



# SAIL DATABANK



Ymchwil Iechyd  
a Gofal **Cymru**  
Health and Care  
Research **Wales**



**2021 - 2022**  
**Annual report**



# CONTENTS

**03** INTRODUCTION

**06** ORGANISATIONAL CHART

**08** KEY PARTNERSHIPS & COLLABORATIONS

**16** ADDITIONAL PARTNERSHIPS & COLLABORATIONS

**19** OUR YEAR IN NUMBERS

**20** WORKSTREAMS

**25** DEVELOPING THE NEXT GENERATION

**26** PUBLIC INVOLVEMENT & ENGAGEMENT

# INTRODUCTION



For SAIL Databank, as for the UK generally, this has been a year of transition and change. 2021 saw a gradual reduction of the restrictions imposed as part of the initial COVID-19 emergency response, followed by further uncertainty and disruption caused by the Omicron variant. SAIL Databank has followed this theme of transition, with many short-term COVID emergency planning projects concluding and being replaced by longer-term research into the effects of the pandemic and how to adapt going forward. We have also seen a significant rise in the numbers of non-COVID related research projects approaching SAIL for support.

Perhaps the most exciting developments have been the innovations from the SAIL Technical and Governance teams, responding to challenges and demands from the UK and international research community. COVID-19 created an increased emphasis on a co-ordinated national approach to research to best serve the needs of the emergency response. Health Data Research UK (HDRUK) became the focus for co-ordination and monitoring of COVID related research. SAIL has been a key partner in the HDR-UK programme since its inception and worked to support the rollout of their UK-wide Innovation Platform to provide a single catalogue and application point for data held in any of the recognised Trusted Research Environments (TREs) across the UK. We have also integrated our governance and application processes to process applications for

access to SAIL data via this route. Whilst the work by HDRUK and others in streamlining the process of getting approvals to access data is a significant step forward, the eventual vision is for researchers to be able to access all the data they need through a single point of access.

During 2021 SAIL was at the forefront of developing governance frameworks and technologies to meet this requirement. The HDRUK convened the International COVID-19 Data Alliance (ICODA) to enable data analysis to be safely undertaken from an external platform (the ICODA Workbench) on data within SAIL using several data access models depending on the data sharing permissions in place. By the end of the pilot project (October 2021) data from 38 contributing centres worldwide was securely hosted within SAIL and analysed by researchers via the ICODA Workbench. This exemplifies the speed and efficiency with which SAIL can enable complex international research, proving the concept that queries can be run on data within one or more TREs from a single external platform, and for the first time allowing researchers to conduct research on data held in a secure TRE using an external access platform.

Whilst a significant step, this model has the disadvantage that, for security reasons, data must remain within its host environment and only the results of queries can travel. This makes gathering and comparing data held within a number of TRES difficult. During 2021-22 the SAIL Technical Team developed a technology platform which potentially enables secure transfer of data between a network of agreed endpoints, subject to a variety of configurable parameters. This model allows data to be compared side-by-side and analysed as a whole, and data can be accessed from any approved TRE. Obviously, development of this enabling technology is just the starting point, and several challenges remain if such a vision is to become reality.

As significant as the challenge is, there is good reason to be hopeful. Creating an environment where researchers can access data from around the globe from a single access point, whilst maintaining the highest standards of security and integrity, would have seemed a distant dream in 2019. The unique challenges of COVID-19 created an unprecedented momentum for change, enabling new working models which would previously have taken decades to develop. SAIL Databank has been at the forefront of this historic shift, helping data science move beyond the realm of academic research and become a front-line, real-time tool for health services and governments in a global pandemic. The impetus to realise the benefits of greater access to data shows no signs of slowing. Just as the challenges of this vision are enormous, the potential benefits to the research community and to society are equally compelling. SAIL Databank stands ready to face these challenges and to continue to maximise the unique opportunities for change, improvement and ongoing technological advancement afforded by the current climate.

*David Ford*

**Congratulations!...**

to **Professor Lyons** on his **OBE** for services to Research, Innovation and Public Health.



**Professor David Ford**  
Co-director  
SAIL Databank



**Professor Kerina Jones**  
Co-investigator  
SAIL Databank



**Professor Ronan Lyons OBE**  
Co-director  
SAIL Databank

# SAIL Databank Co-director becomes fellow of the Academy of Medical Sciences.

Professor Ronan Lyons, Clinical Professor of Public Health at Swansea University Medical School and one of the two Directors at SAIL Databank, became one of 60 outstanding biomedical and health scientists admitted to the Academy's influential Fellowship. The Academy of Medical Sciences is the independent body in the UK representing the diversity of medical science.

During the pandemic, Ronan and the team have used insights from the rich health data in SAIL Databank to support policy decisions to protect the public, including providing intelligence to the Welsh Government's Technical Advisory Group and subsequently feeding into the UK's SAGE (Scientific Advisory Group for Emergencies).

“I am delighted and honoured to be selected as a Fellow by the Academy of Medical Sciences. This undoubtedly reflects the widespread appreciation of research conducted using the SAIL Databank...

...and our dedication to the advancement of health research through our many collaborations across the UK and around the world”.

