RAMSEY SYSTEMS

Harley Johnson

Data Analyst

Harley graduated from Liverpool John Moores with a first in Mathematics and has two years experience as a software developer. She is currently working on a problem to understand and illustrate information architectures, using simulations.

harley@ramseysystems.co.uk

I've been working at my current job for about 3 months now and I'm still never sure how to tell people what I do. My profile on Ramsey systems is as above but I normally tend to tell people…

RAMSEY SYSTEMS

Harley Johnson

Software developer working on creating synthetic health data through simulations

harley@ramseysystems.co.uk

I am a ‘software developer working on creating synthetic health data through simulations’. Which normally elicits a very confused look even among people in the relevant field.

RAMSEY SYSTEMS

Harley Johnson

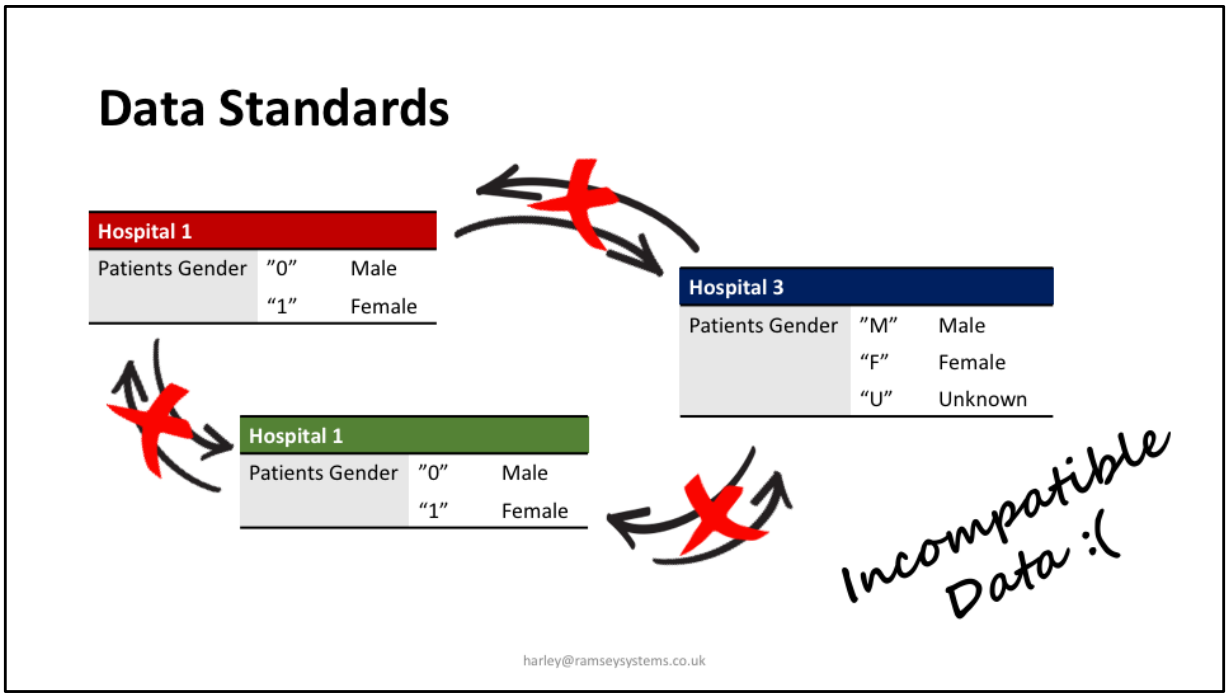
I write code that makes made up data…

harley@ramseysystems.co.uk

