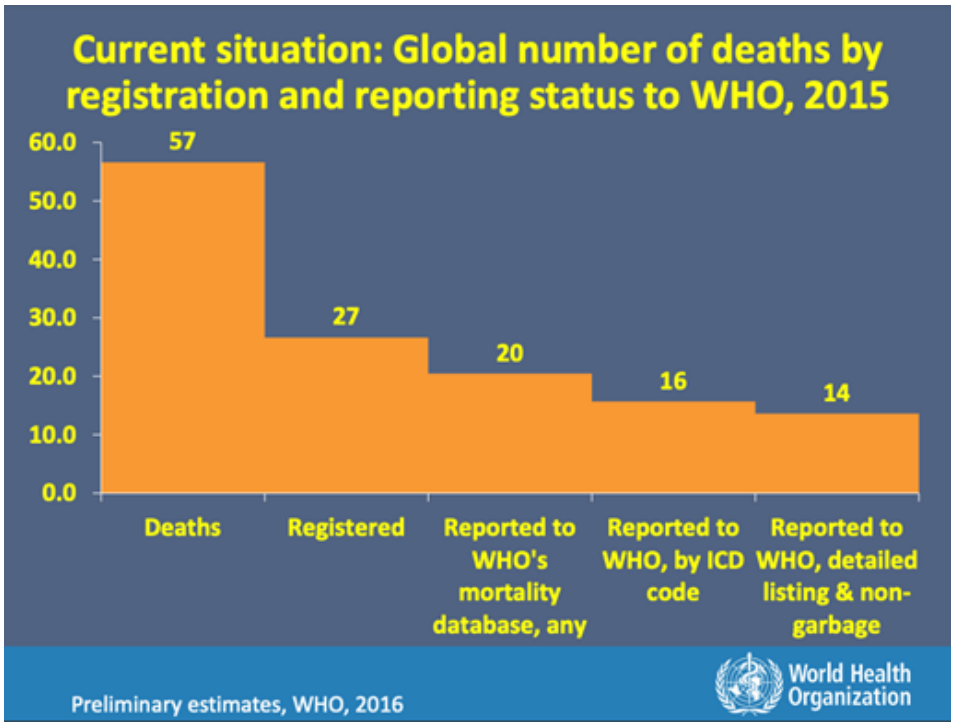
What is Health Data Science?

Hello, my name is Dr Dimitra Blana, and I am a member of the Aberdeen Centre for Health Data Science.

The aim of the Centre is to use data to address the big challenges in health and healthcare. In this video, I will try to give you an idea of what health data science is, using some examples of the work we do at the Centre.

Have a look at this graph. This shows the global number of deaths in millions in 2015, and estimates of how many of them were registered and reported to the World Health Organisation. This graph suggests that three quarters of all deaths worldwide were not reported in terms of medical cause of death. In 2021, this is estimated to be half of all deaths. This is an improvement, but it still leaves substantial numbers of people's lives undocumented. And of course undocumented deaths are not randomly distributed around the world but occur among poor people in poorly resourced areas of less developed countries, where civil registration and vital statistics systems are incomplete or absent. This is an example of data missing, and what we need is strategies to fill these data gaps.