- Professor Ronan Lyons OBE

# FUNDERS

# MANAGEMENT

# OPERATIONAL

# EXTERNAL GROUPS



Scientific External Advisory Board

Information Governance Review Panel

Consumer Panel

Data Management Group

SAIL User Forum

\* Staff supported by multiple funders

# KEY PARTNERSHIPS & COLLABORATIONS

## COVID-19

### School openings and COVID infection risks.

A study which reported in May 2021 provided evidence that the opening of schools in Wales was not associated with an increased COVID-19 risk to staff. The first population-level study of COVID-19 transmission between pupils and staff in a school environment has shown that the opening of schools in Wales between September and December 2020 was not associated with an increased risk of staff testing positive. The study, which was carried out by the One Wales COVID-19 response team, using data hosted within SAIL, found that only pupils—not staff—were at increased risk of testing positive, following cases appearing within their own year group. Findings from the study directly informed Welsh Government’s COVID-19 Technical Advisory Group (TAG) and the UK’s Scientific Advisory Group for Emergencies (SAGE) and were part of the evidence which led to the decision to reopen schools March 2021 following prolonged closures. Full details of this project can be found here: <https://bmjpaedsopen.bmj.com/content/5/1/e001049/>

### SAIL data used to highlight variations in COVID vaccine uptake amongst minority ethnic groups.

In March 2021 investigations led by the One Wales COVID-19 Response team using SAIL data revealed that there was 11.3% less uptake of vaccines in black, asian, mixed and other ethnic groups aged 80+, 12.8% less for adults aged 75-79 years and 10% less for adults aged 70 – 74 years. This finding was particularly significant given that it was already established that minority ethnic groups were disproportionately affected by COVID-19, with infection rates higher and the severity of the reaction in those that tested positive to the virus significantly worse. Malorie Perry, a Senior Epidemiologist at Public Health Wales, at the time said “The availability of this information has enabled NHS colleagues to evaluate and tailor the immunisation programme in Wales making sure that the vaccine programme is available to all. As the vaccine programme continues so does our monitoring using this anonymised data analysis method. We continue to look at equity in vaccination uptake and it is

**“The availability of this information has enabled NHS colleagues to evaluate and tailor the immunisation programme in Wales...”**

- Malorie Perry,  
Public Health Wales

hoped that this can be expanded to aid understanding around other groups disproportionately affected by the COVID-19 pandemic”. Full details are available at:

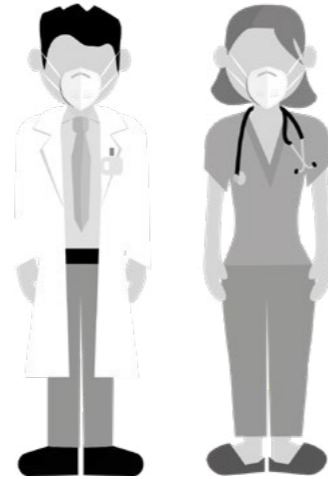
<https://www.sciencedirect.com/science/article/pii/S0264410X21011981?via%3Dihub#b0090>

## Research Using SAIL Databank Finds COVID-19 Vaccines Offer Effective Protection against infection for High-Risk Health Care Workers.

In one of the largest studies of its kind, new research followed nearly 83,000 health care workers in Wales, from 7th December 2020 to 30th September 2021, following first and second doses for up to 26 weeks post-vaccination. Published in the journal, *Vaccine*, the study revealed that there was a 90% vaccine uptake, with 79% receiving the Pfizer-BioNTech vaccine. Investigations found that the Pfizer-BioNTech vaccine was 85% effective after 2 weeks from the second dose and 52% from 22 weeks. The study also found there was equitable effectiveness across age, sex, ethnicity and deprivation groups. The research was undertaken as part of the One Wales collaboration and National Core Studies – Data and Connectivity: COVID-19 Vaccines group Pharmacovigilance (DaC-VaP) led by Professor Aziz Sheikh, which is a UK-wide Team Science collaboration involving members from academia, the NHS and the public. Professor Aziz Sheikh, Principal Investigator for DaC-VaP, “This large national study undertaken across Wales clearly demonstrates the substantial protection vaccination offers to healthcare workers – who are a high risk group – against SARS-CoV-2 infection.

<https://www.sciencedirect.com/science/article/pii/S0264410X21015358?via%3Dihub>

## COVID-19



The findings for comparable effectiveness in people of different ages, both males and females, across ethnic groups and in people of different socioeconomic backgrounds is very encouraging. This research demonstrates the outstanding health data science infrastructure in Wales, which is now providing insights to guide health policy and public health decision making not only for the UK but also the rest of the world.” The study team anticipate these results will provide further evidence for the Welsh vaccination strategy, which advises that frontline health and social care workers

receive a booster vaccine after 6 months. It will also provide national population estimates over time of vaccine uptake and effectiveness for healthcare workers in relation to COVID-19 infection, as well as estimates of uptake and effectiveness across a range of sociodemographic characteristics, most notably health care occupation. The full research paper can be found here: <https://doi.org/10.1016/j.vaccine.2021.11.061>

**“This research demonstrates the outstanding health data science infrastructure in Wales, which is now providing insights to guide health policy and public health decision making not only for the UK but also the rest of the world.”**

- Professor Aziz Sheikh,  
DaC-VaP

## New COVID-19 research on the impact of school closures suggests a widening wellbeing gap between rich and poor children.