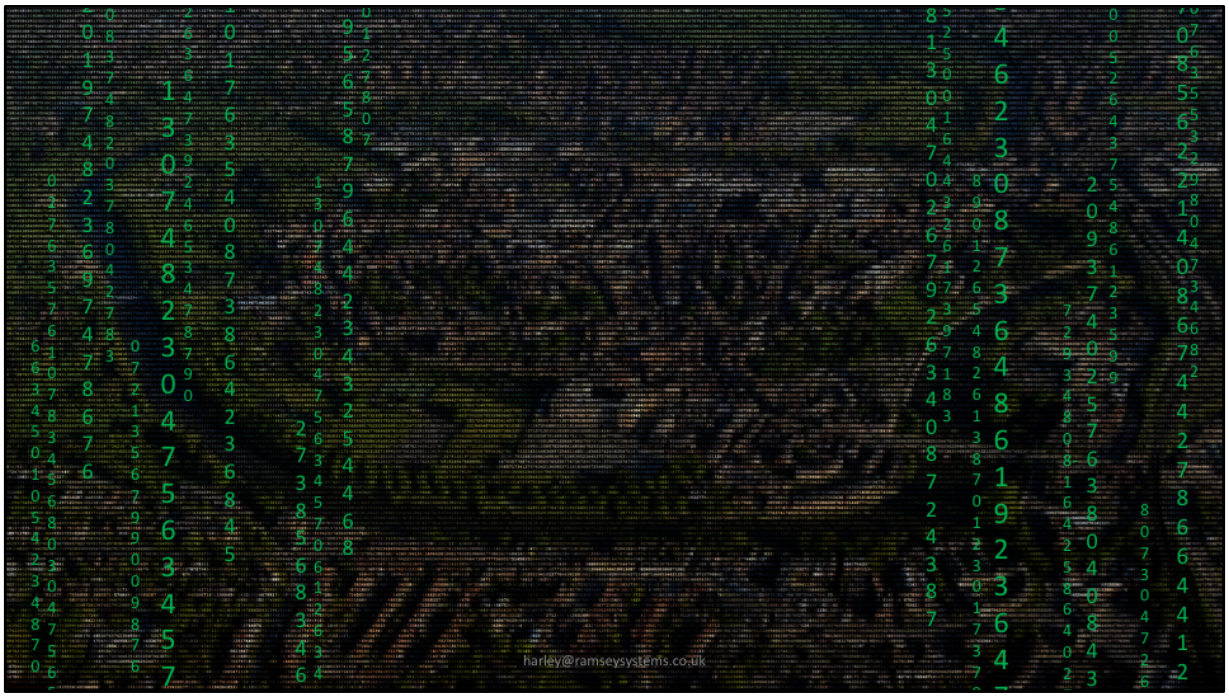
In the end it often boils down to me saying ‘I make stuff up’. But I’m hoping I can give you an idea in this short talk about what I really do and give some food for thought on Simulations.

Synthetic data is great. Having a whole bunch of made up patients with fake names and addresses and relatives and health problems means you don't need to get permission to use the data. You've probably heard about the issues that start arising when keeping real data safe and even anonymised data risks becoming deanonymised. Companies face a lot of trouble when anything gets leaked.

But data is needed by a wide range of organisations for a spectrum of uses.



The data that we produce might be used for testing new software or looking at data standards which is a huge problem in health but one for another time. Today, however, I want to talk about how we use simulations to create synthetic data and why on earth we want to do it.

