The past, present and future of OpenSAFELY

> OpenSAFELY has been a breakthrough in how we responsibly use the UK's health data for public good. By bringing together expertise in software engineering, epidemiology and a deep knowledge of the UK health system, the team has delivered products and services being used by hundreds of researchers to understand how to improve human health. Wellcome is proud to be a long-term supporter of the OpenSAFELY team. They continue to demonstrate how to encode trust into technology and policy, and deliver useful technology that encourages transparency and open working by design.
>
> – Tariq Khokhar, Wellcome

> I've been using GP data for research since 1999. These NHS records are incredibly powerful, but for decades they have been hard to access securely at scale. In OpenSAFELY, finally, there is a platform that enables access, at national scale, with clinical information flowing in near-real-time, while maintaining data security and patient privacy.
>
> – Liam Smeeth, London School of Hygiene & Tropical Medicine

Contents

Foreword

Ben Goldacre, Director and Seb Bacon, CTO
Bennett Institute for Applied Data Science

England's GP records have extraordinary power to do good, when they are brought together at the scale of the whole population. But until OpenSAFELY, this power had never been unleashed. The reasons are simple.

GP records contain billions of rows of detailed information: every diagnosis, every treatment, every test, and more, for every citizen in the country. This data can turbocharge research. But those same records also contain, by definition, the most confidential medical secrets for every one of us.

This book describes how we square that circle: how we can give users efficient and productive access to everyone's data; while also protecting everyone's privacy.

Solving these problems required a new approach. Our community had to create new working methods, then implement those ideas into working tools and services. This, in turn, required deep, creative collaboration between researchers, software developers, policymakers and innovators.

In the past, "data infrastructure" meant beige boxes in large buildings. In the 21st century, data infrastructure is code, and teams with skills, coordinated in networks.

Building this kind of infrastructure is harder than buying a box. But it is much more exciting, and critical to the future of research!

Introduction

OpenSAFELY is a free, award-winning digital platform that helps researchers analyse large, sensitive datasets, safely and securely. It's a huge, highly efficient, highly productive electronic health records platform, built with the NHS, at low cost.

It has a huge user community, with analysts from 32 organisations running 181 projects. Our users have published 86 peer reviewed papers so far, with many more to come.

It operates at unprecedented scale, with access to 58 million full GP patient records inside the secure data centres of TPP and EMIS, the two main suppliers of electronic health record services to GP practices. It also provides linked access to other important datasets including:

- Hospital Episode Statistics (HES)
- Secondary Uses Service (SUS)
- Emergency Care Data Set (ECDS)
- COVID-19 Therapeutics
- ISARIC (International Severe Acute Respiratory and emerging Infection Consortium)
- The UK Renal Registry
- The Office of National Statistics (ONS)
- The COVID-19 Infection Survey.

Starting in the frantic early days of the COVID-19 pandemic, OpenSAFELY was built to solve one of the biggest problems in medical science: overcoming the inherent tension between the needs of researchers who want to use health data for science, and the needs of patients, who expect their personal information to be kept secret and secure.

OpenSAFELY was designed to balance the needs of both sides. It helps researchers use data to generate insights, while maintaining patients' personal privacy. It's open, but it's safe.

This booklet sets out:

- how OpenSAFELY works
- the story of how it was made during the pandemic
- what results it has brought about
- some stories from researchers who have used it
- some thoughts for the future.

58 million patient records
86 published peer reviewed papers
181 projects
32 organisations

1

How OpenSAFELY works

Previous efforts to extract public health data into a central database – and then disseminate it to multiple locations – caused huge public disquiet, and 3 million people chose to opt out of their records being used in research. OpenSAFELY uses technology to mitigate those concerns. None of the raw patient data ever leaves the secure data centres where it already lives. OpenSAFELY provides a secure way for researchers to submit questions, run them against the data, and get back aggregated results about groups of people.

To use OpenSAFELY, researchers first prepare the raw GP data in a form where it can be used in an analysis. They do this by writing code in Electronic Health Records Query Language (known as ehrQL for short – it rhymes with 'circle') – that helps to extract and shape data from the available data sets.

When they've prepared their dataset for analysis, they then write analysis code (in standard languages like Python, R or Stata) to produce graphs and tables, or to run statistical tests.

All the code users write is made up of individual units called actions, and those actions are organised into a pipeline. By working in this way, we ensure every users' code is well organised.

OpenSAFELY generates dummy data, so that researchers can test their assumptions, and make sure their code is likely to work, all on their own computer. This is a critical design feature for privacy: it means that users don't interact directly with real patient data when writing their code. Users can also import their own dummy data, if they prefer.

Codelists are collections of short codes that match specific clinical terms in the data – they're a useful tool for designing research projects. We've built an online tool called OpenCodelists to help researchers create and share their codelists.

Once the code has been run on dummy data, researchers select a button to submit it for running on the real data. All code submitted to run in OpenSAFELY must first be made available online using GitHub, along with contextual information about the project.

Each package of work is known as a job. OpenSAFELY automatically keeps track of all the jobs, including every action being run, what it does, who requested it, and when it happened – there's a live public dashboard on the web at jobs.opensafely.org, where anyone can keep an eye on what's happening.

OpenSAFELY then runs that research code automatically, at arm's length, inside a secure environment, meaning that researchers never need to access sensitive patient data directly.

When each job is complete, researchers can see summary results (mostly in the form of tables and graphs) inside the secure environment, using a tool called Airlock. Inside Airlock, users can see log files (useful for debugging and problem-solving), and data outputs, which must not contain any identifiable information. Airlock has automatic controls to restrict data (such as very large files, or certain file types).

For some (but not all) projects, the researcher might want to move selected outputs outside the secure environment, perhaps for use in a draft paper. Before that happens, we have to make sure that nothing leaving the secure environment could potentially identify any individual patients – what's known as disclosive information.

This is where our output checking service comes in. After a researcher requests that some outputs be released from the secure environment – some graphs, or results tables – then at least two trained and qualified humans will manually check that they aren't accidentally releasing anything that could possibly contain any information about any individual, even an anonymous individual.

Those approved outputs are then moved to a secure job server, outside the secure environment, from where they can be released to the outside world.

The output checking process is also fully audited, including requests for changes made by output checkers, and responses from the researchers. It's called 'Airlock' for a reason: it's a secure place where outputs can be viewed, understood and output-checked. Some of those outputs will be released, but many aren't.

Some important things to keep in mind:

- The private patient-level data never leaves the secure data centre; only aggregated outputs.
- The researchers never get direct, unconstrained access to interact with private patient data; instead they develop their code using randomly generated dummy data; and their finished code is then run against real patient data.
- OpenSAFELY was designed to encourage scientific rigour and best practice. There are a few hoops to jump through, but they exist for good reasons: to ensure the safety of the outputs, and to help users write high-quality code that's capable of running on unprecedentedly huge national datasets.
- Newcomers using OpenSAFELY for the first time are given a helping hand from experienced co-pilots.
- We have strict information governance policies, and a team of in-house experts to make sure everyone sticks to them.
- OpenSAFELY was created in close collaboration with teams from the main private sector suppliers of data services for GP surgeries – TPP and EMIS – and in close collaboration with research users who have deep expertise in working with electronic health records.
- NHS England is the Data Controller for the whole service, and the GP practices whose records we are using remain the Data Controller for patients' records, with our tools integrated onto their systems.
- All the documentation for using OpenSAFELY is published on the web, so anyone can start learning how to use it.

How OpenSAFELY is different

We don't give researchers huge extracts of pseudonymous data, either direct to their computer or inside the remote secure environment, because we don't believe that pseudonymisation is secure enough. It's often possible to identify individual patients, even in pseudonymised data.

The dummy data that OpenSAFELY generates is a unique and important feature – it means that researchers can check that their code works as expected, before using it with real data. This encourages a more hypothesis-driven approach, and discourages mid-research iterations that could potentially introduce biases and affect research findings.

OpenSAFELY has earned the trust of all the big names in medicine and medical privacy, including the British Medical Association (BMA), the Royal College of General Practitioners (RCGP), their Joint GP IT Committee, Citizens' Juries and privacy campaigners such as medConfidential.

OpenSAFELY is designed to be open:

- all the code that makes OpenSAFELY work is in the public domain
- all the code that researchers write is open, making it easier for others to re-use as part of their research
- there's a live dashboard on the web, showing the current status of every job that's running, or has been run before.

Openness is part of the deal: researchers can develop their code privately, but have to agree to making it open once any results are shared, as a condition of using the service. We've won awards for our commitment to openness. We think it's very important, because it means:

- anyone can scrutinise the work, to check a researcher's findings,
- and anyone can re-use the work, as part of their own studies, which makes the science faster and more efficient.

Openness is part of the deal

Co-pilots give newcomers a helping hand

Regardless of the means of accessing electronic health records (EHR) data, working with it can be hard. While we designed OpenSAFELY to be as simple as possible, and while we want its users to work independently, we're conscious that its novel and highly secure approach comes with a learning curve. So, in addition to the full set of detailed documentation, examples and walkthroughs to help new users get to grips with it, we have designed and developed an end-to-end onboarding and support service: the OpenSAFELY Co-Pilot Programme.

So far, it's been very successful. The pilot phase consisted of 50 projects (60+ users), from across 22 academic and government organisations, enrolled between July 2021 and April 2023. Already, these projects have:

- made over 10,000 job requests (equating to over 50,000 jobs)
- had over 5,000 files released from the secure server
- produced a huge number of peer-reviewed publications, with more still in the pipeline.

Once enrolled in the programme, external users (pilots) are strategically paired with an experienced in-house OpenSAFELY researcher (co-pilot), who helps them understand the OpenSAFELY philosophy, learn the various software tools, and work through each step required for a successful project.

Every new user is different. For example: some are more familiar with software tools like Git, and others have to learn how to use them from scratch. So the co-pilot's role varies a lot. In addition to direct support via email and Slack, co-pilots organise regular calls where pilots and co-pilots set goals, discuss progress made and resolve any issues. Co-pilots also support pilots through sessions on specific topics, such as those on paired programming, implementing quality assurance steps and statistical disclosure control. But of all the support that co-pilots offer, one of the most important aspects is a calm, friendly face: moral support is as important as technical support.