On 07/09/21 BBC News reported findings of a project undertaken by the Health and Attainment of Pupils in a Primary Education Network (HAPPEN). Findings of the research suggests that school closures result in widening health inequalities and as schools return, measures will need to be in place to re-address the widened gap in physical health and wellbeing.

HAPPEN researchers collected data about children’s wellbeing during school closures between April and June 2020, using the ‘HAPPEN At Home’ survey. Participating HAPPEN Network primary schools in Wales were encouraged to invite parents and guardians to facilitate children taking part in the survey from home. Researchers then compared this information with similar data for the same period from 2019 and 2018.

Of the 1000+ eight to eleven-year olds surveyed, some children reported increased physical activity levels, sleep time and happiness when compared to previous years. But the most disadvantaged children; those receiving free school meals, were impacted most negatively by the school closures, the survey results suggest. The children receiving free school meals reported less physical activity levels and more takeaway consumption than their more affluent counterparts.

These children also reported generally eating less fruits and vegetables and achieved a lower level of educational self-assessment compared with previous years. It’s hoped that this research will provide policymakers and educational leaders with evidence that can inform future decision-making and ensure that adequate levels of support are given to those who need it most as schools welcome back their pupils.

The full research publication can be found here - <https://bmjopen.bmj.com/content/bmjopen/11/10/e051574.full.pdf>



### Award Successes

**SAIL Databank** and the ‘**One Wales**’ team have enjoyed award success in recognition of the response to COVID-19 through ‘flexibility and adaptability in a changing landscape.’



**Outstanding Impact on Health and Wellbeing**  
(Swansea University Research & Innovation Awards 2022)



**Recognised for Outstanding Response to COVID-19**  
(ADR UK Partnership Conference Awards 2022)

## Non-COVID-19

### SAIL Databank reveals link between household mental ill-health and developmental disorders in children.

New research has discovered the risk of children developing mental ill-health increases significantly when they have lived with someone who also has a common mental health disorder. The study, a collaboration between the Centre for Development, Evaluation, Complexity and Implementation in Public Health Improvement (DECIPHer) at Cardiff University's School of Medicine, The Wolfson's Centre for Young People's Mental Health at Cardiff University, Swansea University Medical School and Bangor University, showed children who had grown up living with someone with mental health difficulties were 63% more likely to experience a mental health issue, which includes but is not limited to anxiety, depression, anti-social behaviour and personality disorders. The study, also established links between household members who had mental ill health and other conditions in children such as personality or eating disorders.

They found that mental ill health among household members was related to a 42% increase in developmental disorders, which includes learning disabilities or attention-deficit disorders. Alongside mental health, children who experienced victimisation such as maltreatment or assault were 90% more likely to suffer childhood mental health issues and were 65% more likely to have developmental disorders. Researchers used SAIL Databank to securely analyse anonymised hospital admission and GP records which tracked 190,000 children living in Wales from birth to age 15. The study examined mental health symptoms, diagnoses and treatments, and associated mental health issues and developmental disorders such as learning disabilities or attention deficit. Read the full paper at: <https://bmcmmedicine.biomedcentral.com/articles/10.1186/s12916-021-02045-x>

### New study revealed that home adaptation interventions help to reduce emergency fall admissions in older people.

A new study has revealed that home adaptation interventions help to reduce emergency fall admissions amongst older people. Falls are common among older people, with 30% of people aged 65 and over and 50% of people aged 80 and over falling at least once per year. Falls cause increased morbidity, mortality and use of health care services and are a growing concern, with falls costing the NHS an estimated £2.3 billion per year. This study linked administrative, geographical, Care & Repair Cymru (C&RC) data and electronic health record data within the SAIL Databank to investigate fall outcomes following home adaptation interventions. The key objectives were to determine if home adaptation interventions carried out by C&RC led to a reduction in falls resulting in emergency admissions to hospital and investigate if there were differences in the risk of a fall based on area. The research involved 657,536 individuals aged 60 and over living in Wales between 1st January 2010 and 31st December 2017, 123,729 of which received a home adaptation service. The main findings of the paper were:

- Compared to the control group (who didn't receive any interventions), C&RC clients had higher odds of falling -this indicates that C&RC are successfully targeting a more vulnerable sub-population of older adults.
- C&RC adaptations/interventions in older people's homes reduced the odds of falling.
- Falls were more likely in females.
- Older age increased the chance of a fall.
- People living in deprived areas were more likely to fall.
- Increased frailty severity was linked to a higher chance of falling.