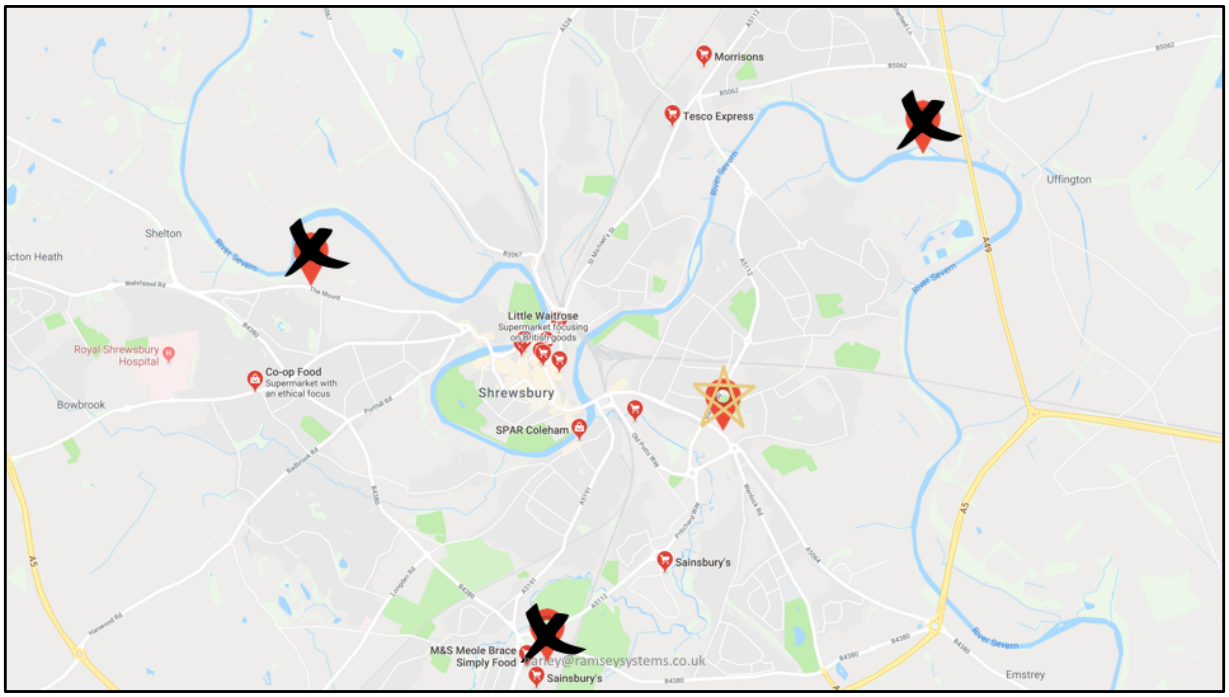


I’m not working on creating a version of the Matrix, simulating an entire reality is a bit beyond my capabilities, the point of what we are doing is to facilitate projects in health. For example, if there is a problem in the local area with an aging population, we want to look at projects and proposals and provide insight into what will help their ‘measurable outcomes’ which is essential in health and other fields to prove their work is worthwhile.

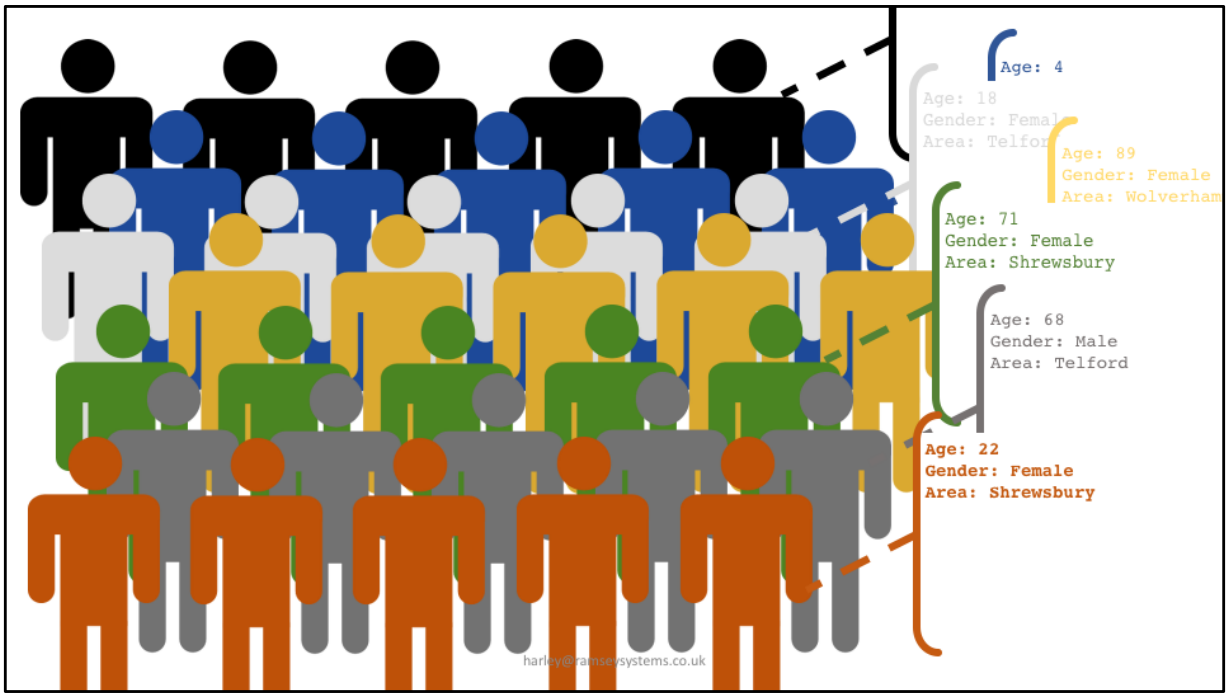
The idea is to try and predict the results of the investments made in healthcare.

Let me give you an example of how this is being used in retail to inform management decisions.