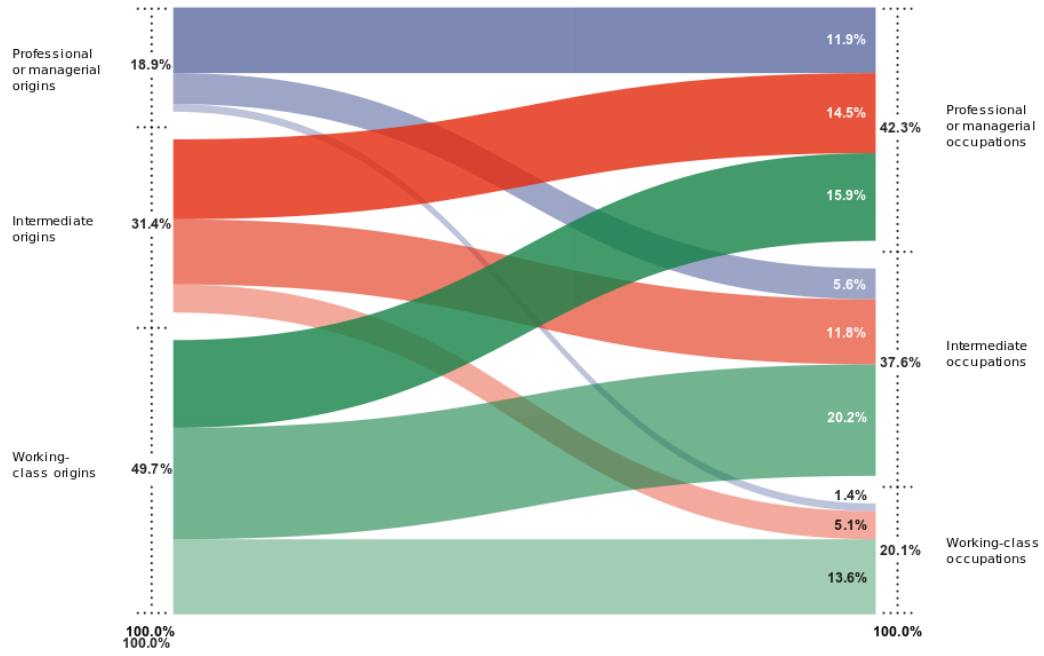


Currently, the only realistic alternative to medical certification of deaths is verbal autopsy. This is a survey-based method in which trained fieldworkers gather information from final caregivers on signs and symptoms of the deceased prior to death and from these data,

probable medical cause(s) of death are separately interpreted. At the Aberdeen Centre for Health Data Science we have developed verbal autopsy systems over many years to understand burden of disease in otherwise unregistered populations, filling important gaps on the global health data map.

Closer to home, we want to understand more about our local population and their needs. A great tool for this is birth cohort studies, that follow large numbers of people from birth and across their entire lives. One of these is the Aberdeen Children of the 1950s. These are 12,150 people born in Aberdeen between 1950 and 1956. The data collected from this group is used to answer a wide variety of questions. One of our researchers has looked into their social mobility. Have a look at this graph. Here, people are grouped by their trajectories of origin and destination. Origin is their father's socio-economic group based on their occupation, when the children were 7-11 years old, so between 1962 and 1964. The destination is the person's own socio-economic group, at age 45-51 in 2001.

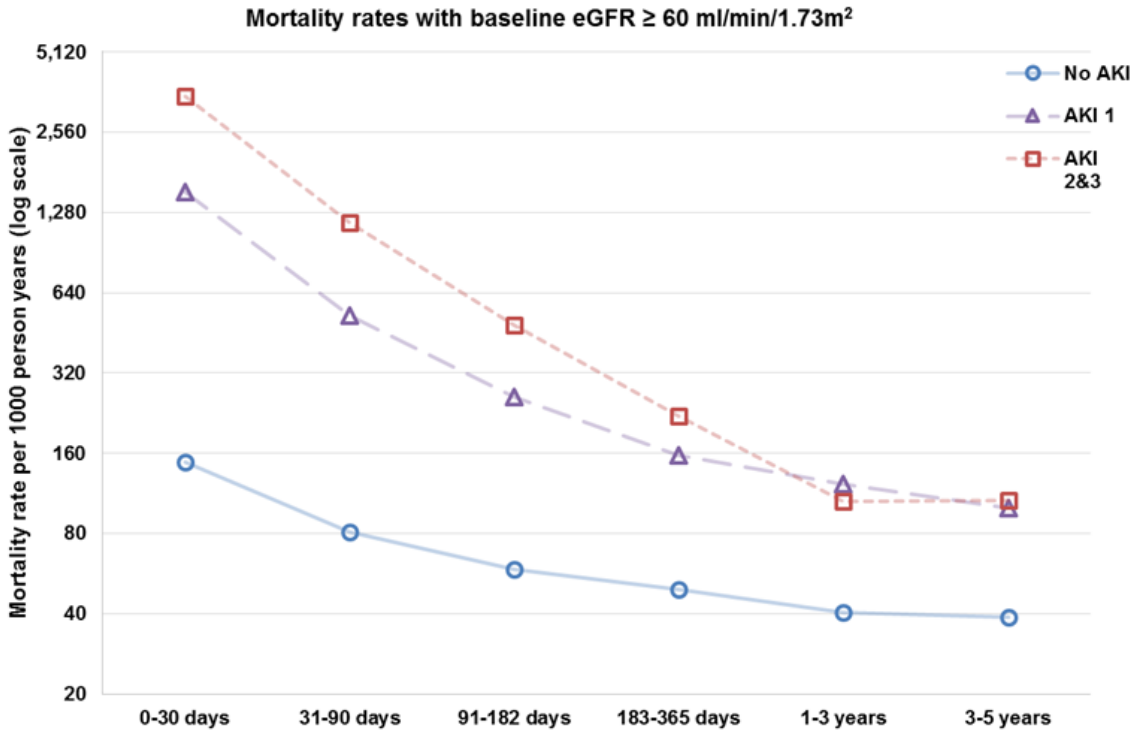
Figure 1. Mobility flows from origins to destinations in the ACONF cohort study (n = 6555).