Read the complete publication here: <https://academic.oup.com/ageing/advance-article/doi/10.1093/ageing/afab201/6399893>

### SAIL Databank supports eight projects funded by Health & Care Research Wales' £6.5m funding award (October 2021)

Eight out of 23 newly funded research projects in Wales, funded by Health and Care Research Wales, are to use SAIL Databank based in Population Data Science at Swansea University in their pursuit to uncover potentially life-saving insights into a range of health and social care challenges. These key investments from their 2020-21 funding call are designed to 'fund high-quality research that addresses different research needs across Wales'. In response to the announcement, SAIL Databank Director and Professor of Informatics at Swansea University Medical School, Professor David Ford, said, "The commitment and investment from Health and Care Research Wales over the years has enabled SAIL Databank to grow and develop its data assets, data support services and its secure analytical environment. As a result, we are able to offer the research community a first-class data service that has helped to position Wales as a leader in the field of population data science research. We're excited to be collaborating with world-class researchers on these newly funded projects announced by Health and Care Research Wales and look forward to seeing the benefits they bring to the patients and people of Wales". Further details can be found here: <https://popdatasci.swan.ac.uk/sail-projects-funding/>

### British Heart Foundation fund research to explore link between depression and heart disease in Wales.

Previous research has shown that people who experience depression are more likely to develop heart disease than those who don't. The reasons behind this are not fully understood. This new study to investigate cardiovascular risk factors and treatments in patients with depression aims to find out why. The study, supported by a British Heart Foundation investment grant of over £120,000, will use the SAIL Databank Trusted Research Environment (TRE) as part of this ground-breaking investigation early next year. Preliminary work from this group has shown that patients with depression are less likely than patients without depression to have their cholesterol checked after coronary angioplasty and stenting treatment; a procedure to unblock the main blood vessels supplying the heart. This was also most likely to be the case in women with depression than in men. This new research aims to explore whether people with depression in the wider population without pre-existing heart disease are less likely to have these risk factors assessed and treated as effectively as those without depression. In addition, the researchers want to examine whether socio-economic status and where someone lives, for example in urban or rural settings, influences an individual's likelihood of risk factor assessment and effective treatment. Further details can be found here: <https://popdatasci.swan.ac.uk/british-heart-foundation-funded-research-to-explore-link-between-depression-and-heart-disease-in-wales/>

# Non-COVID-19

## Study into link between mental health and substance abuse issues on young people's mortality.

New research which has been published online in the journal, *Clinical Epidemiology*, has revealed that young people with a history of mental health and drug and alcohol issues are more likely to die than those with a mental illness or who use substances alone. The study found that more than 70 per cent of young people with substance use mentioned in their health data also had a record relating to mental illness.

The study also found that males and those from the most deprived areas are at greatest risk of death. Researchers also found that while incidence of young people with both mental illness and substance use in their hospital admissions data remained stable, in GP practice records it decreased in the most deprived part of Wales and increased in the least deprived areas. This may be because the number of young people with these conditions in those deprived areas is reducing; alternatively, it may be because fewer young people in these areas are accessing services or are having this information recorded in their patient history. The full paper can be found here: [https://www.dovepress.com/articles.php?article\\_id=72260](https://www.dovepress.com/articles.php?article_id=72260)

**“The work of ADR Wales and SAIL Databank on ADR UK themes such as housing and communities, children and young people, health and well-being and particularly Covid-19 have constructively informed policy decisions and improved the lives of many people in Wales.”**

-Dr Emma Gordon,  
ADR UK Director

## £17m funding for ADR Wales' collaborative data research in Wales.

Administrative Data Research Wales (ADR Wales) has been awarded £16,985,944 refunding until 2026 as part of the £90 million UK wide Administrative Data Research UK (ADR UK) investment by the Economic and Social Research Council (ESRC), part of UKRI.

ADR Wales is a well-established programme which aligns its work with the key areas identified in the Welsh Government's Programme for Government 2021-2026 such as education, mental health and housing. It utilises the academic independence and expertise of a team of specialist researchers, analysts, and data scientists, working in partnership with SAIL Databank and played a pivotal role in informing understanding and subsequent decision making at policy level during the pandemic in both Wales and across the UK.

Dr Emma Gordon, ADR UK Director said: “We are delighted to award ADR Wales funding to continue their important administrative data research and data linkage programmes. The work of ADR Wales and SAIL Databank on ADR UK themes such as housing and communities, children and young people, health and well-being and particularly Covid-19 have constructively informed policy

decisions and improved the lives of many people in Wales. This funding extension is testament to the fantastic impact that ADR Wales and SAIL Databank have delivered so far”. Read more about this at: <https://popdatasci.swan.ac.uk/17m-funding-for-collaborative-data-research-in-wales/>

## University helping to improve medicine safety in early motherhood.

Researchers from Swansea University and SAIL Databank have teamed up with colleagues from across Europe to produce their first publication as part of the IMI ConcePTION project.

This £28 million Europe-wide project is funded by the Innovative Medicines Initiative (IMI); a joint undertaking, which receives support from the European Union's Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations (EFPIA).

IMI seeks to bridge the knowledge and data gap in the safety of medicines prescribed during pregnancy and breastfeeding. It aims to do this by creating a trusted information ecosystem using anonymised data from multi-national data sources.

Congenital Anomaly Register and Information Service (CARIS) Manager David Tucker said: “This has been an exciting project. It uses the routine data collected by CARIS to its full potential. This research is answering long-asked questions about children's survival, quality of life and educational achievements. These are questions which every parent asks and now we are beginning to provide some answers.”