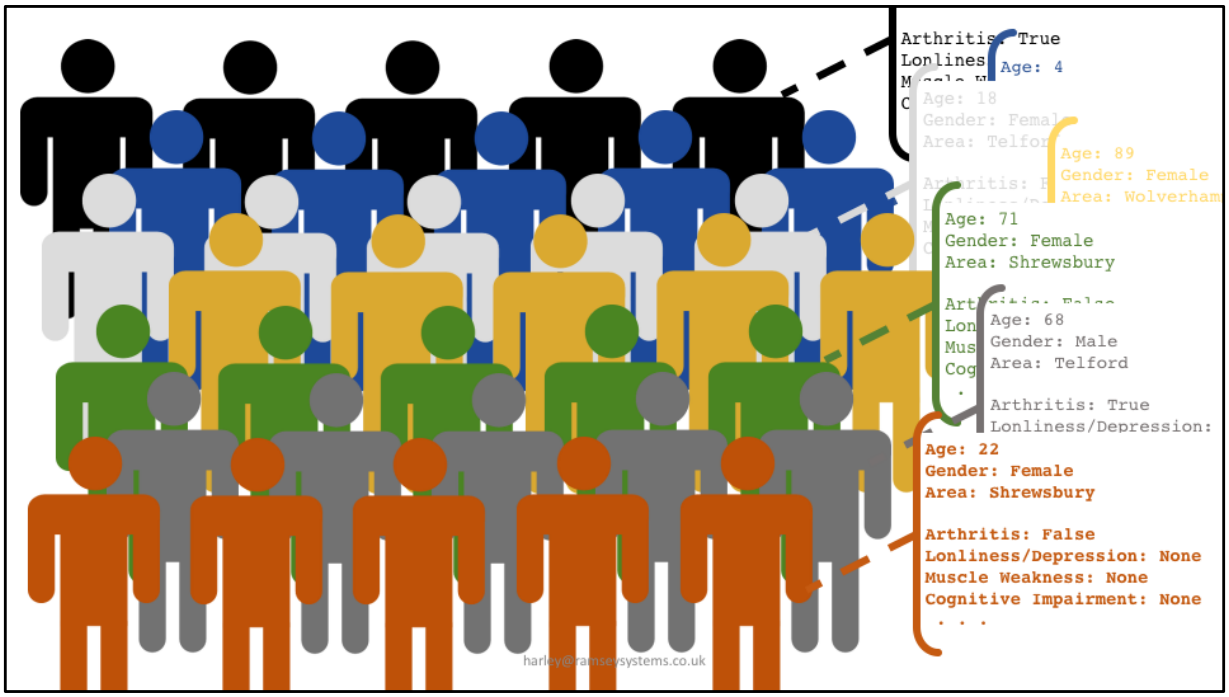
Stores want to know where their customers go, what conversations they are having, anything that influences them to start purchasing. They can't get hold of all of this individual data (though some companies are closer to getting it than we might like) so one way to try to predict the best location for a new store, for example, would be to simulate a population of fake people and see what they do.



You can use all the data you have about the general population, ages, gender, shopping habits to create a population similar to the one that already exists. They will make decisions about what to buy based on their traits. Running a simulation multiple times with different new stores will inform the decision on where would be best for the store to go.



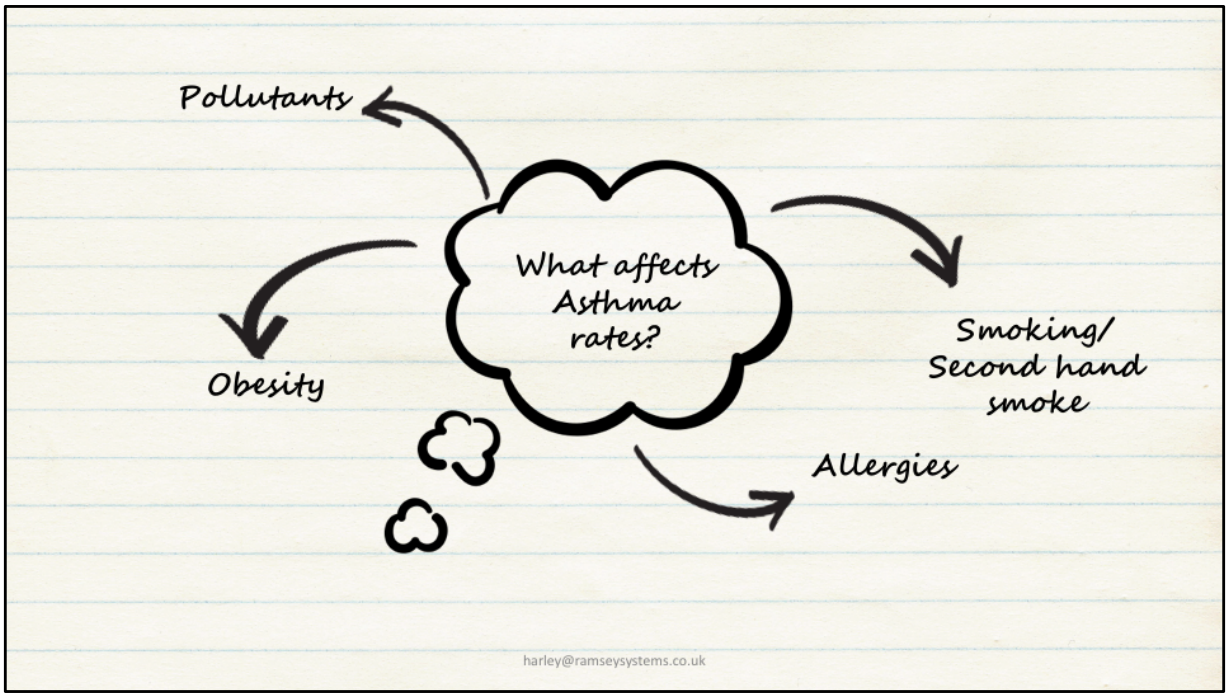
The way to start building these simulations is to start with the basics. What is our populations age, gender, race, socioeconomic status? From this our simulations can back up some very common-sense things. That hiring more paediatricians will not help the aging population or in the retail example, opening a new toy store may not work out too well either. At this point I have spent quite a bit of time programming a system that can tell us things we already knew. This might not seem very useful. When it does become useful however is when you start looking at more variables.