The support that pilots receive from co-pilots is initially very intensive, but fades over time as pilots start to gain experience, and make use of the available community resources (such as the OpenSAFELY discussion forum and the #opensafely-users Slack channel). But co-pilots don't just disappear overnight. Additional support is offered where needed until project completion (and even beyond). This is largely in terms of continued output checking and manuscript review, but also through ad-hoc meetings which will be offered in order for co-pilots to check progress and to help with any issues that may still arise. Often, even after publishing a paper, a co-pilot and pilot will keep in touch to discuss new ideas and collaborations.

Co-piloting isn't a one-way thing, either. The programme has enabled OpenSAFELY to learn from its users, use their feedback to drive and prioritise development, and thus increase the productivity of the platform. As a result, OpenSAFELY has been able to build an analysis platform that works for more people, faster.

> For me, the co-pilot programme was so important to the on-boarding process. My co-pilot was committed to me and was always on hand to help.
>
> OpenSAFELY user Rachel Seeley, Head of Analytics at PrescQIPP

57 co-piloted projects
91 pilots onboarded
17 co-pilots

Standard tools for data preparation, and "federated analytics"

Electronic health records are made for doctors and patients; and preparing those records for analysis is hard. This process of turning raw records into analysis-ready tables is a common challenge, and often the biggest part of any project using EHR data.

In the past, different teams working on EHR data would each approach the same data preparation job in their own esoteric way. This causes huge problems: even if you could access their code, it would be hard to understand. Furthermore, code for data preparation was generally written to run on one database, in one computer: so if the database changed, or the data moved, then often that hard work had to be thrown away.

We set out to solve those problems by creating new, bespoke, standardised, open source tools that all analysts can use for data preparation. These standard OpenSAFELY data preparation tools are an important feature of the platform's success.

Firstly, this standardisation means that all researchers are doing the same job in the same way, which is great for efficiency. Users can read and understand each others' standardised code more easily, so they can check its quality, and re-use it or adapt it if they want to.

Secondly, it means that users' data preparation code can run in any data centre where OpenSAFELY tools have been installed: it's no longer tethered to one specific database, in one specific data centre. That's critical, because it avoids problems like "vendor lock-in" to any one type of database. It also means that it's worth writing great data preparation code, because you'll be able to use it for a long time.

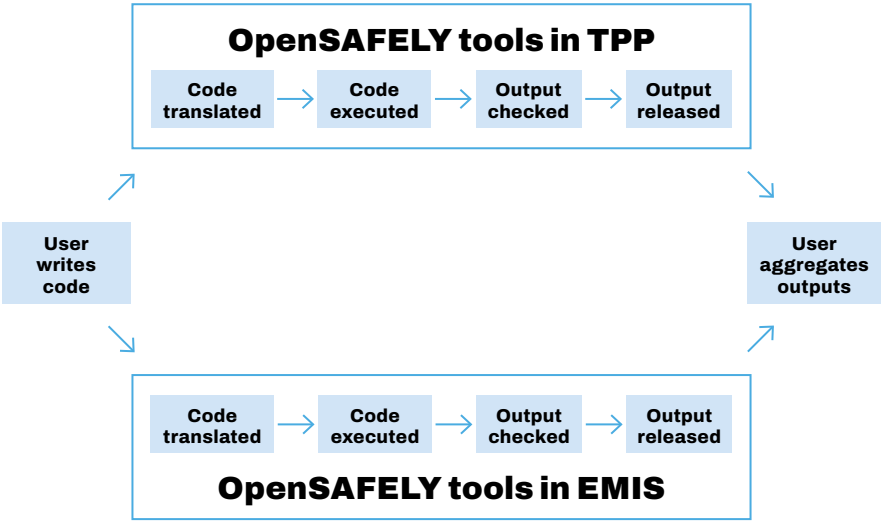
But it also helps us deliver one last key innovation from OpenSAFELY: federated analytics.

GPs' records about their patients are stored in two different places: the two data centres of the two major electronic health record system suppliers, TPP and EMIS. Each GP choses which system to use. We were under pressure not to extract and move large amounts of disclosive

GP data around to new locations and data centres, because of issues around privacy and transparency.

So instead of extracting data to give to users, we devised OpenSAFELY as a way for the code to go to the data. Our standardised data preparation code is a key part of what enables that “nice idea” to work in practice.

You can see this in the picture, below. The same data preparation code is written once, by users. It then travels out to EMIS and TPP’s data centres, where it is automatically translated by the OpenSAFELY tools into code that can work on that data centre’s specific machines and databases. The code runs to completion, and the analyst stitches together the results using their preferred methods.



Federated analytics: how OpenSAFELY takes the code to the data

Making this happen was hard work, requiring deep and creative collaboration between software developers and researchers with extensive knowledge of how electronic health records work, and how they are used in research.

But the hard work paid off. For many years, people have talked about the power of GP data in the NHS, but it's never previously been possible to access this data at a national scale. By developing tools that can do federated analytics, and manage privacy challenges, we were able to produce landmark papers analysing the whole population's GP records for the first time in history.

Using these tools, papers have now been published to explore all kinds of important questions: which kinds of patients have, and have not, been vaccinated against COVID-19; which types of patients are being coded as having "Long COVID" most commonly; and whether there were changes in adherence to guidelines on safe prescribing during and after the pandemic. We look forward to working with our users to produce many more.

Standardisation means that all researchers are doing the same job in the same way

Output checking helps to keep private data safe

The whole point of OpenSAFELY is that it enables researchers to make use of private health data, without compromising anyone's privacy.

Researchers don't get direct access to the raw data, but they can still use it to produce research outputs in the form of graphs and tables. These outputs contain aggregated summaries of individual patient data which can provide useful insights without revealing private personal information.

But even summary outputs could accidentally disclose information that might identify individuals or groups of people. Output checking is the work we do to minimise the risk of that happening.

Output checkers are experienced data science professionals. Current and former members of the output checking team include people from the Bennett Institute, the London School of Hygiene and Tropical Medicine and the University of Bristol. Even with their data science backgrounds, all checkers go through formal training and have to pass an exam to complete it. That training is provided by colleagues from the Data Research, Access and Governance Network at the University of the West of England.

An output checker's job is to examine research outputs from OpenSAFELY, and judge whether or not they pose any risk of disclosing identifiable private information. Just like the researchers, the output checkers don't have direct access to the raw data – but they do have the experience and expertise to spot the kinds of outputs that could be problematic.

It's important that output checking happens, and that it isn't done in a rush. No-one wants to delay scientific progress, either. We aim to strike a sensible balance, and the majority of requests submitted by researchers have been checked in fewer than seven days. Most checks are done in under 30 minutes.

The process for output checking looks like this:

1. Having run a job within OpenSAFELY, researchers use Airlock to view the initial outputs within the secure environment. They are expected to carry out disclosure checking on these themselves, before passing anything up to the output checking service. (All OpenSAFELY users have to complete training on this before they're allowed to start using the platform.) Then they'll fill in a request form, in which they explain what each output file shows, and any disclosure controls they've already applied.
2. The request is tracked as an issue on GitHub, and two output checkers are assigned to take a look.
3. Each output checker gets access to the same output results that the researcher saw, and marks each file with a grade: approved, approved subject to change, or rejected.
4. The reviews are sent back to the researcher who proceeds accordingly: if any files were marked as "approved subject to change", the output checker will explain what change is necessary, and the researcher will have to re-submit for another output check after making the changes.
5. Once the outputs are approved by both output checkers, they are released from the secure environment to the researchers, who can continue with any further analysis.

12,700 outputs checked
1,100 requests made
45 output checkers trained

The legal basis: ethics, controls and building trust

Information governance (IG) is the term given to a set of rules about how researchers access patient data. They exist to help us maintain the highest standards of patient privacy, whilst still adhering to the necessary legal frameworks and best-practice ethical principles. IG is a vital component of OpenSAFELY – without those rules, and a system for maintaining and checking that they're adhered to, OpenSAFELY simply couldn't function. It would no longer be considered 'safe.'

The governance of OpenSAFELY is a complex and, above all, collaborative process. NHS England is the Data Controller for the service as a whole. The GP practices themselves remain the Data Controller for the raw GP data that the OpenSAFELY tools operate on.

Day-to-day, our IG team supports researchers from one end of the process to the other – from applying to use OpenSAFELY, to publishing a paper. We help to make sure that researchers are properly trained; have the correct permissions to access data; and are given access to the relevant policies. We also check that every project using the rules for COVID-19 data access meets the relevant criteria.

We work across the whole platform to ensure that all relevant permissions are in place. This entails close work with NHS England and other external bodies such as the Health Research Authority (HRA), ONS and the Department of Health and Social Care (DHSC).

We help to identify the legal basis (under UK GDPR and Common Law) for processing patient data, supporting NHS England to complete all the necessary documentation, including the Data Protection Impact Assessment (DPIA); Data Processing Agreements (DPAs) with EMIS and TPP; Data Sharing Agreements (DSAs) with data providers; and the Data Provision Notice to GP practices explaining the legal obligation they are under to share patient data to OpenSAFELY. Yes, there are a lot of important forms.

We work with colleagues across the Bennett Institute – for example, with OpenSAFELY co-pilots – to explain the controls and checks applied to OpenSAFELY applications. And with developers, to

check that everyone's clear about agreements with external data providers regarding access to specific datasets, and help develop new OpenSAFELY features for auditing and monitoring data access and processing activities.

A lot of our time is spent talking to GPs, patients and the public, policymakers and other groups, to learn about their concerns, and to collaboratively develop solutions that will manage their concerns around data access and maintain support for OpenSAFELY across the wider community.

Lastly, we work closely with the BMA, the RCGP, NHS England and privacy campaigners (such as medConfidential) to provide OpenSAFELY with the legal basis and wider stakeholder support to expand analyses beyond COVID-19, that will bring benefits to patients, clinicians and the wider NHS across all of human health.

The clue is in the name when we're talking about "information governance". We're here to help govern access to data, because the rules exist for good reasons. Someone has to check that the rules make sense, that they are workable, and that everyone's sticking to them. That's our job.

From the beginning we have sought advice and input from technical experts, privacy specialists, and those well-versed in public opinion on the use of health data

Earning and maintaining trust: PPIE and more

It's essential that OpenSAFELY is trusted by everyone: that includes patients, funders, policymakers, professional bodies, privacy campaigners, and other stakeholders.

PPIE is a technical term used by medical researchers: it stands for "Patient and Public Involvement and Engagement". Meaningful and recurring public engagement really matters, which is why we set up our Digital Critical Friends (DCF) Group.