Further details can be found at: <https://popdatasci.swan.ac.uk/university-helping-to-improve-medicine-safety-in-early-motherhood/>

**“This research is answering long-asked questions about children's survival, quality of life and educational achievements. These are questions which every parent asks and now we are beginning to provide some answers.”**

- David Tucker,  
Public Health Wales





# ADDITIONAL PARTNERSHIPS & COLLABORATIONS

## Professor Ronan Lyons discusses 'COVID Passes' on BBC Radio Wales.

Professor of Public Health and Director of the Secure eResearch Platform (SeRP), SAIL Databank and Population Data Science at Swansea University, Professor Ronan Lyons, was interviewed in October 2021 by Oliver Hides on BBC Radio Wales. Professor Lyons discussed the evidence gathered from the summer 2021 unlocking of restrictions to reintroduce large events and how this had informed the Welsh Government's decision as winter approached, to introduce COVID passes to allow people to access premises where large numbers of people could be present and social distancing would be impractical. <https://popdatasci.swan.ac.uk/professor-ronan-lyons-discusses-covid-passes-on-bbc-radio-wales/>

## World-class dementia data repository secures re-funding.

The Dementias Platform UK (DPUK) Data Portal - a world-class data repository for dementia research that operates on SeRP, technology developed by the SAIL Databank team – has secured re-funding in a multi-million-pound investment by The UK's Medical Research Council (MRC) in partnership with industry and the third sector. Further reading: <https://popdatasci.swan.ac.uk/swansea-university-awarded-2-4-million-renewal-funding-for-world-class-dementia-data-repository/>

## SAIL Databank to host DATAMIND.

In May 2021 HDR UK and the MRC announced a £2 million investment to establish a new UK-wide Hub to provide innovative data resources for mental health research and innovation. The Hub will improve the discoverability and usability of diverse data sources for research to help improve the lives of people with mental health problems. DATAMIND will operate across the four nations of the UK, bringing together expertise from the NHS, universities, charities, research institutes and industry. <https://popdatasci.swan.ac.uk/datamind-a-new-uk-data-hub-for-mental-health-research-joins-population-data-science-group/>

## SAIL Databank to support feasibility study for new 70-year national birth cohort study.

Researchers from University College London, Swansea, Edinburgh, Ulster, Queen's University Belfast, and Manchester Metropolitan universities are set to test approaches to setting up a major new UK-wide study that will follow babies born in the 2020s over many decades to understand how societal circumstances and events affect them. A £3 million investment, made by the ESRC, part of UKRI, will allow researchers to develop a two-year-long birth cohort feasibility study. This study will develop and test the design, methodology and viability of a full-scale Early Life Cohort Study that is likely to follow participants for more than 70 years, starting from 2024. <https://popdatasci.swan.ac.uk/swansea-researchers-involved-in-pilot-plans-for-new-70-year-national-birth-cohort-study/>

## The International COVID-19 Data Alliance (ICODA) programme.

ICODA has almost completed its initial pathfinder / proof of concept project, International Perinatal Outcomes in the Pandemic (iPOP), which focussed on investigating the impact of the COVID pandemic on key perinatal outcomes in over 38 countries around the world. Initial findings from the iPOP study have been published by the *Wellcome Trust* (<https://wellcomeopenresearch.org/articles/6-21>)

## QCovid – The UK's COVID-19 mortality prediction model evaluated in Wales using SAIL Databank.

This study uses Welsh population data to validate QCovid mortality prediction model and is a collaboration between Swansea Oxford, Edinburgh and Leicester Universities, as well as the Office for National Statistics. It resulted in validation of the QCovid model as being highly accurate in identifying individuals at high risk of death from COVID-19 within the Welsh population. Read the full study here: <https://ijpds.org/article/view/1697>

**New research published July 2021 reveals parents of children in care proceedings experience greater levels of health vulnerabilities, including mental health, substance use and injury related conditions.**

The research was carried out by the Family Justice Data Partnership, using SAIL data, and revealed in the years leading up to their children being involved in care proceedings parents in Wales were associated with higher levels of overall healthcare use, included elevated levels in emergency healthcare settings compared to parents with no family court involvement. <https://popdatasci.swan.ac.uk/wales-parents-of-children-in-care-proceedings-experience-greater-levels-of-health-vulnerabilities-including-mental-health-substance-use-and-injury-related-conditions/>

**Never too late to quit smoking for people with Multiple Sclerosis.**

New research recently published finds that quitting smoking may cause slowing of Multiple Sclerosis (MS) disease progression. For more information on smoking and MS visit: <https://www.mssociety.org.uk/care-and-support/everyday-living/smoking-and-ms>

**SAIL Databank to support five new COVID-19 projects funded by UKRI'S National Core Studies.**

Five of the nine successful projects from a £2m UKRI funding call to support UK Government's NCS programme are using SAIL Databank as a key support resource, and include research into vaccine impacts, caring for children with COVID-19 and the way recorded ethnicity is used in research. <https://popdatasci.swan.ac.uk/research-blog-covid-19-projects-funded-by-ukris-national-core-studies-to-use-sail-databank/>

**Welsh research finds Covid-19 infection rates lower amongst people who had experienced homelessness than the general population.**

Between 1 March 2020 and 1 March 2021, COVID-19 infection rates among people experiencing homelessness were 5%, compared to 6.9% amongst the general population of similar demographics suggesting that the £50m investment by Welsh Government to change homelessness policies during the pandemic may have had a significant positive impact in reducing infection. Read the publication here: <https://popdatasci.swan.ac.uk/welsh-research-finds-covid-19-infection-rates-lower-amongst-people-who-had-experienced-homelessness-than-the-general-population/>.