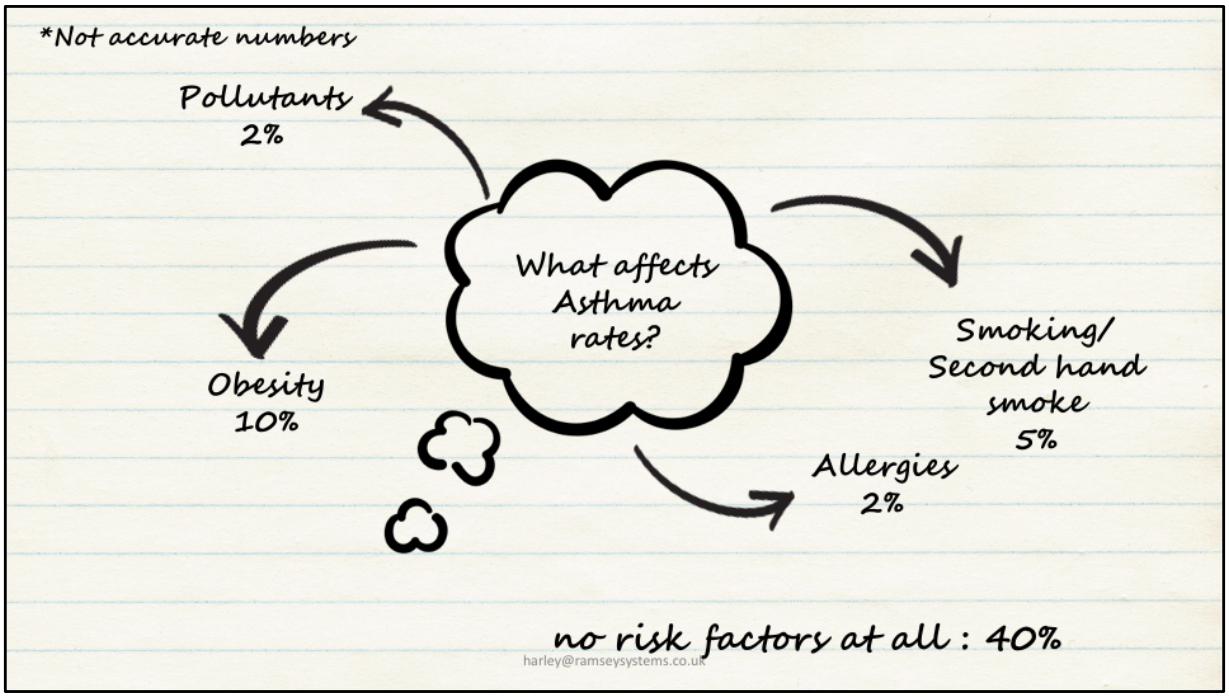
'How likely is this person of this gender, at this age in this area to fall?' turns into 'how likely is this person of this gender, at this age, in this area, with arthritis and a history of falls, in a single person household to fall?'
If you have enough data to base these simulations on they should be a good predictor for what will happen. We can't use everything however…

… it would be fantastic to copy and paste the universe, make our little change and see what happens next. Trial and error to get the best result. That would be great for our purposes. If you could do that however we start going into moral dilemmas and an existential crisis about how it's more than likely than not that we are simulated too…

So, what we do is look at our current problem and hope to get a reasonable prediction from RELEVANT data. That means we need to decide what is relevant. So, what affects whether someone gets asthma?



And which of these might have the most effect?