This group is independently chaired by Andy Gibson (Professor of Patient and Public Involvement and Engagement at the University of the West of England) and John Kellas (a community technology and data consultant), and aims to represent the population of England by age, gender, ethnicity and geography.

The DCF Group helps us run OpenSAFELY, meeting every month to offer invaluable inputs in project discussions, raising pertinent questions and providing informed critiques. It has covered operational and strategic topics such as: the life of OpenSAFELY beyond COVID-19-focused research, co-designing communications, NHS consultation processes, and more.

Involvement and engagement with patients and the public has provided considerable insights into the day-to-day running and development of OpenSAFELY. It's also helped us spot the enablers and barriers to conducting high-quality involvement and engagement activities for data-intensive research when the focus is on platform design, not specific research questions.

We hope that this work has been helpful, and the evidence suggests that it has.

In 2021, the NHS, NIHR and the National Data Guardian commissioned a series of Citizens' Juries to review various data-sharing initiatives introduced during the COVID-19 pandemic. Of the initiatives reviewed, the Juries were most supportive of OpenSAFELY (77% of jurors very much in support). Most jurors considered OpenSAFELY to be the most

transparent, trustworthy and secure of the sharing initiatives. When asked at the end of the jury process whether OpenSAFELY should continue after COVID-19, 87% of jurors said yes.

OpenSAFELY also needs trust and support from a wider community of stakeholders.

From the beginning we have sought advice and input from technical experts, privacy specialists, and those well-versed in public opinion on the use of health data.

OpenSAFELY has received ongoing support from the professional community, including from the BMA, the RCGP and the Joint GP IT Committee, as well as from NHS England's Advisory Group for Data (AGD). This is not because of lobbying: it's because we have involved them from the outset, sharing our plans, listening to their concerns, and then – critically – building practical and technical solutions to manage any concerns they raised.

Similarly, privacy campaigners medConfidential – who are often loudly critical of projects accessing citizens' data – have been strongly, publicly and actively supportive of OpenSAFELY. Again, this is not because of lobbying: they give detailed technical feedback, and we have modified the way our platform works in response.

Lastly, we maintain active partnerships with all of our wider stakeholders including Data Controllers, funders, partner organisations, users, patient representatives, privacy campaigners, the BMA and RCGP, not least in our formal governance. All these constituencies are represented in OpenSAFELY governance through the OpenSAFELY Oversight Board.

Above all, though, our trust is earned through actions, rather than through meetings. Our goal is to be provably trustworthy. We earn trust through one simple thing: we set out to deliver new technical methods, tools and services that materially protect patients' privacy.

Our goal is to be provably trustworthy

2

How OpenSAFELY began

OpenSAFELY was born out of the COVID-19 pandemic. Within weeks of the first official confirmation of cases of "viral pneumonia" by Chinese authorities in late 2019, governments and health teams all over the world began to make preparations for pandemic response.

At the Bennett Institute – then still called "the DataLab" – we did the same. We immediately started asking ourselves: what would colleagues in healthcare need to tackle a full-blown pandemic? How could we make use of the National Health Service's existing datasets? What could we do, on a practical level, to contribute?

Two days after the official pandemic announcement, we published an article on the British Medical Journal website, in which we said:

> There are numerous sources of data that can be better exploited from primary and secondary care, each with their own attendant barriers [...]. COVID-19 shows more clearly than ever that we can and must deliver clean, real-time, standardised data to support direct care and all aspects of system planning and response. This is not a 'back-office expense' to be minimised, but a core part of delivery.

The team immediately started brainstorming ideas. The most interesting was for a "data platform to give health researchers what they will need [...] fast, secure access to large volumes of COVID-19 related data". It was initially known as "The Open COVID Research Platform".

We proposed a platform of some kind: something open for research, but safe for patient privacy

Jessica Morley, who was our Policy Lead at the time, later wrote a detailed history of those early days, and reflected:

> We realised that there was an increasingly urgent need to answer questions such as: which demographic characteristics or medical conditions made people more vulnerable to COVID-19? Which drugs might help or hinder the treatment? And what happens to people after they have recovered from initial infection? We also realised that to answer these questions quickly, researchers would need rapid access to unprecedented volumes of clinical data, and a means of conducting high-quality analytics in a collaborative fashion. We concluded that rather than a single study or data source, the NHS needed a platform that would enable many data analysis studies to be conducted in a single secure environment.

Within days we wrote a joint letter, with colleagues from the London School of Hygiene and Tropical Medicine, to the Secretary of State for Health. In it we proposed a platform of some kind: something open for research, but safe for patient privacy. Soon after that, we asked our contacts at health records company TPP if they might be interested in collaborating on something. They immediately said yes.

On the same day, the government had issued a "Control of Patient Information" (COPI) notice to the Chief Executive of the NHS, which gave legal backing in principle to access electronic health records on behalf of NHS England – a prerequisite for making the platform work. But a legal basis is only part of the story: there would still need to be a trustworthy way to achieve that access. Things were moving extremely fast.

One week after the first UK lockdown was announced, we committed the first line of code to GitHub – work on the platform had begun. But it still didn't have a finalised name. That took a few more weeks of debate and friendly argument to decide. Before the end of April 2020, we'd settled on "OpenSAFELY."

On 7 May 2020, the first scientific paper written using OpenSAFELY went to pre-print: "OpenSAFELY: factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients."

This was a big moment. We were showing that OpenSAFELY worked as a fully open-source, privacy-preserving software platform, capable of running open and reproducible analytics across electronic health records, all held securely in situ. Working with colleagues from the Electronic Health Records research group at the London School of Hygiene and Tropical Medicine, NHS England, and TPP, we'd got it up and running in just 42 days.

As 2020 rolled on, that paper was formally published in Nature, and we expanded OpenSAFELY's reach to include data from the other major electronic health records company, EMIS, as well as TPP. Now the platform provided researchers with access to more than 55 million patient records – more than 95% of all patient data in England.

Towards the end of that year – partly in response to requests from privacy campaigners – we created the Jobs Dashboard at jobs.opensafely.org, so that anyone could see the work running on the platform.

In December, the first COVID-19 vaccine was approved for use in the UK. The NHS vaccination programme began just six days later. OpenSAFELY delivered its first live dashboards showing which kinds of patients were and were not receiving the vaccine within ten days of the first vaccine being administered.

The next big milestone came in 2021, when we published the first federated analysis using data from both TPP and EMIS: "Trends and clinical characteristics of COVID-19 vaccine recipients: a federated analysis of 57.9 million patients' primary care records in situ using OpenSAFELY."

Jessica Morley noted:

> This federated analysis was a truly massive technical achievement […] it was driven, as ever, by the combination of skills that no single individual, or even team, is ever likely to embody alone, across EHR data analysis, EHR system design, software development, data management, open science, and more.

Since then, it's been that collaborative team thinking that has pushed OpenSAFELY forward as a continuously evolving and improving digital tool for researchers. Our goal has always been to be a team that goes beyond research – we wanted to use data to build machines that act in the world, machines that make a practical, tangible difference. Tools that you can use to get things done. OpenSAFELY has got a great deal of practical, tangible work done: 86 peer-reviewed and published papers so far, and counting.

OpenSAFELY began with the pandemic, and has been largely funded to date on the understanding that research conducted through it relates to COVID-19 – but we think it has a bright future ahead. We'd like to expand its reach, to get more researchers using it in more organisations. We're already looking at ways we could use the OpenSAFELY model in other aspects of health care, beyond electronic health records. Or beyond health care entirely.

The journey so far has not been easy. There have been downs, as well as ups. But the team is committed, hard-working and determined. We have a lot more work to do yet.

We wanted to use data to build machines that act in the world

3

Stories from OpenSAFELY users

Consequences of COVID-19 and the role of vaccination

by Dr Venexia Walker, Senior Research Fellow,
Bristol Medical School, University of Bristol

“No other platform comes close”

by Dr Mark Russell,
King’s College London

The ‘unreal’ speed of OpenSAFELY

by Dr Ed Parker, Assistant Professor,
London School of Hygiene & Tropical Medicine

Using OpenSAFELY to fight antimicrobial resistance

by Professor Diane Ashiru-Oredope,
Lead Pharmacist for healthcare-associated infections and antimicrobial resistance,
UK Health Security Agency

OpenSAFELY and antibiotics

by Francine Jury, Project Manager,
School of Health Sciences, University of Manchester

Using OpenSAFELY to carry out a randomised trial

by Ben Ainsworth, Associate Professor,
Digital Interventions Group, University of Southampton

Consequences of COVID-19 and the role of vaccination

by Dr Venexia Walker, Senior Research Fellow,
Bristol Medical School, University of Bristol

On 11 February 2020, the disease associated with SARS-CoV-2 was named by the World Health Organisation as COVID-19. Over time, SARS-CoV-2 strains have mutated, resulting in new variants. Different strains differ in their infectiousness, symptom profiles, and potentially their consequences, so each warrants detailed study.

The rapid rollout of COVID-19 vaccination was a crucial component of the public health response to the pandemic. Rollout began on 8 December 2020, initially aimed at workers in care homes and people over 80 years old. From there, it gradually expanded to cover more people, new vaccines were authorised, and there were several rounds of booster shots.

Right from the start of the pandemic, we knew it was important to identify the consequences of COVID-19 for physical and mental health, and, once vaccines were available, to better understand vaccine coverage (i.e., who received vaccines) and effectiveness (i.e., how well were people protected from the virus). We knew OpenSAFELY was a good place to answer these questions, as the OpenSAFELY-TPP database allowed us to analyse the full primary care records of 45% of the English population with relevant linkages, while the platform enabled cross-institution team science – the likes of which we could not achieve pre-pandemic.

Ours was a large cross-institution team, mainly at the Universities of Bristol and Cambridge. Previously, it would have been very difficult for all of us to work on the same electronic health record dataset, because permissions are usually institution-based. You typically have to arrange honorary contracts, and sometimes even travel to another institution that you want to collaborate with. OpenSAFELY changes all that.

We wanted to find answers to questions such as:

- Are people getting the benefits of vaccination observed in trials?
- How long do the benefits of vaccination last?
- Do the consequences of COVID-19 differ if you are vaccinated when you get it?