**Number of teenagers in care proceedings in Wales rises by almost 50% in last decade.**

According to new data analysis published by Nuffield Family Justice Observatory (NFJO), the number of adolescents subject to care proceedings rose from 219 to 323 between 2011/12 and 2019/20. The data also revealed high levels of mental health diagnoses among this group. Just over half of those in proceedings also had a mental health disorder recorded by their GP, compared to a third of their peers. Hospital admissions and A&E attendance for mental health disorders, injuries, and poisoning-related conditions were also markedly higher. <https://popdatasci.swan.ac.uk/number-of-teenagers-in-care-proceedings-in-wales-rises-by-almost-50-in-last-decade/>

**19 OUR YEAR IN NUMBERS**

**Our Year In Numbers**



**Research income secured**



**Created in Wales**

**Core Metrics**  
Reporting period: **2020/2021**

**Health and Care Research Wales infrastructure award to the group**



Direct funding awarded

**£882k**

Jobs created through direct funding



**Grants won during reporting period**

Grants won	Led by group	Group collaborating
Number	<b>3</b>	<b>134</b>
Value	<b>£4.8m</b>	<b>£32m</b>
Funding to Wales	<b>£4.8m</b>	<b>£4.2m</b>
Funding to group	<b>£681k</b>	<b>£1.2m</b>
Additional jobs created for Wales	<b>30</b>	<b>9</b>
Additional jobs created for group	<b>24</b>	<b>9</b>



Number of publications



Number of public engagement events



Number of public involvement opportunities

# WORKSTREAMS

## WORKSTREAM 1

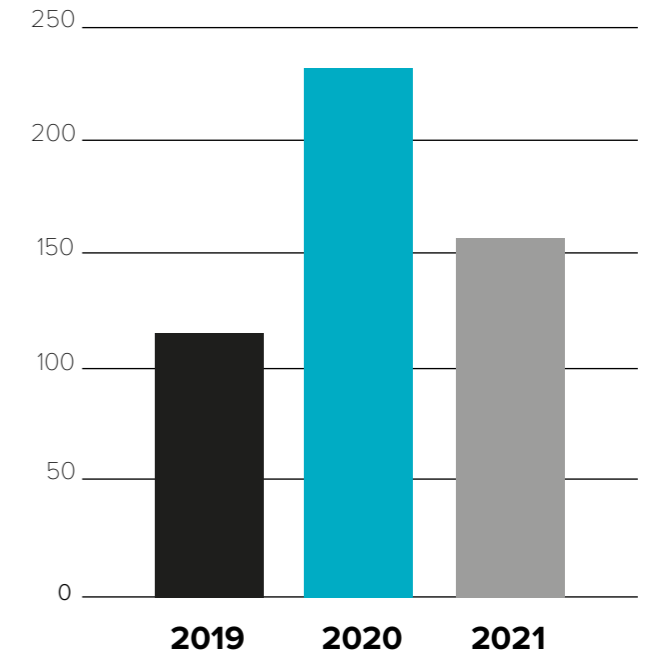
To run the core SAIL service, providing a comprehensive support service to research projects to enable access to SAIL data.

Throughout the year, the SAIL team has continued to offer a comprehensive range of support services to a wide portfolio of research projects. Key activities include:

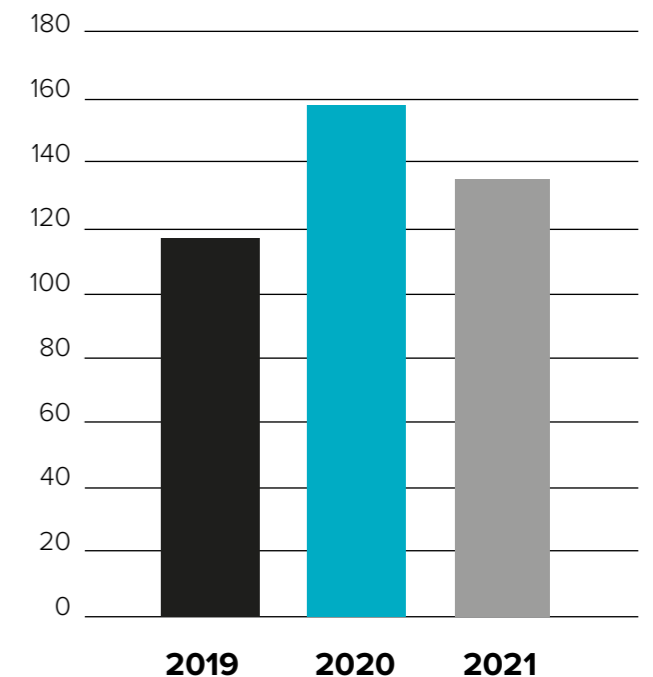
**Project Scoping** - The first contact that researchers have with SAIL is generally when they approach us with a project which requires scoping. Exceptional scoping volumes were seen during 2020 as a result of the pandemic showing the critical role SAIL had during the pandemic. 2021 saw a drop in overall scoping requests to around the same level as pre-pandemic. While the overall scoping number reduced in 2021, the number of scopes that were led by a member of the core SAIL team, rather than via an affiliated research platform, increased by around 35% compared to pre-pandemic levels. This suggests that more researchers are coming directly to SAIL for support for their project, possibly reflecting more complex or wide-ranging requirements.