I love this graph. It clearly tells the story of social mobility. It shows, for example, how it's more likely to end up in the top occupations if you are already there or nearby. But it shows you something else as well. It shows you how the structure of occupations itself has changed. So professorial or managerial occupations more than doubled in size from 19% of

all occupations in the 1960s to a staggering 42% in 2001. It is a great example of a visualization that is both informative and beautiful.

The amount of data produced by our healthcare systems is constantly increasing. Using this data to improve patient care is exciting. But it's important to look at the data carefully. Acute kidney injury is a sudden episode of kidney failure and it is associated with a very high mortality rate. However, if we focus on a single number (like the mean/ or median survival) we miss a very important point, which is that the excess mortality risk changes over time.



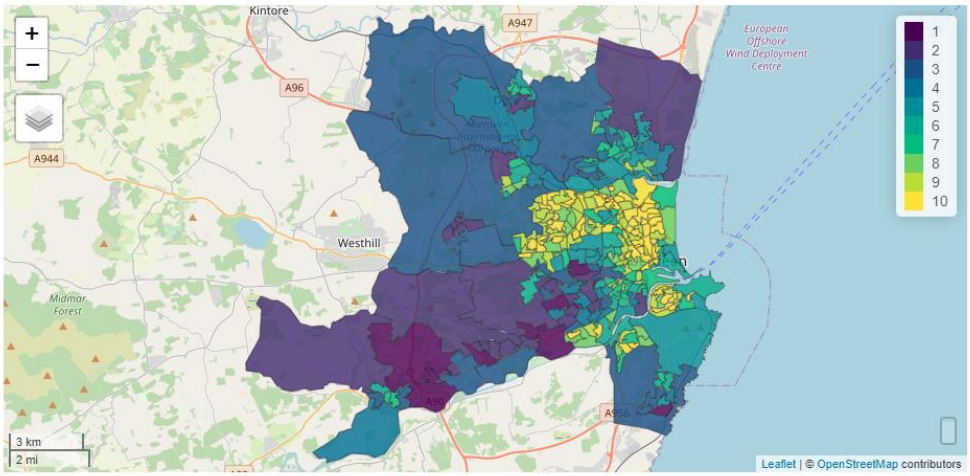
One of our researchers put together this graph that shows mortality rates over time, for people without acute kidney injury, with mild injury (stage 1) and more severe kidney injury (stages 2 and 3). What you can see is that people with acute kidney injury are indeed 8-16 times more likely to die in hospital than those around them, but that terrible prognostic message changes massively among those who are still alive 1 year after hospitalisation. The excess risk is now more like twofold (and most of it is due to other health problems). The implications are vast when you consider the type of conversations people are having with patients in hospital, without modifying the message for those people who are obviously getting better, or without modifying the information people receive in a follow up clinic compared to in-hospital. This impacts policy. Our researchers are working on rewriting clinical guidelines for people with acute kidney injury more strategically based on follow up

time. So, targeting people for intensive monitoring in the first few months, then relaxing monitoring and providing more reassurance as time goes on.

At the Centre, we work closely with colleagues from our National Health Service board, NHS Grampian. The Covid-19 pandemic made this collaboration our number 1 priority. The pandemic forced rapid changes across the NHS to care for COVID-19 patients, while balancing this with the care it has always provided. At the Centre, we developed a variety of modelling and monitoring tools. The aim was to provide timely information on the local progression of COVID-19 to NHS Grampian to ensure there were beds, staff, and equipment where they were needed for both COVID-19 and non-COVID-19 patients.

One of our researchers developed a social distancing score to identify areas that might be at greater risk of transmitting COVID-19. The score combines three measures: percentage of people with essential jobs; percentage of people living in overcrowded homes; local population density.

Here is this score calculated for the city of Aberdeen. It's a very effective visualisation: you can clearly see the differences across the city. I want to highlight that this was put together using publicly available data. But it's the combination of datasets and the presentation that has turned this into such a valuable tool. NHS Grampian used these maps to inform COVID-19 testing provision.



There you have it: a few examples of how researchers at the Aberdeen Centre for Health Data Science use data to improve health and healthcare. We combine data from different datasets, we analyse, model and visualise it. This enables us to notice what data is missing from the global health data map, understand more about our populations and their needs, improve patient care, and respond to health emergencies.