Over the years that followed, we published several papers on these and other topics. Some of our findings included:

- The effectiveness of both the Pfizer and AstraZeneca vaccines against infection waned to almost nothing after six months – although they were still very effective against the severe outcomes of hospitalisation or death.
- COVID-19 booster vaccination, compared with no booster vaccination, provided substantial protection against COVID-19 hospitalisation and death, but only limited protection against testing positive for SARS-CoV-2.

Using OpenSAFELY influenced and encouraged good working practices within our team. We made our protocols (documents setting out the plans for research before it happens) more detailed, we implemented code review, and some of our team became qualified output checkers – which provided a greater understanding of what 'good outputs' look like.

OpenSAFELY's built-in reproducibility was a very important factor: code from our early projects became the starting point for later projects, and some of our frequently used code was developed into a reusable action. (In OpenSAFELY, an action is a piece of code written for one study, that can be reused in other studies without copying-and-pasting between them). The ability to reuse code in this way sped up the whole process, which was extremely useful when the pressure of the pandemic on the health service – and on health researchers – was at its height.

OpenSAFELY's built-in reproducibility was a very important factor

“No other platform comes close”

by Dr Mark Russell, King’s College London

Our team used OpenSAFELY to research how the COVID-19 pandemic affected routine healthcare for people with inflammatory arthritis conditions, such as rheumatoid arthritis.

The standard way of studying – and therefore improving – routine healthcare in this and other areas of medicine is via national audits, in which clinicians are asked to fill out forms manually. That’s a time-consuming task, often seen by clinicians as an unwelcome burden. We wondered: could we use OpenSAFELY to change how some of that monitoring is done?

We used OpenSAFELY to replicate some of the quality-of-care metrics that are typically generated by the National Early Inflammatory Arthritis Audit. We looked at data points such as:

- number of new diagnoses
- time to assessment by a hospital specialist
- time to prescription of a disease-modifying treatment.

We then compared the standards before and after the pandemic, nationally and regionally.

What we found was really interesting. New diagnoses of inflammatory arthritis dropped in the first year of the pandemic, then went back up, but they still haven’t rebounded above pre-pandemic levels. Which suggests that there might be lots of undiagnosed patients out there.

We also found that diagnosed patients got good care, despite the pandemic. The time to see a hospital specialist continued to improve, and the proportion of patients who were prescribed disease-modifying drugs remained stable. The health service did an impressive job of adapting to the pressures and restrictions of working through the pandemic and all the lockdowns.

What was most striking to us was that we were able to gain all these insights using data already present in OpenSAFELY – no additional, time-consuming data collection had to happen. We think this research demonstrated the potential for using OpenSAFELY and other platforms alongside audits, and perhaps in the long term, replacing the need for manual data collection entirely.

Equally striking was how using OpenSAFELY introduced us to new ways of doing research. While many of us had experience of writing code, it was the first time I'd used GitHub, or Python. So our whole team was very grateful for the resources made available to help us learn – the documentation for OpenSAFELY is better than anything I've seen on any other platform. OpenSAFELY's docs are detailed, well written and very helpful. What's more, there's a huge archive of other people's code that you can reuse when writing your own code. The dummy data OpenSAFELY provides makes the whole process safer – not just from a privacy perspective, but also because it helps improve the integrity of the research.

We also really liked working with our co-pilot. At least once a week, we had a chance to ask questions, get our code reviewed, and get help working our way through the process. Once our co-pilot had walked us through some of the trickier tasks, it was easier to do them solo next time round.

We enjoyed working with OpenSAFELY. It's actually not as hard to pick up as you might think. There's a proactive team behind it, and a supportive community around it. That makes writing code easier, especially when you're actively encouraged to reuse code that's already been written. No other platform comes close in terms of data coverage, privacy and safety.

Using OpenSAFELY introduced us to new ways of doing research

The ‘unreal’ speed of OpenSAFELY

by Dr Ed Parker, Assistant Professor,
London School of Hygiene & Tropical Medicine

It was clear from the early days of the COVID-19 pandemic that people with kidney disease were particularly vulnerable to the effects of the virus.

That’s why these people were among the first to be offered vaccines when they became available. They’ve since been offered extra booster doses for further protection.

We were interested in finding out the effects of different vaccines (specifically those developed by AstraZeneca and Pfizer) during the early stages of the vaccine rollout. We found that after two doses, the AstraZeneca vaccine was slightly less effective than the Pfizer vaccine at preventing infection, hospitalisation, and death among people with kidney disease. But when the same group was given a third dose of the Pfizer vaccine, regardless of which vaccine they’d been given initially, the playing field was levelled. The third dose seemed to close the gap in protection. All of this played-out between the spring and winter of 2021, when the Delta wave of the virus was dominant, and just as the Omicron variant took over towards the end of the year. Our work helped highlight the value of additional doses of RNA vaccines in those at highest risk of COVID-19.

What was important here was access to the right sort of data, at the right sort of scale. That’s where OpenSAFELY was really powerful – it gave us access not just to 24 million up-to-date patient records from GPs via OpenSAFELY-TPP, but also to data in the UK Renal Registry, which keeps a record of every patient on dialysis, or who has received a kidney transplant.

Stitching these two datasets together was the best way to accurately identify people with severe kidney disease before looking at the effectiveness of different vaccines. OpenSAFELY was the only platform that gave us all the necessary data in one place.

I was mentored throughout this work by Laurie Tomlinson at the London School of Hygiene & Tropical Medicine, and got a lot of help and input from other collaborators, including the Electronic Health Records Research Group here at LSHTM, the UK Renal Registry, the University of Bristol, and the team at the Bennett Institute.

I'm a data scientist so I felt comfortable writing code, but things were made easier because there was already so much existing code for key steps in the analysis. I could take code that someone had written to ask one question, and repurpose it to ask another. When I got stuck with anything, the Bennett team was on hand with support – a quick 10-minute call and job done, we could move on.

I couldn't believe how fast it was – I started in January 2022 and had a paper online by June. That's unreal in my experience.

OpenSAFELY might feel daunting at first, but I'd say it's as accessible as you could hope for considering the complexity of the data involved. As a vaccine researcher, I have felt very privileged to use OpenSAFELY to address public health issues related to the uptake and effectiveness of COVID-19 vaccines in different groups.

What was important here was access to the right sort of data, at the right sort of scale

Using OpenSAFELY to fight antimicrobial resistance

by Professor Diane Ashiru-Oredope, Lead Pharmacist for healthcare-associated infections and antimicrobial resistance, UK Health Security Agency

The more we use antibiotics, antifungals, antivirals and antiparasitics, the less effective they can become. It's a phenomenon known as "antimicrobial resistance," or AMR. One way to mitigate it is by avoiding inappropriate repeat antibiotic prescriptions in primary care.

Working collaboratively using OpenSAFELY, our team at the UK Health Security Agency (UKHSA) worked with the Bennett Institute on a study to evaluate repeat antibiotic prescribing from general practices during the pandemic. Overall, we found that antibiotic prescribing decreased during the pandemic, for many reasons including restrictions/lockdowns, changes in healthcare delivery, and healthcare-seeking behaviour. This decline was particularly pronounced for one-off prescriptions, compared with repeat prescriptions.

Because OpenSAFELY has lots of data from GPs, we were also able to investigate in granular detail the clinical and demographic characteristics of small groups of patients, and shine new light on health inequalities and AMR. We found that older patients and care home residents were more likely to receive antibiotics, especially repeat antibiotic prescriptions. With OpenSAFELY, we can match the prescription with the diagnosis and here, we found repeat prescribing was highest for people with chronic obstructive pulmonary disease (COPD) and urinary tract infections.

OpenSAFELY's built-in reusability was an important part of this project. NHS England has a published approach to help reduce healthcare inequalities at national and systemic levels, known as Core20PLUS5. We were able to reuse code for this that was already available in OpenSAFELY, which helped us target specific health conditions, more deprived population groups, and ethnic minorities.

The result was that we were able to conduct one of the most granular antibiotics-prescribing studies ever undertaken. It showed a general trend of decreasing one-off and repeat prescriptions, but also huge variation in the changes across ethnic minorities:

- Compared to before the pandemic, there was a 40% decrease in one-off antibiotic prescriptions among patients of Chinese ethnicity.
- On the other hand, for people with mixed ethnic backgrounds, such as White and Black African, White and Black Caribbean, and other mixed backgrounds, we actually noticed an increase in repeat prescriptions during the pandemic.
- There were also distinct differences in patterns of prescribing between Indian and Pakistani communities, two groups which are normally aggregated together in smaller studies, hiding these differences.

The findings from this study played an important part in developing new AMR tools for general practice teams, such as a set of “How to” guides to help manage and review adults on long-term or repeated antibiotics for Acne Vulgaris and COPD exacerbations. The guides were published online at the Royal College of GPs TARGET eLearning hub.

As a result of this project, we have a better understanding of how OpenSAFELY can help us address health inequalities related to infection and AMR. Building on this work, we also collaborated with the University of Manchester to assess the association between non-COVID-related sepsis incidence, mortality, and health inequality factors.

OpenSAFELY has been a productive tool for us so far, and we’re planning more work to investigate the risk factors for resistant bloodstream infections, to support the development of more interventions to slow AMR.

OpenSAFELY has been a productive tool for us

OpenSAFELY and antibiotics

by Francine Jury, Project Manager,
School of Health Sciences, University of Manchester

The OpenSAFELY team at the University of Manchester (UoM) consisted mostly of early career researchers. Using OpenSAFELY gave us a new opportunity to analyse millions of anonymised patient records, but initially there was a steep learning curve. However, the support of the active OpenSAFELY user community, and ability to share programs and codes quickly, enabled the team to learn the necessary skills, and conduct a substantial and impactful amount of research.

Initially we used OpenSAFELY to analyse the impact of COVID-19 on the management of common infection and antibiotic prescribing in primary care. Our research highlighted the substantial challenges around antibiotic prescribing. Key findings included:

- A lack of risk-based prescribing of antibiotics (i.e., patients at low risk of infection-related complications were as likely to be prescribed an antibiotic as high-risk patients).
- A substantial number of patients receiving frequent antibiotics over time. Our research further highlighted increased concerns about the safety of this (patients with a history of frequent antibiotic prescribing with multiple types were at highest risk of developing severe COVID-19).
- Community-acquired sepsis is a significant problem, particularly in patient groups that suffer from health inequalities.

There's a need to strengthen clinical practice with better advice to clinicians, in order to improve targeting and risk-based prescribing of antibiotics.

