

Hello, everyone. It's great to be here. I'm going to talk about using people's medical records, other people's medical records. Basically, lots of people to thank, many more than listed there. The main thing, I think we're living in an age when so much of our data, and our health data, is getting computerised and we can use it for research. We can do cheaper and quicker research, but what matters really, I think is that we do better research. I've been using this sort of data for a long time for research, and it's pretty clear it's quite easy to quickly produce bad research and it's quite difficult to produce good research. For me, to focus on the good end. I thought I'd start with probably, the first project I was involved in using these kinds of computerised data, and this was about measles, mumps, rubella vaccine. There's a paper published in *The Lancet* in 1998 suggesting that this vaccine causes autism, and it caused a bit of an outcry, to the least, measles vaccine coverage fell across the world. Outbreaks occurred at the time and have continued to occur as a result of those falls. This was the UK data on vaccine coverage. The aim is by the time of your second birthday, babies will have all been vaccinated.

Up here is a 90 per cent figure and 90 per cent is a magic figure for measles vaccine, because if we vaccinate more than 90 per cent of the population, we get something called the herd immune threshold. We're a herd of animals and the virus is there, if more than 90 per cent of you are immune, then the virus can't get hold in the population and is kept under control. If that coverage of vaccination falls below 90 per cent, which it did following the paper in *The Lancet*, we're going to get measles outbreaks. It fell quite sharply as you can see, this is when the study raised its concerns, the public faith in the vaccine dropped off. This was mapped internationally as well. We were funded by the Medical Research Council in the UK to do a large study of this problem using computerised health data, with similar studies in the USA and Denmark. These studies were only possible because we had computerised health records; I'll come back to that point in a minute. This is a somewhat complicated graph but it basically shows the results of the individual studies. This was our study here and their confidence intervals here, and essentially, if you pooled the data from all the studies that were done at around that time, you found a convincing lack of evidence.

You found a convincing suggestion that there was no problem at all between this vaccine and autism. This is one of those graphs you see as a teaching example of someone being completely over the top, because if I put when our study was published, it looks like we saved the world. In fact, there was a worldwide effort to restore confidence in this vaccine, of which my study was one tiny little bit, but what matters is that confidence in the vaccine was restored. If we hadn't had the computerised health records of people, if we'd had to go out and see 5,000 kids with autism around the UK, similar numbers in Denmark and America, and measured everything and done everything from scratch, we'd still be doing that work today. I dread to think where vaccine coverage would be, probably somewhere down here. We could do this work because we had the records. I'm going to jump straight to diabetes. This is diabetes control and heart disease. HbA1c is a measure of how well your diabetes is controlled, it's a long-term sugar measure. There have been international conferences and huge amounts of money spent on deciding whether the cut-off or the threshold should be for 48 or 58.

I've even seen conferences discussing whether it should be less than or equal to 58 or just equal to 58, and it's costing lots of money. We did a study, Anoop Shah and others did a study showing using computerised health records, what we found was that, here's these people, this is their risk of developing heart disease, you can see it goes up as your blood sugar control is worsened. What we found was this very, very bad group and if you look at this group, they've got by far the worst risk of heart disease, much the highest risk of anybody. Who are these people? They haven't got a value of 58 or 59. These are people who haven't been to the doctor to get their sugar measured. These are the people who are getting the heart attacks and the people who are most at risk. We should be having international conferences on them. We couldn't have done this work without the health records. These are the people who are not in our studies and don't keep turning up and we couldn't have done this study without the health records. Now, one of the things I think Biobank can really do is help us look at a whole complexity of exposures that make some people ill or some people better. We heard this morning about genetics and this idea that, things are much more complex than that.

Well, here's a very complex exposure that a lot of people are prone to, which is hearing reports on the news and reading newspapers as a determinant of health. In 2013, there were a couple of articles in the British Medical Journal suggesting that statins have these terrible effects and they don't work, and there was a lot of media coverage. The lay-media got hold of this story in a big way. It's worth saying that there was no evidence for any of this, it was just an opinion piece or two. What happened was that, there was this big explosion of interest in statin side-effects. This is a fairly simplified graphic from Google Analytics. You can see how commonly people are searching for things. What you see are these two big-peaks in how often, worldwide, people were searching for statin side-effects as a reflection of public concern, as a reflection of what they were seeing in the newspapers and on TV. What we did was look at whether exposure to this sort lay-media reports and the concern expressed in Google search is translated into how people behave, and it did. I could spend many happy hours explaining this slide, but I won't, except to say that, what it shows is this little white window here, is the time period of just following the maximal time when these stories were getting reported in the press.

What it in particular shows is that, people were less likely to get started on their statins following a heart attack, and they were more likely to stop their statin if they were just taking them to prevent disease. These media reports were directly impacting on people's behaviours and whether they were taking a statin or not. It's quite a complex exposure. We worked out that this overall, would probably cause something like between 2,000 and 6,000 extra heart attacks just in the UK alone over the following ten-years, as a response to the change in behaviour we saw as a response to the media coverage. Again, we could only do this work because we had the health records. This was a success in more ways than one, it led to a change in the editorial line of the *Daily Mail* who'd been virulently anti-statin and for some reason, overnight, they changed their line. I've had absolutely no luck getting them to change their line on anything else, ever. The next day, they had some terrible headline about migrants or something, so it was brief respite. One of the things we've

been involved in is linking up different sources of data. It isn't just one source of data, and what we've done is shown what you can achieve by linking up different sources of data.

This is a very simple study. All we're trying to do is count how many heart attacks happen. You'd have thought that would be easy; it's not that easy. What we did was, we linked up records from general practice, records from hospital discharge, called hospital episode statistics, mortality data from the Office for National Statistics, this is what people are dying of, and data from a disease register called MINAP. Which is the Myocardial Ischaemia National Audit Project. It's data that is collected in emergency hospitals about people with chest pain and then presenting what happens to them. What we showed is, when you count how many heart attacks happen, not surprisingly, the lowest count you get from the death registry, because fortunately, a lot of people don't die when they have a heart attack. This was a disease registry; how many you get. This is the count there, this is from just from hospital admissions and this was from the primary care and general practice data, which seemed to capture far more of the heart attacks that happened. What happens is, when you combine the data sets together is when you actually start to get the best estimates of the incidence of disease, just showing the value of linking up different data sources, even for quite a simple question and that value is even more so, if you're addressing complex questions.

I will finish in a minute but I think I would say, health records are important because they can help us answer really, really difficult questions and help us to embrace this complexity of what it is that makes some people unhealthy and others healthy. I think we need to focus on making sure we're doing better research, not just more of it. What do we need for that? I think we need expertise. It is surprisingly difficult to do this research, it looks easy until you do it. We need to keep developing better methods, we need to make sure we're using the best quality data we can get, and of course, most importantly, remember always that we're using people's data, we're using very, very personal data. Peoples health records are the most personal data that they have and we need to look after it, make sure it's secure, make sure people feel reassured about the confidentiality and public trust and have this as a shared exploit. We're trying to improve human health and your data and everyone else's data and my data can help scientists to achieve that, and that's what we're trying to achieve here. To do that, we need to keep this trust going and keep people's data secure and safe and make sure it's a joint endeavour. Thank you.

[END OF TRANSCRIPT]