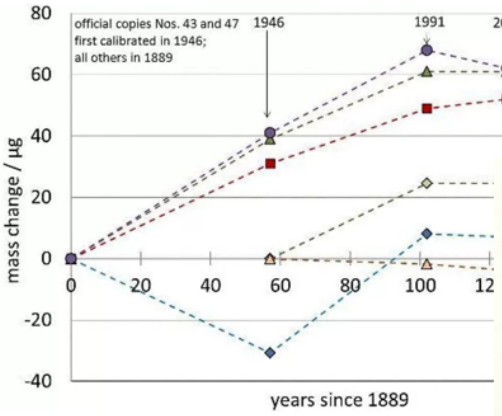
I'm asking you because this is exactly what we had to do when we started. We make our best guess and then we try to figure out how wrong we are. Actually, there was a lot of researching and talking to clinicians too but there was a fair bit of guessing.

We use our best guesses and start running simulations with just the chosen attributes.

When we disregard all of the things that we don't care about, we do sacrifice accuracy and this can be a serious concern in healthcare. There is a risk that if the simulations suggest the wrong thing and a decision is made based on them then we may be putting lives at risk.

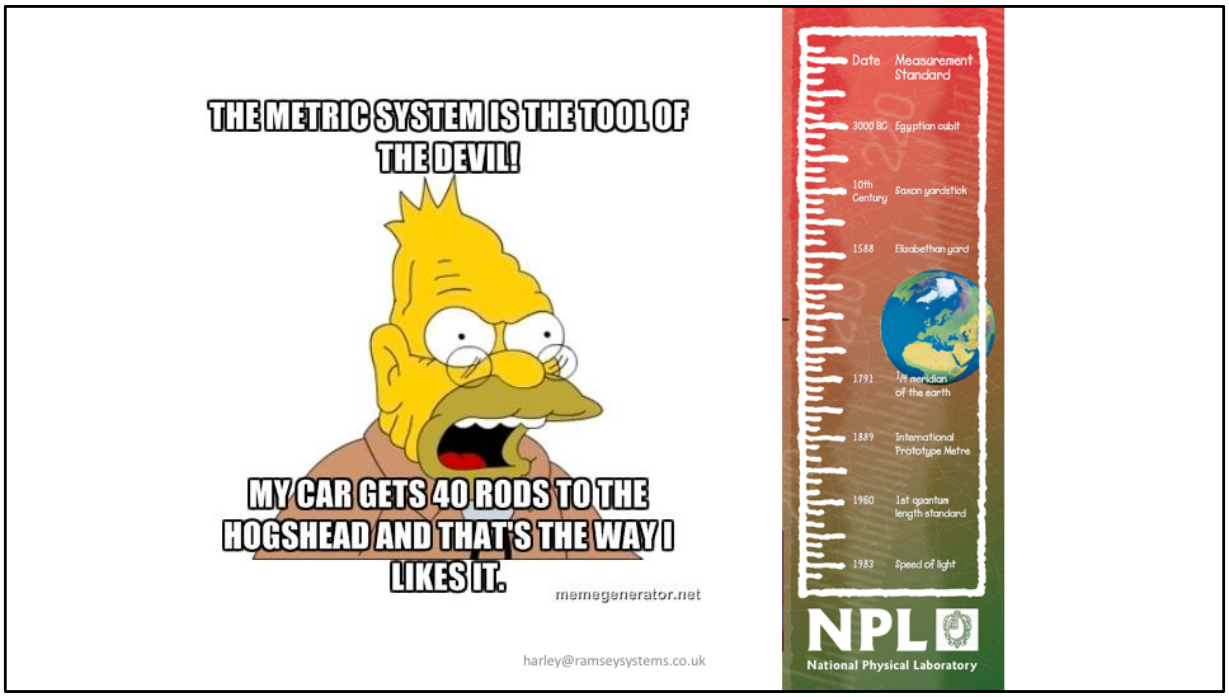
Throughout the whole process we always keep in mind that we will get it wrong. When you work with statistics you will always have some degree of error.