The UoM eHealth research team used the OpenSAFELY analysis outputs to develop a number of tools with integrated algorithms. These include a

series of dashboards that we built on top of local GP software systems in Manchester, that can identify:

- patients frequently prescribed antibiotics
- inappropriately prescribed antibiotics
- practice prescribing by patient risk profiles.

These dashboards are now being used across the north west of England, giving clinical teams an opportunity to audit prescribing for common infections, and identify areas for improvement interventions.

In addition, we've built the BRIT2 knowledge support system for clinicians, now being trialled across the north west. It integrates risk scores for infection-related complications that can support clinical decision-making in consultations with patients. The risk scores developed from OpenSAFELY analysis include risks of hospitalisation, antibiotic failure, increased resistance and risk of adverse events.

Being able to analyse millions of patient records through OpenSAFELY allowed the UoM Research team to present to clinicians risks based on hundreds of patients with the same risk profiles as their patient, making visualisation and presentation of information easy to understand and act upon. The UoM eHealth research team will continue to use OpenSAFELY to investigate future applications of these tools that can impact other significant health challenges.

Using OpenSAFELY gave us a new opportunity to analyse millions of anonymised patient records

Using OpenSAFELY to carry out a randomised trial

by Ben Ainsworth, Associate Professor,
Digital Interventions Group, University of Southampton

During the COVID-19 pandemic, it became clear that supporting people to change their behaviours was going to be crucial in reducing the spread of the virus. Simple things like wearing masks, opening windows and washing hands would help reduce the impact of COVID, buying the NHS valuable time while a vaccine was developed and tested.

During the earlier H1N1 pandemic in 2009, we had developed a web app, 'Germ Defence,' which encouraged people to improve hygiene practices at home. Between 2010 and 2013 this was tested in a large randomised controlled trial of over 20,000 people, and the test showed that people who used Germ Defence reduced their chances of developing a cold or respiratory infection by 14%.

So when the COVID-19 pandemic struck, we knew that some relatively minor updates to Germ Defence would be likely to make it an effective tool against the COVID virus. These were updates around the technology, but also to the behavioural advice.

After making those changes, we worked to disseminate Germ Defence via any mechanism we could. However, as with any medical intervention, we also recognised the need to evaluate it. We wanted to see whether GP practices that asked their patients to use Germ Defence reduced the amount of patients who were diagnosed with COVID-19 (and other infections), as well as other healthcare outcomes.

However, there are challenges to evaluating interventions that are already being used. The evaluation needs to happen quickly (so that it can inform ongoing intervention) and it needs to be unobtrusive – so that it doesn't interfere with the healthcare that is already being delivered.

We were able to do this using OpenSAFELY. We designed the largest clinical randomised control trial on the topic to have ever been conducted (that we know of!). In Autumn 2020, we randomised every single GP practice in England to text their patients a link to Germ Defence (3,297 practices), or continue usual practice until Spring 2021 (3,282 practices). This gave a maximum sample size of every NHS patient in England!

We designed a trial that avoided many of the usual burdens that are placed upon already overstretched GP practices taking part in research. OpenSAFELY meant we could analyse data for 1,246 intervention practices and 1,252 control practices, representing 11.9 million and 12.3 million registered patients.

In the trial, only 460 practices sent Germ Defence on to their patients (16%), which was far below what we had hoped, and we didn't see any reduction in infection rates. However, we were able to link patient website usage with infection rates at a practice level. OpenSAFELY allowed us to evaluate a large-scale public health intervention in real time, while it was actually being used.

We were able to deliver gold-standard randomised control trial results in a short enough time frame to inform current practice – in a fraction of the expected time, at a fraction of the expected cost, and with far fewer people working on it than expected. In the future, this method can be used for evaluating a huge range of new interventions – technological, social, behavioural, pharmaceutical and organisational.

OpenSAFELY allowed us to evaluate a large-scale public health intervention in real time

4

The OpenSAFELY Collaborative

OpenSAFELY is a huge collaborative project that spans a diverse range of organisations including universities, the NHS, and GP system suppliers.

The core technical platform teams at Oxford now have around 50 software developers, researchers, clinicians, and data policy experts. This core team designs and runs the network of components that make up the OpenSAFELY service: the technical tools (some made by our teams, some re-purposed) and the human services like output checking and co-piloting.

These teams have been built over the past decade specifically to deliver digital tools and services on top of health data. We set out to pool skills, knowledge, and best practice from our constituent communities. Our software engineers arrive with great developer skills, but they acquire deep knowledge on epidemiology research, electronic health records and more. Conversely our traditional epidemiology researchers have acquired deep technical skills on software development, not to replace developers, but to work closely alongside them, in one technically creative community.

These OpenSAFELY tools can be built in any data centre containing electronic health records data. For the NHS England OpenSAFELY GP Data Service, the tools are integrated onto the databases and technical services inside the machines owned by the major **GP system suppliers** – the companies that make the electronic health records systems that GP practices use to store their patient notes. We don't just integrate into these companies' technical services: we also pool skills, knowledge and ideas with their teams, because their developers literally designed the underlying primary care datasets that our research community want to analyse.

NHS England are the Data Controllers of the service as a whole, and the core funder of the NHS England OpenSAFELY GP Data Service. They resource us to support 50 new projects a year (we'd always love to support more, so do get in touch if you have any funding opportunities!). But they also work closely with us on getting new users and projects into the platform, managing all the information governance, and generating and delivering ideas for new services or analyses. We are working with NHS England to help them take on more of the service aspects in the platform.

GP practices remain the Data Controllers for all the raw records that our tools run against, inside the GP system suppliers' machines. They are also a core part of our community: we work closely with the RCGP, the BMA and the Joint GP IT Committee to make sure that our tools and services meet their needs, and they contribute energetically to our oversight and governance.

Patients and the public are at the heart of what we do: they live the lives that are encoded in this data, and any findings that our researchers make are only possible because of that data, so we are closely focused on earning and retaining trust in all our design choices. Our Digital Critical Friends group is one way that we make sure we stay close to the needs of patients and the public, alongside Citizens Juries and other policy and public engagement initiatives.

Then there are **the users of the platform**: epidemiologists, health service analysts, and people from think-tanks and other organisations. Some of these teams – like LSHTM, Bristol and Manchester – are very close collaborators. They do their own research analyses, of course; but in addition, many of their researchers also give extra time, helping us to develop and test new ways of working, writing their own tools, and taking on tasks like output checking.

Every user contributes reusable shared code to the community

Code for one is code for all

Alongside all this, the design of OpenSAFELY means that every user contributes re-usable shared code to the community: every subsequent user can see it, evaluate it, and re-use it, efficiently and productively.

This works in two ways. Firstly, the design of OpenSAFELY means that all the code that's been run is automatically shared openly. But that wouldn't be enough, on its own, to build a growing ecosystem of technical collaboration across so many different groups. Preparing electronic health records data is a complex business: in the past everybody did similar tasks – like data preparation – in their own esoteric way, on their own machine. So even if they did share their code, it would be hard for any other group to disentangle what they have done.

In OpenSAFELY we have imposed light-touch pragmatic standardisation of common tasks like data preparation, using tools like ehrQL. This brings many efficiency benefits, but also one critical community benefit: everybody is doing the same task in the same way, which means they can more easily read, understand, evaluate and re-use each other's code. This is called "legibility" in the coding community, and you can see the benefits in the huge productivity of our users. It means that every use of OpenSAFELY is contributing to the growing stack of knowledge, code and tools needed to prepare and analyse GP data. For that, we are hugely grateful to all our wonderful, energetic users.

How to work with us

OpenSAFELY was designed from the outset as a huge collaborative project, with structures that capture and share all the work of everyone using the platform. There are lots of ways that teams and individuals can contribute and benefit our large community: some are obvious, and some less so.

Use the NHS England OpenSAFELY GP Data Service to do your own research

There is currently no charge for access to the platform, and the application process is a short form, followed by a short wait. NHS England pays for 50 projects per year, and they prioritise the projects, with input from our teams on technical issues.

Implement our tools in your own data centre

The OpenSAFELY platform is constituted from a detailed network of components, tools and services that each deliver key tasks like data preparation, remote execution of code, federated analytics and so on. Our philosophy is that the data stays put, and the code comes to the data. Our tools were built to be portable, and run in any data centre containing electronic health records data. If you would like to use our tools and methods, we are always happy to talk. You might want your data centre to be fully "OpenSAFELY-enabled"; or just use single components, like our standardised tools for preparing raw GP data into analysis-ready tables.

Our philosophy is that the data stays put, and the code comes to the data

Build your own tools into our platform

OpenSAFELY is best thought of as a network of components. Some are technical (a particular bit of code that does a particular job); and some are services, run by teams of people following templates (like our output checking service, or our co-pilot service). We spend our time thinking carefully about how to build networks of these components: identifying existing components to re-use; and, where necessary, building new components in this carefully designed network. We're always keen to have more people contributing. We can't open all the core code up for anyone to make occasional changes, because the network is complex, and the platform is busy, with critical private data inside. So we've specifically created new ways for people to contribute components for their own tasks:

- Contribute to the actions pipeline. Our actions framework is a well documented set of services that allow users to write their own components and stitch them into an actions pipeline. Our documentation has detailed guides for contributors, and lots of examples of actions that users have produced already, to help their pipelines work the way they want.
- Talk to us about deep changes and feature requests. Sometimes, users want new features or changes that require deeper integration into the network of components. We're always happy to talk with teams about how best to do this, either in collaboration with our own teams, or working more on their own if they prefer.

Fund us, so we can do more!

Lastly, you can fund us. We are always keen to help more users, whether that's more projects in our existing services, or new features and different types of data. We are here to help, on team@opensafely.org

5

Some reflections about funding

Running data infrastructure can be hard, and funding it is hard too. We have been able to slowly grow the service, with some great news in just the past few months: today, the OpenSAFELY team has stable funding. But there have been several times when we nearly had to close the platform, and disband the team. People are sometimes shy about discussing funding in public, but we are happy to share our funding journey, in the hope that it might help others trying to build tools and services.

We started at the beginning of the pandemic with no resource. The Oxford team had a slice of core NIHR BRC money – £185,000 per year, to cover the OpenPrescribing service. We instantly switched this to start OpenSAFELY. Soon after, we were able to get core money from UKRI MRC covering some COVID research costs and platform development costs.