**Governance Review** - When researchers want to use SAIL data for their project, they must make an application to gain approval from our independent Information Governance Review Panel (IGRP), and they must apply again if any aspect of their project changes. In 2021, the number of new applications has returned to 2019 levels, however the number of applications for approval for amendments is still increasing. The SAIL team provide a review service for each IGRP application before it is submitted to the Panel, to ensure it is of high quality and the required data is clearly defined to support IGRP review, data provisioning and future file output review. In 2022, we have also started monitoring the reasons for amendments to see if any common themes should be highlighted at initial application review, to minimise avoidable amendments.

Number of requests scoped



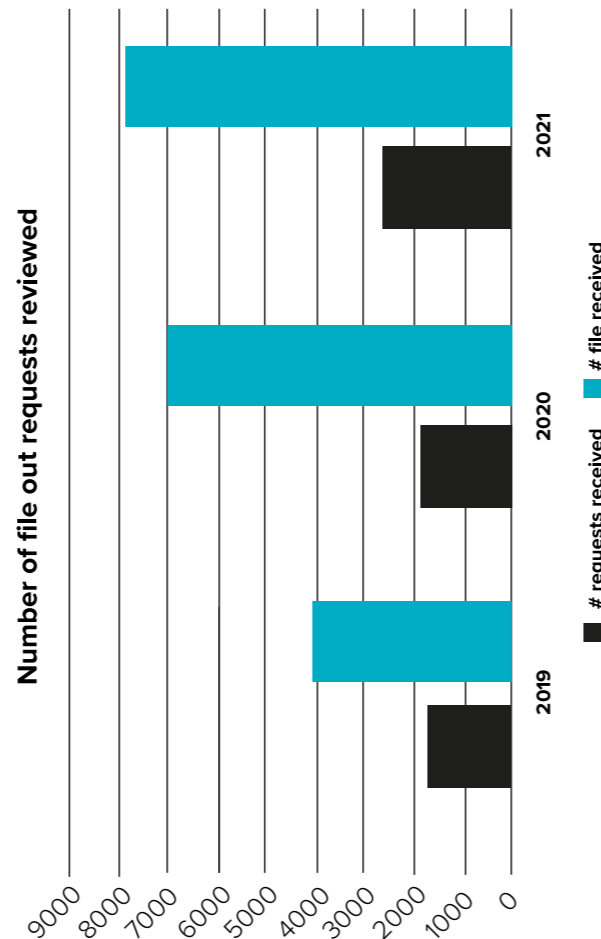
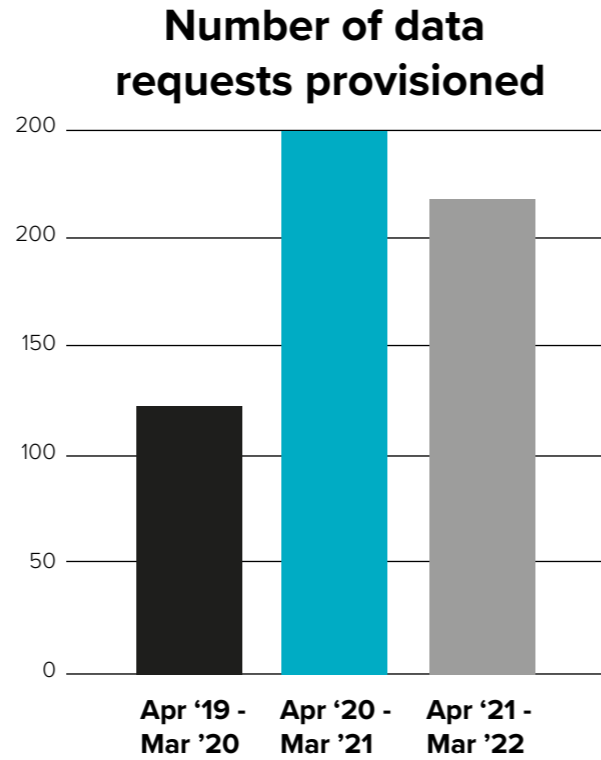
Number of internal IGRP reviews undertaken



**Data Provisioning** – Once a project is approved, we construct the project data views before they are made available to the researchers. This process ensures that the researchers receive only the data approved within their IGRP application and ensures adherence to the data providers requirements. We aim to provision project data within 30 days of application approval, excepting restricted or bespoke datasets. Between April 2021 and March 2022, 98% (158/161) of the project data provisions completed were done within 30 days of approval &/or data available. 2021/22 has seen a slight reduction in the number of project provision requests compare to 2020/21, however, the demand is still much higher than pre-2020.

**Project audits** – Our SAIL governance framework and external accreditation standards such as the ISO 27001 and the Digital Economy Act require us to randomly audit a proportion of projects to ensure they are operating within the terms of their IGRP approval. In total, 20 projects were randomly audited as part of the annual SAIL projects audit programme. No material breaches were identified, however a list of minor areas for improvement has been kept and is fed back to the user community to improve overall best practice.