The kilogram prototype in France has been losing weight, so it's time for a new definition

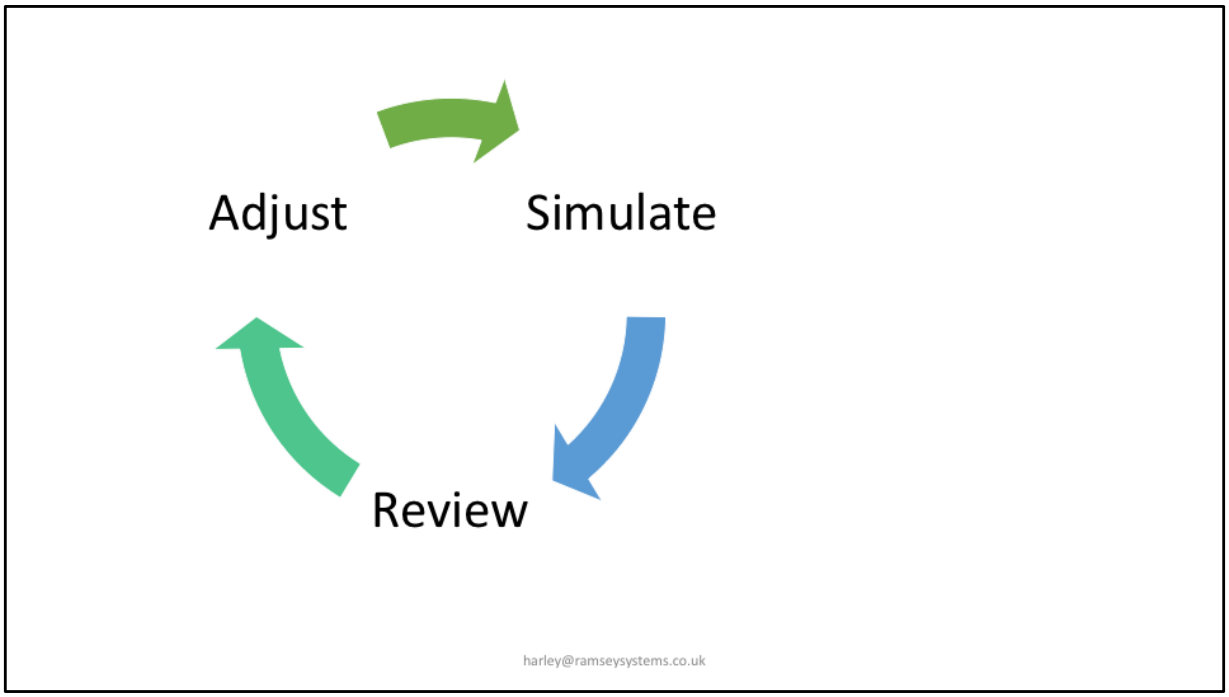


harley@ramseysystems.co.uk

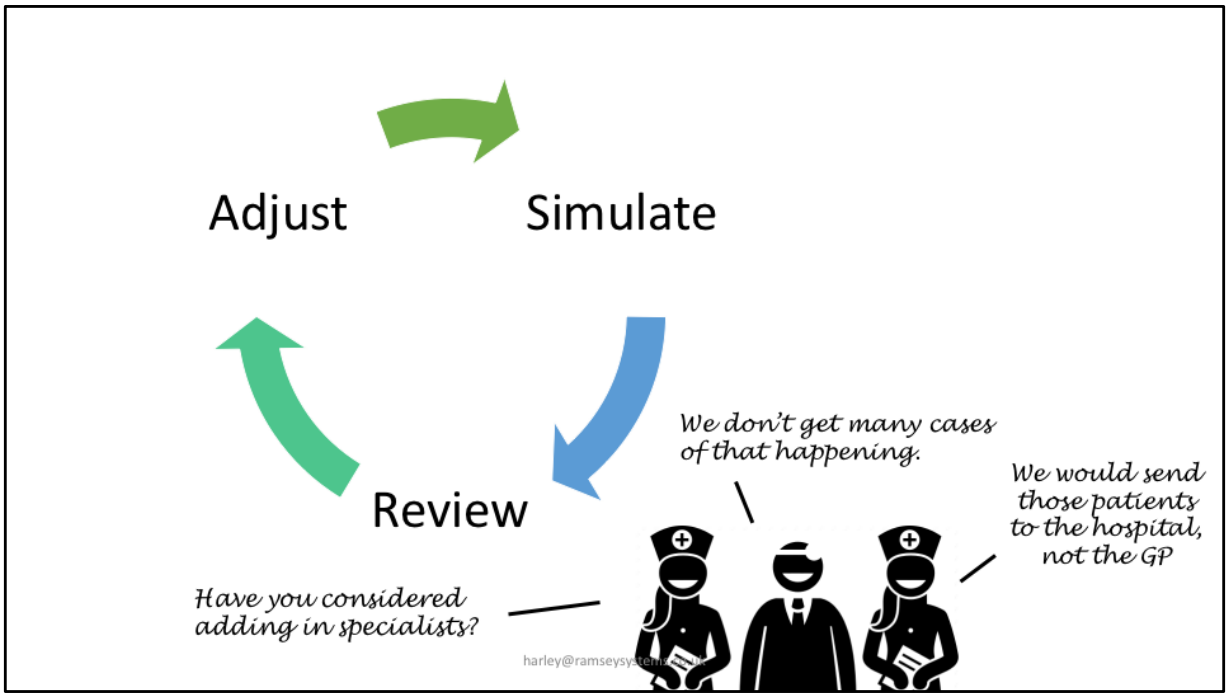
Even things we might think of as fixed can be improved. So instead of aiming for the 100% guaranteed answer and sticking with it, you accept that there is a better answer out there and aim for better than the last answer we had.