For the next period we were able to grow "in the gaps", by finding money for platform work inside grants that were principally focused on delivering COVID-related research papers. Many people included costs for development of shared data infrastructure (with special thanks to Professor Nish Chaturvedi, and her NIHR, MRC and National Core Studies grants). But throughout this period we struggled to get funding that was specifically focused on data infrastructure, as a shared resource for all users.

There was one notable success: Wellcome's data team actively sought us out, and gave us our first and biggest ever data infrastructure grant, at £2.3m. This "discretionary award" was solely and explicitly focused

on stabilising the core team, and developing and delivering shared code, methods, tools, and services to make data more accessible for all users. This grant from Wellcome changed the future of the platform. Later, we also got £1.8m of UKRI/MRC money to expand our user-base through the COVID National Studies 'Data and Connectivity Programme', for which we are hugely thankful.

This platform funding ran down quickly, as we took on huge numbers of users. As COVID funding dried up, we were no longer able to find gaps in other teams' research grants to grow in, or infrastructure funding. Things got very close. But then, in October 2023, NHS England committed £4m per year to pay for 50 projects per year for our users, and some platform changes. For this, we are deeply grateful: not just for the ability to support more users, but also the opportunity to have deeper links with the NHS, its teams, and its expertise.

All academics are familiar with the challenges of funding. Infrastructure seems harder, because it's a square peg. Despite the roadblocks and anxiety (and the rejected applications) we can empathise with the challenges that funders face: they were mostly designed to fund traditional research paper outputs; data infrastructure has shifted rapidly from "beige boxes in concrete buildings" to "code and teams"; and it's hard, especially in a fast-moving space, to evaluate new options.

We are very conscious that all our funding for tools and services has been through bespoke, unconventional arrangements. If we have one request, it's this: that funders could have open, competitive funding calls, where all can submit their own great creative ideas for services, tools and teams that make data accessible for researchers. We know that there are many great teams out there, across the country, in many other places, who could deliver brilliant data infrastructure work, if they could access the funds to scale their ideas.

However, as of autumn 2024, things are looking good for OpenSAFELY. This booklet is published alongside the first big OpenSAFELY User Community conference, and we are delighted to announce that OpenSAFELY has just been awarded two major grants from Wellcome, amounting to £17m over 7 years, starting in 2025. We're hugely grateful for their support.

OpenSAFELY funders to date are:

NHS England

Wellcome

NIHR

MRC

Core Bennett Institute costs are supported by the Peter Bennett Foundation

6

What's next?

OpenSAFELY has been substantially stabilised over the past year. We're always cautious about publicising big hopeful ideas. We prefer to share simple small prototypes directly with our users, get feedback, and iterate through delivery. But here is some of what we know, or hope, is coming soon.

Non-COVID permissions

In November 2023 NHS England, DHSC and Secretary of State announced that OpenSAFELY will cautiously expand to support analyses on non-COVID topics. The paperwork on this has progressed well, with support from all stakeholders: we hope to make a positive announcement soon.

Return to whole population coverage

OpenSAFELY delivers federated analytics, with the same code running in any OpenSAFELY-enabled data centre, covering all GP practices in TPP and EMIS. We were delighted to enable access to the whole nation's GP data for many publications. However, while TPP have been able to work pro bono (for which we are very grateful), the EMIS half of the service has been switched off (very understandably) while awaiting funding from NHS England or similar to enable us to use their resources. We hope this will be resolved soon.

Wellcome core funding and new features

Our new core Wellcome grant is very exciting: it stabilises the core team, to keep developing new methods to join up new datasets and new data centres, and to implement those ideas in working tools and services. This grant is deliberately not tethered to a shopping list of features or datasets. It lets us follow the amber lights, in the desired direction of travel, across diverse non-health and non-UK datasets, and new functionality in our tools. We are hugely grateful.

Wellcome mental health funding

The second of our two Wellcome grants is much more specific. We are working closely with the NHS Talking Treatments team to link up mental health data with electronic health record data, and to develop a new kind of health data research service. Stay tuned.

OpenSAFELY-Schools with the National Institute of Teaching

In late 2024 we started a new deep collaboration funded by Nuffield Trust and the philanthropy fund at XTX Markets to develop new tools and services on top of detailed schools data. More soon.

Lastly:

Not just OpenSAFELY!

The OpenSAFELY tools were specifically built to address the particular challenges of privacy, transparency, efficiency and reproducibility when users are accessing national-scale GP data. But other datasets present very different challenges, and need very different solutions.

We have a range of ideas around different datasets, different ways of linking data centres, and different models of access, that we are discussing with new partners. These include: new ways to do old-fashioned remote desktop Trusted Research Environments (TREs) more safely and efficiently; new ways of passing data carefully between data centres; and more.

They are very different to the OpenSAFELY model, but they keep our core ethos:

- collaborate and re-use all good existing tools or services
- put creative technical delivery teams at the heart of the work, tightly coupled to research domain expertise
- and work around barriers like privacy with ideas and engineering, rather than advocacy alone.

If you want to work with us, get in touch!
Email us: team@opensafely.org

We prefer to share simple, small prototypes directly with our users, get feedback, and iterate through delivery

7

OpenSAFELY papers so far

Costello RE, Pedersen L, Henderson AD, Tazare J, Sorensen HT, et al. Impacts of the COVID-19 pandemic on deprivation-level differences in cardiovascular hospitalisations: a comparison of England and Denmark using the OpenSAFELY platform and National Registry Data. *BMJ Open* 2024; 14: e088710.

Fonseca M, MacKenna B, Mehrkar A, The OpenSAFELY Collaborative, Walters CE, et al. The Use of Online Consultation Systems or Remote Consulting in England Characterized Through the Primary Care Health Records of 53 Million People in the OpenSAFELY Platform: Retrospective Cohort Study. *JMIR Public Health Surveill* 2024; 10: e46485.

Wiedemann M, Speed V, Cunningham C, Higgins R, Curtis HJ, et al. Impact of COVID-19 on recorded blood pressure screening and hypertension management in England: an analysis of monthly changes in the quality and outcomes framework indicators in OpenSAFELY. *Open Heart* 2024; 11: e002732.

Walker VM, Patalay P, Cuitun Coronado JI, Denholm R, Forbes H, et al. COVID-19 and Mental Illnesses in Vaccinated and Unvaccinated People. *JAMA Psychiatry* 2024; e242339.

Taylor K, Eastwood S, Walker V, Cezard G, Knight R, et al. Incidence of diabetes after SARS-CoV-2 infection in England and the implications of COVID-19 vaccination: a retrospective cohort study of 16 million people. *Lancet Diabetes Endocrinol* 2024; 12: 558–568.

Bechman K, Green A, Russell MD, Yang Z, Zheng B, et al. The safety of antivirals and neutralising monoclonal antibodies used in prehospital treatment of Covid-19. *J Infect* 2024; 106227.

Andrews CD, Mathur R, Massey J, Park R, Curtis HJ, et al. Consistency, completeness and external validity of ethnicity recording in NHS primary care records: a cohort study in 25 million patients' records at source using OpenSAFELY. *BMC Medicine* 2024; 22: 288.

Zhong X, Palin V, Ashcroft DM, Goldacre B, MacKenna B, et al. Risk of emergency hospital admission related to adverse events after antibiotic treatment in adults with a common infection: impact of COVID-19 and derivation and validation of risk prediction models. *BMC Med* 2024; 22: 277.

Schaffer AL, Park RY, Tazare J, Bhaskaran K, MacKenna B, et al. Changes in sick notes associated with COVID-19 from 2020 to 2022: a cohort study in 24 million primary care patients in OpenSAFELY-TPP. *BMJ Open* 2024; 14: e080600.

Schaffer AL, Andrews CD, Brown AD, Croker R, Hulme WJ, et al. Changes in opioid prescribing during the COVID-19 pandemic in England: an interrupted time-series analysis in the OpenSAFELY-TPP cohort. *Lancet Public Health* 2024; 9: e432–e442.

Horne EMF, Hulme WJ, Parker EPK, Keogh RH, Williamson EJ, et al. Effectiveness of mRNA COVID-19 Vaccines as First Booster Doses in England: An Observational Study in OpenSAFELY-TPP. *Epidemiology* 2024; 35: 568–578.

Lin L-Y, Henderson AD, Carlile O, Dillingham I, Butler-Cole BFC, et al. Healthcare utilisation in people with long COVID: an OpenSAFELY cohort study. *BMC Med* 2024; 22: 255.

Yang Y-T, Wong D, Zhong X, Fahmi A, Ashcroft DM, et al. Exploring Prior Antibiotic Exposure Characteristics for COVID-19 Hospital Admission Patients: OpenSAFELY. *Antibiotics* 2024; 13: 566.

Cunningham C, Fisher L, Wood C, Speed V, Brown AD, et al. Incidence and treatment of group A streptococcal infections during covid-19 pandemic and 2022 outbreak: retrospective cohort study in England using OpenSAFELY-TPP. *BMJ Med* 2024; 3: e000791.

Henderson AD, Butler-Cole BFC, Tazare J, Tomlinson LA, Marks M, et al. Clinical coding of long COVID in primary care 2020–2023 in a cohort of 19 million adults: an OpenSAFELY analysis. *eClinicalMedicine* 2024; 72: 102638.

Lemanska A, Andrews C, Fisher L, Bacon S, Mehrkar A, et al. During the COVID-19 pandemic 20 000 prostate cancer diagnoses were missed in England. *BJU Int* 2024; 133: 587.

Carlile O, Briggs A, Henderson AD, Butler-Cole BFC, Tazare J, et al. Impact of long COVID on health-related quality-of-life: an OpenSAFELY population cohort study using patient-reported outcome measures (OpenPROMPT). *Lancet Reg Health Eur* 2024; 40: 100908.

Santhakumaran S, Fisher L, Zheng B, Mahalingasivam V, Plumb L, et al. Identification of patients undergoing chronic kidney replacement therapy in primary and secondary care data: validation study based on OpenSAFELY and UK Renal Registry. *BMJ Med* 2024; 3: e000807.

Brown AD, Fisher L, Curtis HJ, Wiedemann M, Hulme WJ, et al. OpenSAFELY: The impact of COVID-19 on azathioprine, leflunomide and methotrexate monitoring, and factors associated with change in monitoring rate. *Br J Clin Pharmacol* 2024; 10.1111/bcp.16062.

Cezard GI, Denholm RE, Knight R, Wei Y, Teece L, et al. Impact of vaccination on the association of COVID-19 with cardiovascular diseases: An OpenSAFELY cohort study. *Nat Commun* 2024; 15: 2173.

