbidities. Real-world data will look at everybody.

One of the disadvantages of real-world data is you might need to link it with other data sources. That can take a bit of time and there might be a cost associated with this. You might need to think of the time period, so how much follow-up time you've got, especially when you’re looking at cancer, if you've been looking at people for long enough to determine truly whether they do or don't develop a cancer, or whether they do or don't die, or do or don't have a cancer recurrence.

And also potentially your control group, if you are relying on people having contact with health care services, those people who are less likely to have contact with health care services and become unwell, mean you might not actually have some of the other basic information on, for example smoking status, that might be crucial to your risk prediction model, things that you have to take into account when thinking of routinely collected data and development of risk prediction models. And there’s a huge clinical need, so the importance of cancer risk prediction models.

Once you do develop an accurate prediction model, you could potentially stratify your population, so you, as a clinician, know who is more likely to get a cancer than those who aren't. And also crucially, if there are screening programs, those people at higher risk can go into either breast or colorectal cancer screening programs and potentially be picked up earlier. Because you've entered their data into a risk prediction model, they can then talk about symptom awareness for individuals or modifying potentially risky health behaviours.