

**The only independent UK movement of patients, relatives and carers focussed on the use of patient data to save lives and improve outcomes**

## **Supporting the use of synthetic data help to make research more secure, as a key component of Trusted Research Environments (TREs)**

**use MY data** endeavours to highlight the many benefits that appropriate usage of healthcare data can make, to save lives and improve care and well-being for all.

**use MY data** supports and promotes the protection of individual choice and privacy in the sharing of healthcare data to improve patient treatments, outcomes and experience.

**use MY data** aims to educate and harness the patient voice to understand aspirations and concerns around the use of data in healthcare delivery, in service design and improvement, and in research, aimed at improving clinically informed patient decision making, treatment and experience.

### **Overview**

In a world where there is often significant risk-aversion to the sharing of patient data, there has been increasing interest in the topic of synthetic data, sometimes called simulated data.

In essence, synthetic data is designed to look closely like the real data it represents, but is completely fictitious.

We describe below how we think synthetic data should be used to make research more secure.

### **What is synthetic data?**

If you were to draw a human stick figure, it would clearly represent a person. You would be able to count the number of arms and legs, and if it were drawn to scale you could estimate the height. It would have the same form as a human figure, the same basic body structure, and the same characteristics.

You could derive a fair amount of information from the stick figure about a human, and you wouldn't need to see a real person. By making the stick figure more detailed, adding eyes, fingers, toes, it becomes more realistic - but it is still simulated.

Synthetic data follows those same principles.

A synthetic dataset looks like real data, with the same structure, composition, data fields and similar data entries. Of course, synthetic data environments need to be transparent as to which methodologies they use to generate the synthetic data so that the features and limitations of the synthetic data are understood by researchers.

### **So how can synthetic data improve research?**

A significant benefit is that the synthetic dataset allows researchers who want to develop queries for very detailed health record data to be allowed to do so, safe in the knowledge that while the data looks and feels like the real thing, there is no danger of them breaching patient confidentiality.

It allows researchers to write computer code to examine (or query) the dataset, to develop complex data extraction queries, and to test these to make sure they run, and produce sensible results without directly accessing real patient data.

The queries can then be given to the organisations which hold patient data securely, who then run the query, inside their own secure environment, on the real data. The results of this are then sent back to the researchers as a completely anonymous set of statistics for them to draw meaningful health conclusions based on real data with the researchers never having had access to that real data

Where synthetic data could provide an increasingly useful contribution is in the representation of synthetic versions of the data held in Trusted Research Environments (TREs). Given recent commitments by [government](#) and [health agencies](#) to invest steeply in new TRE infrastructure, we would encourage the development of parallel synthetic data environments to facilitate data-driven research whilst reassuring the public that researchers are not directly accessing their personal data.

If you want to know more about how synthetic data has been used with a powerful example for cervical cancer, there is a useful blog by Lucy Ellis-Brookes, called [How do you do pioneering research without ever seeing the data?](#)

### **How our Members should adopt and use this Position Statement**

- We support the use of synthetic data as a tool to better understand complex data, particularly alongside a Trusted Research Environment (TRE)
- For any organisation that you are working with that is developing or using a Trusted Research Environment, make sure they are aware of this Position Statement
- Find out whether the organisation is working in this way, and if not, highlight the areas which they might want to look at
- Let the Secretariat know about any concerns or comments that the organisation may have, or that you have
- If you need advice on taking this forward, contact the Secretariat and we will either help you directly, or put you in touch with another Member who can help

# The only independent UK movement of patients, relatives and carers focussed on the use of patient data to save lives and improve outcomes

---

## Our vision

Our vision is of every patient willingly giving their data to help others, knowing that effective safeguards to maintain the confidentiality and anonymity of their data are applied consistently, transparently and rigorously.

## About use MY data

- **use MY data** is a movement of patients, carers and relatives.
- **use MY data** endeavours to highlight the many benefits that appropriate usage of healthcare data can make, to save lives and improve care for all.
- **use MY data** supports and promotes the protection of individual choice, freedom and privacy in the sharing of healthcare data to improve patient treatments and outcomes.
- **use MY data** aims to educate and harness the patient voice to understand aspirations and concerns around the use of data in healthcare delivery, in service improvement and in research, aimed at improving patient decision making, treatment and experience.

## What we do

- We promote the benefits of collecting and using patient data to improve patient outcomes with sensible safeguards against misuse.
- We work to bring a patient voice to all conversations about patient data.
- We have developed the Patient Data Citation, which acknowledges that patients are the source of the data. Details are available [here](#).
- We act as a sounding board for patient concerns and aspirations over the sharing and using of data in healthcare and health research.
- We provide learning resources for patient advocates on patient data issues, including:
  - Hosting events for patients and the public, focussing on patient data topics
  - a library of resources of data security, consent
  - narratives from individuals about the value of collecting and using patient data.
- We advocate public policy that supports the effective use of patient data within appropriate frameworks of consent, security and privacy, and with the aim of providing benefit to patients and their health care services.

---

[www.useMYdata.org.uk](http://www.useMYdata.org.uk)  
[join@useMYdata.org.uk](mailto:join@useMYdata.org.uk)  
[@useMYdata](https://twitter.com/useMYdata)