Samuel M, Park RY, Eastwood SV, Eto F, Morton CE, et al. Weight trends amongst adults with diabetes or hypertension during the COVID-19 pandemic: an observational study using OpenSAFELY. *Br J Gen Pract* 2024; BJGP.2023.0492.

Homan K, Seeley R, Fisher L, Roberts CA, Khatri S, et al. Safety of direct-acting oral anticoagulant (DOAC) prescribing: OpenSAFELY-TPP analysis of 20.5 million adults' electronic health records. BJGP Open 2024; *BJGPO.*2023.0163.

Green MA, McKee M, Massey J, Mackenna B, Mehrkar A, et al. Trends in inequalities in avoidable hospitalisations across the COVID-19 pandemic: a cohort study of 23.5 million people in England. *BMJ Open* 2024; 14: e077948.

The OpenSAFELY Collaborative, Wood C, Speed V, Fisher L, Curtis HJ, et al. The impact of COVID-19 on medication reviews in English primary care. An OpenSAFELY-TPP analysis of 20 million adult electronic health records. *Br J Clin Pharmacol* 2024; 90: 1600–1614.

Nab L, Schaffer AL, Hulme W, DeVito NJ, Dillingham I, et al. OpenSAFELY: A platform for analysing electronic health records designed for reproducible research. *Pharmacoepidemiol Drug Saf* 2024; 33: e5815.

Ainsworth B, Horwood J, Walter SR, Miller S, Chalder M, et al. Implementing Germ Defence digital behaviour change intervention via all primary care practices in England to reduce respiratory infections during the COVID-19 pandemic: an efficient cluster randomised controlled trial using the OpenSAFELY platform. *Implement Sci* 2023; 18: 67.

Zhong X, Ashiru-Oredope D, Pate A, Martin GP, Sharma A, et al. Clinical and health inequality risk factors for non-COVID-related sepsis during the global COVID-19 pandemic: a national case-control and cohort study. *EClinicalMedicine* 2023; 66: 102321.

Zheng B, Tazare J, Nab L, Green AC, Curtis HJ, et al. Comparative effectiveness of nirmatrelvir/ritonavir versus sotrovimab and molnupiravir for preventing severe COVID-19 outcomes in non-hospitalised high-risk patients during Omicron waves: observational cohort study using the OpenSAFELY platform. *Lancet Reg Health Eur* 2023; 34: 100741.

Zheng B, Campbell J, Carr EJ, Tazare J, Nab L, et al. Comparative effectiveness of sotrovimab and molnupiravir for preventing severe COVID-19 outcomes in patients on kidney replacement therapy: observational study using the OpenSAFELY-UKRR and SRR databases. *Clin Kidney J* 2023; 16: 2048–2058.

Russell MD, Massey J, Roddy E, MacKenna B, Bacon S, et al. Gout incidence and management during the COVID-19 pandemic in England, UK: a nationwide observational study using OpenSAFELY. *Lancet Rheumatol* 2023; 5: e622–e632.

Yang Y-T, Zhong X, Fahmi A, Watts S, Ashcroft DM, et al. The impact of the COVID-19 pandemic on the treatment of common infections in primary care and the change to antibiotic prescribing in England. *Antimicrob Resist Infect Control* 2023; 12: 102.

Horne EMF, Hulme WJ, Keogh RH, Palmer TM, Williamson EJ, et al. Challenges in Estimating Effectiveness of 2 Doses of COVID-19 Vaccines Beyond 6 Months in England. *Am J Epidemiol* 2023; 193: 227-231.

Macdonald O, Green A, Walker A, Curtis H, Croker R, et al. Impact of the COVID-19 pandemic on antipsychotic prescribing in individuals with autism, dementia, learning disability, serious mental illness or living in a care home: a federated analysis of 59 million patients' primary care records in situ using OpenSAFELY. *BMJ Ment Health* 2023; 26: e300775.

Lemanska A, Andrews C, Fisher L, Bacon S, Frampton AE, et al. Healthcare in England was affected by the COVID-19 pandemic across the pancreatic cancer pathway: A cohort study using OpenSAFELY-TPP. *eLife* 2023; 12: e85332.

Fisher L, Curtis HJ, Croker R, Wiedemann M, Speed V, et al. Eleven key measures for monitoring general practice clinical activity during COVID-19: A retrospective cohort study using 48 million adults' primary care records in England through OpenSAFELY. *eLife* 2023; 12: e84673.

McElroy E, Herrett E, Patel K, Piehlmaier DM, Gessa GD, et al. Living alone and mental health: parallel analyses in UK longitudinal population surveys and electronic health records prior to and during the COVID-19 pandemic. *BMJ Ment Health* 2023; 26: e300842.

Zhong X, Pate A, Yang Y-T, Fahmi A, Ashcroft DM, et al. The impact of COVID-19 on antibiotic prescribing in primary care in England: Evaluation and risk prediction of appropriateness of type and repeat prescribing. *J Infect* 2023; 87: 1–11.

Yang Y-T, Wong D, Ashcroft DM, Massey J, MacKenna B, et al. Repeated antibiotic exposure and risk of hospitalisation and death following COVID-19 infection (OpenSAFELY): a matched case-control study. *EClinicalMedicine* 2023; 61: 102064.

Costello RE, Tazare J, Piehlmaier D, Herrett E, Parker EPK, et al. Ethnic differences in the indirect effects of the COVID-19 pandemic on clinical monitoring and hospitalisations for non-COVID conditions in England: a population-based, observational cohort study using the OpenSAFELY platform. *EClinicalMedicine* 2023; 61: 102077.

McInerney CD, Kotzé A, Bacon S, Cutting JE, Fisher L, et al. Postoperative mortality and complications in patients with and without pre-operative SARS-CoV-2 infection: a service evaluation of 24 million linked records using OpenSAFELY. Anaesthesia 2023; 78: 692–700.

Lemanska A, Andrews C, Fisher L, Butler-Cole B, Mehrkar A, et al. A National Audit of Pancreatic Enzyme Prescribing in Pancreatic Cancer from 2015 to 2023 in England Using OpenSAFELY-TPP. *Semin Oncol Nurs* 2023; 39: 151439.

Zhong X, Pate A, Yang Y-T, Fahmi A, Ashcroft DM, et al. Impact of COVID-19 on broad-spectrum antibiotic prescribing for common infections in primary care in England: a time-series analyses using OpenSAFELY and effects of predictors including deprivation. *Lancet Reg Health Eur* 2023; 30: 100653.

Hulme WJ, Williamson E, Horne EMF, Green A, McDonald HI, et al. Challenges in Estimating the Effectiveness of COVID-19 Vaccination Using Observational Data. *Ann Intern Med* 2023; 176: 685–693.

Parker EPK, Horne EMF, Hulme WJ, Tazare J, Zheng B, et al. Comparative effectiveness of two- and three-dose COVID-19 vaccination schedules involving AZD1222 and BNT162b2 in people with kidney disease: a linked OpenSAFELY and UK Renal Registry cohort study. *Lancet Reg Health Eur* 2023; 30: 100636.

Nab L, Parker EPK, Andrews CD, Hulme WJ, Fisher L, et al. Changes in COVID-19-related mortality across key demographic and clinical subgroups in England from 2020 to 2022: a retrospective cohort study using the OpenSAFELY platform. Lancet Public Health 2023; 8: e364–e377.

Curtis HJ, MacKenna B, Wiedemann M, Fisher L, Croker R, et al. OpenSAFELY NHS Service Restoration Observatory 2: changes in primary care clinical activity in England during the COVID-19 pandemic. *Br J Gen Pract* 2023; 73: e318–e331.

Hulme WJ, Horne EMF, Parker EPK, Keogh RH, Williamson EJ, et al. Comparative effectiveness of BNT162b2 versus mRNA-1273 covid-19 vaccine boosting in England: matched cohort study in OpenSAFELY-TPP. *BMJ* 2023; 380: e072808.

Parker EP, Tazare J, Hulme WJ, Bates C, Carr EJ, et al. Factors associated with COVID-19 vaccine uptake in people with kidney disease: an OpenSAFELY cohort study. *BMJ Open* 2023; 13: e066164.

Hopcroft LE, Curtis HJ, Brown AD, Hulme WJ, Andrews CD, et al. First dose COVID-19 vaccine coverage amongst adolescents and children in England: an analysis of 3.21 million patients' primary care records in situ using OpenSAFELY. *Wellcome Open Res* 2023; 8: 70.

Green ACA, Curtis HJ, Higgins R, Nab L, Mahalingasivam V, et al. Trends, variation, and clinical characteristics of recipients of antiviral drugs and neutralising monoclonal antibodies for covid-19 in community settings: retrospective, descriptive cohort study of 23.4 million people in OpenSAFELY. *BMJ Med* 2023; 2: e000276.

Fisher L, Hopcroft LE, Rodgers S, Barrett J, Oliver K, et al. Changes in medication safety indicators in England throughout the covid-19 pandemic using OpenSAFELY: population based, retrospective cohort study of 57 million patients using federated analytics. *BMJ Med* 2023; 2: e000392.

Curtis HJ, Fisher L, Evans D, OpenSAFELY Collaborative, Bacon S, et al. Interpreting pathology test result values with comparators (< , >) in Electronic Health Records research: an OpenSAFELY short data report. *Wellcome Open Res* 2023; 8: 541.

Wing K, Grint DJ, Mathur R, Gibbs HP, Hickman G, et al. Association between household composition and severe COVID-19 outcomes in older people by ethnicity: an observational cohort study using the OpenSAFELY platform. *Int J Epidemiol* 2022; 51: 1745–1760.

Russell MD, Galloway JB, Andrews CD, MacKenna B, Goldacre B, et al. Incidence and management of inflammatory arthritis in England before and during the COVID-19 pandemic: a population-level cohort study using OpenSAFELY. *Lancet Rheumatol* 2022; 4: e853–e863.

Zheng B, Green ACA, Tazare J, Curtis HJ, Fisher L, et al. Comparative effectiveness of sotrovimab and molnupiravir for prevention of severe covid-19 outcomes in patients in the community: observational cohort study with the OpenSAFELY platform. *BMJ* 2022; 379: e071932.