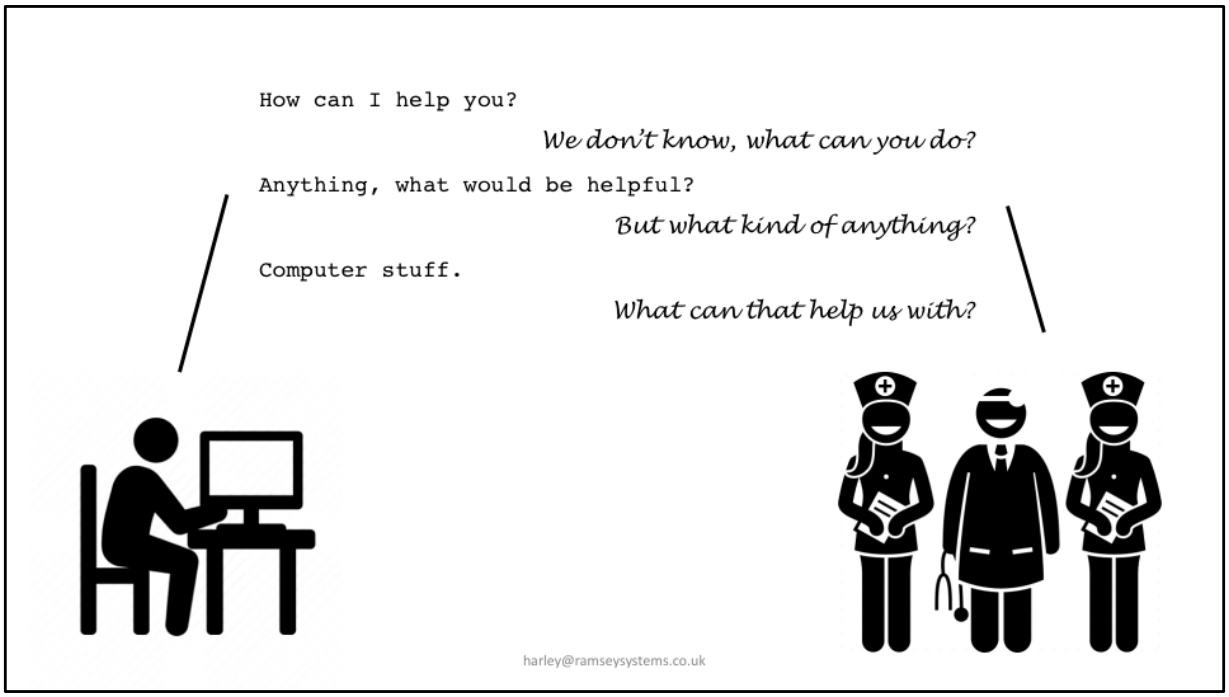
Measuring distance with the meters is better than yards is better than cubits…



We iterate through our process as many times as we can, running simulations, reviewing them, adjusting them, running, reviewing, adjusting. If we can get more accuracy then we keep going. If our model is not mimicking the current population then we are looking at the wrong data so we adjust it and run again. If we aren't producing the most useful data to look at the measurable outcomes, we need to change what we are measuring so we adjust it and run again.



All the while we work alongside the clinicians and experts as much as we can when we create simulations. They give us the proper context that informs the iterative process.



This work is about helping inform decisions in healthcare but also about provoking discussions about health data that need to be had. If we don't know what is the biggest factors are affecting asthma, affecting obesity, affecting falls in the elderly then how can we look at preventative measures? If we don't know what doctors need to assess a patient's health then how can IT providers give a useful service? If we don't test how the health systems work then how can we know what needs to be done to connect them together so that you can go to Portugal or Greece or Belgium and the doctors get your medical data so they know not to give you a drug that you are allergic to. These sorts of conversations are what Ramsey Systems is interested in is as a whole. Simulations are useful in their own right but they are also a way to spark the discourse is an interesting way. When I manage to explain what it is that I do anyway…