Curtis HJ, Inglesby P, MacKenna B, Croker R, Hulme WJ, et al. Recording of 'COVID-19 vaccine declined': a cohort study on 57.9 million National Health Service patients' records in situ using OpenSAFELY, England, 8 December 2020 to 25 May 2021. *Euro Surveill* 2022; 27: 2100885.

Walker JL, Schultze A, Tazare J, Tamborska A, Singh B, et al. Safety of COVID-19 vaccination and acute neurological events: A self-controlled case series in England using the OpenSAFELY platform. *Vaccine* 2022; 40: 4479–4487.

Hulme WJ, Williamson EJ, Green ACA, Bhaskaran K, McDonald HI, et al. Comparative effectiveness of ChAdOx1 versus BNT162b2 covid-19 vaccines in health and social care workers in England: cohort study using OpenSAFELY. *BMJ* 2022; 378: e068946.

Horne EMF, Hulme WJ, Keogh RH, Palmer TM, Williamson EJ, et al. Waning effectiveness of BNT162b2 and ChAdOx1 covid-19 vaccines over six months since second dose: OpenSAFELY cohort study using linked electronic health records. *BMJ* 2022; 378: e071249.

Green A, Curtis H, Hulme W, Williamson E, McDonald H, et al. Describing the population experiencing COVID-19 vaccine breakthrough following second vaccination in England: a cohort study from OpenSAFELY. *BMC Med* 2022; 20: 243.

Wong AY, Tomlinson L, Brown JP, Elson W, Walker AJ, et al. Association between oral anticoagulants and COVID-19-related outcomes: a population-based cohort study. *Br J Gen Pract* 2022; 72: e456–e463.

Shah SA, Brophy S, Kennedy J, Fisher L, Walker A, et al. Impact of first UK COVID-19 lockdown on hospital admissions: Interrupted time series study of 32 million people. *EClinicalMedicine* 2022; 49: 101462.

MacKenna B, Kennedy NA, Mehrkar A, Rowan A, Galloway J, et al. Risk of severe COVID-19 outcomes associated with immune-mediated inflammatory diseases and immune-modifying therapies: a nationwide cohort study in the OpenSAFELY platform. *Lancet Rheumatol 2022*; 4: e490–e506.

Thompson EJ, Williams DM, Walker AJ, Mitchell RE, Niedzwiedz CL, et al. Long COVID burden and risk factors in 10 UK longitudinal studies and electronic health records. *Nat Commun* 2022; 13: 3528.

Schultze A, Nightingale E, Evans D, Hulme W, Rosello A, et al. Mortality among Care Home Residents in England during the first and second waves of the COVID-19 pandemic: an observational study of 4.3 million adults over the age of 65. *Lancet Reg Health Eur* 2022; 14: 100295.

Fisher L, Speed V, Curtis HJ, Rentsch CT, Wong AYS, et al. Potentially inappropriate prescribing of DOACs to people with mechanical heart valves: A federated analysis of 57.9 million patients' primary care records in situ using OpenSAFELY. *Thromb Res* 2022; 211: 150–153.

Williamson EJ, Tazare J, Bhaskaran K, McDonald HI, Walker AJ, et al. Comparison of methods for predicting COVID-19-related death in the general population using the OpenSAFELY platform. *Diagn Progn Res* 2022; 6: 6.

MacKenna B, Curtis HJ, Walker AJ, Goldacre B. Trends and clinical characteristics of COVID-19 vaccine recipients: a federated analysis of 57.9 million patients' primary care records in situ using OpenSAFELY. *Br J Gen Pract* 2022; 72: 10.

Curtis HJ, MacKenna B, Croker R, Inglesby P, Walker AJ, et al. OpenSAFELY NHS Service Restoration Observatory 1: primary care clinical activity in England during the first wave of COVID-19. *Br J Gen Pract* 2022; 72: e63–e74.

Curtis HJ, Inglesby P, Morton CE, MacKenna B, Green A, et al. Trends and clinical characteristics of COVID-19 vaccine recipients: a federated analysis of 57.9 million patients' primary care records in situ using OpenSAFELY. *Br J Gen Pract* 2022; 72: e51–e62.

Bhaskaran K, Rentsch CT, Hickman G, Hulme WJ, Schultze A, et al. Overall and cause-specific hospitalisation and death after COVID-19 hospitalisation in England: A cohort study using linked primary care, secondary care, and death registration data in the OpenSAFELY platform. *PLoS Med* 2022; 19: e1003871.

Tazare J, Walker AJ, Tomlinson LA, Hickman G, Rentsch CT, et al. Rates of serious clinical outcomes in survivors of hospitalisation with COVID-19 in England: a descriptive cohort study within the OpenSAFELY platform. *Wellcome Open Res* 2022; 7: 142.

Andrews C, Schultze A, Curtis H, Hulme W, Tazare J, et al. OpenSAFELY: Representativeness of electronic health record platform OpenSAFELY-TPP data compared to the population of England. *Wellcome Open Res* 2022; 7: 191.

Walker AJ, MacKenna B, Inglesby P, Tomlinson L, Rentsch CT, et al. Clinical coding of long COVID in English primary care: a federated analysis of 58 million patient records in situ using OpenSAFELY. *Br J Gen Pract* 2021; 71: e806–e814.

OpenSAFELY Collaborative, Curtis HJ, MacKenna B, Walker AJ, Croker R, et al. OpenSAFELY: impact of national guidance on switching anticoagulant therapy during COVID-19 pandemic. *Open Heart* 2021; 8: e001784.

Wong AYS, Tomlinson LA, Brown JP, Elson W, Walker AJ, et al. Association between warfarin and COVID-19-related outcomes compared with direct oral anticoagulants: population-based cohort study. *J Hematol Oncol* 2021; 14: 172.

Williamson EJ, McDonald HI, Bhaskaran K, Walker AJ, Bacon S, et al. Risks of covid-19 hospital admission and death for people with learning disability: population-based cohort study using the OpenSAFELY platform. *BMJ* 2021; 374: n1592.

Wong AY, MacKenna B, Morton CE, Schultze A, Walker AJ, et al. Use of non-steroidal anti-inflammatory drugs and risk of death from COVID-19: an OpenSAFELY cohort analysis based on two cohorts. *Ann Rheum Dis* 2021; 80: 943–951.

Bhaskaran K, Bacon S, Evans SJ, Bates CJ, Rentsch CT, et al. Factors associated with deaths due to COVID-19 versus other causes: population-based cohort analysis of UK primary care data and linked national death registrations within the OpenSAFELY platform. *Lancet Reg Health Eur* 2021; 6: 100109.

Mathur R, Rentsch CT, Morton CE, Hulme WJ, Schultze A, et al. Ethnic differences in SARS-CoV-2 infection and COVID-19-related hospitalisation, intensive care unit admission, and death in 17 million adults in England: an observational cohort study using the OpenSAFELY platform. *Lancet* 2021; 397: 1711–1724.

Forbes H, Morton CE, Bacon S, McDonald HI, Minassian C, et al. Association between living with children and outcomes from covid-19: OpenSAFELY cohort study of 12 million adults in England. *BMJ* 2021; 372: n628.

Rentsch CT, DeVito NJ, MacKenna B, Morton CE, Bhaskaran K, et al. Effect of pre-exposure use of hydroxychloroquine on COVID-19 mortality: a population-based cohort study in patients with rheumatoid arthritis or systemic lupus erythematosus using the OpenSAFELY platform. *Lancet Rheumatol* 2021; 3: e19–e27.

Bhaskaran K, Rentsch CT, MacKenna B, Schultze A, Mehrkar A, et al. HIV infection and COVID-19 death: a population-based cohort analysis of UK primary care data and linked national death registrations within the OpenSAFELY platform. *Lancet HIV* 2021; 8: e24–e32.

Schultze A, Bates C, Cockburn J, MacKenna B, Nightingale E, et al. Identifying Care Home Residents in Electronic Health Records - An OpenSAFELY Short Data Report. *Wellcome Open Res* 2021; 6: 90.

Rowan A, Bates C, Hulme W, Evans D, Davy S, et al. A comprehensive high-cost drugs dataset from the NHS in England - An OpenSAFELY-TPP Short Data Report. Wellcome Open Res 2021; 6: 360.

Schultze A, Walker AJ, MacKenna B, Morton CE, Bhaskaran K, et al. Risk of COVID-19-related death among patients with chronic obstructive pulmonary disease or asthma prescribed inhaled corticosteroids: an observational cohort study using the OpenSAFELY platform. Lancet Respir Med 2020; 8: 1106–1120.

Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature 2020; 584: 430–436.

Work with us:
team@opensafely.org

Print projects go out of date quickly; everything in this booklet was correct when we went to print in early November 2024.

For more up-to-date information, find us online:

Bennett Institute website:	bennett.ox.ac.uk
OpenSAFELY website:	www.opensafely.org

You can follow us on social media. We're active on:

LinkedIn	www.linkedin.com/company/bennettoxford/
Bluesky	bsky.app/profile/bennettoxford.bsky.social
X	x.com/BennettOxford

With warmest thanks to all our supporters and funders.
You're all wonderful.

Thanks to these people for helping us make this booklet:

Ben Ainsworth
Diane Ashiru-Oredope
Seb Bacon
Klaudia Budniak
Louis Fisher
Millie Green
Ben Goldacre
Liam Hart
Eli Holderness
William Hulme
Francine Jury
Frederica Longfoot
Brian MacKenna
Amir Mehrkar
Jessica Morley
Ed Parker
Mark Russell
Andrea Schaffer
Matt Shinn
Tjeerd van Staa
Catherine Stables
Pete Stokes
Laurie Tomlinson
Sonia Turcotte
Giles Turnbull
Caroline Walters
Venexia Walker

Published by The Bennett Institute for Applied Data Science, 2024

> The team at OpenSAFELY demonstrates an approach to sensitive health data research that puts patients and the public's expectations at the heart of its work. It has gone far above the threshold of compliance to show what a trustworthy research environment can be, from its development and use of synthetic data, public logs and reusable tools, to its patient panel and explainer video. I am looking forward to seeing how its use extends beyond COVID-19 and paves the way for future health data research.
>
> – Nicola Hamilton, Understanding Patient Data

> medConfidential have often scrutinised data projects that tried to fudge their way around privacy concerns. Having tried everything else, finally in OpenSAFELY there is a platform currently making the right decisions, protecting patients and serving researchers by being consensual, safe, and transparent in its whole design. All it took was the right ethos, and the right technical skills!
>
> – Sam Smith, medConfidential


OpenSAFELY