**File out Reviews** – The only way researchers can access and analyse SAIL data is by logging into our secure systems. Our agreements with the NHS and other data providers mean that data itself cannot be extracted from this secure environment, but obviously researchers need a way to release the reports and results from their research for publication. This is done via the ‘file out’ review process, where a member of the SAIL team reviews material before authorising its release, to ensure that the content is line with the IGRP permissions in place for the project. Between January 2021 to December 2021 the team received over 2600 file out requests (an increase of ~42% on 2020) and reviewed over 7800 files (an increase of ~12% on 2020), each month exceeding our target of reviewing 90% of requests within 2 working days of receipt.



## WORKSTREAM 2

**Data acquisition to meet expressed requirements of our user base.**

In the period April 2021-March 2022 SAIL worked extensively to improve its data acquisition and views to meet the demands to the researchers using our service for both COVID and non-COVID related projects. Over this period SAIL secured access to the following 8 new data sources:

- Covid Sequencing Data
- Suspected Cancer Pathway
- Cancer network information system dataset
- Additional Looked After Children datasets from Welsh Government including data on adoptions, care leavers, birthday 19 and educational qualifications
- Higher Education Statistics Agency (HESA) Data
- National Survey for Wales 2020-21 dataset
- Pupil Level Annual Census
- Ministry of Justice data

In line with researcher demand, SAIL has also negotiated a more regular refresh of GP data (now monthly), making this dataset to as close as feasibly possible to ‘real-time’.

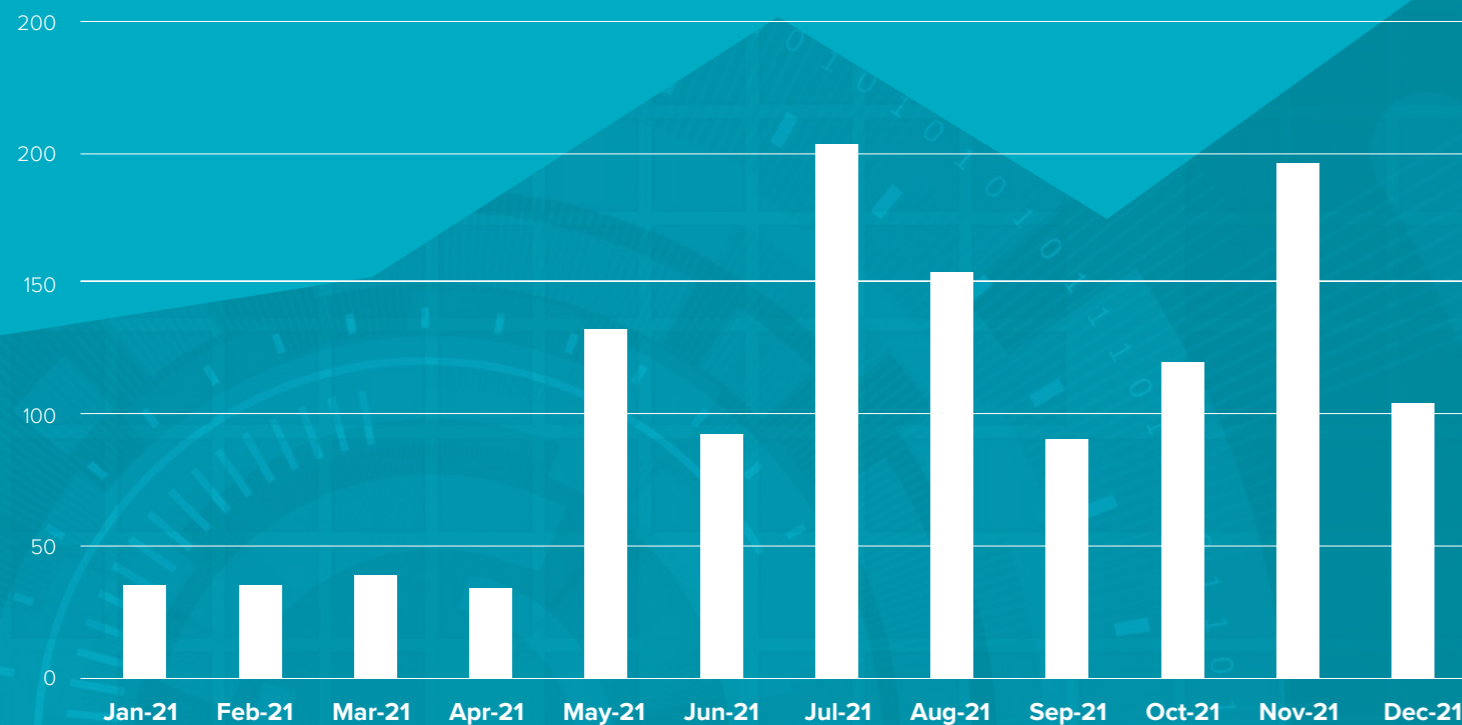
A key underpinning step in gaining access to a number of these datasets was that Digital Health and Care Wales (DHCW, our NHS partner organisation which provides data anonymisation services), gained Digital Economy Act accreditation during 2021, and a great deal of work was done by SAIL and DHCW staff to achieve this.

SAIL currently holds over 50 data sources, many of which are updated regularly. As new or updated data is received from data providers, SAIL undertake quality checks to ensure that it meets agreed criteria and does not include any elements that we do not have permission to hold, or which could be deemed as identifiable. Around 80% of all data sources brought into SAIL during 2021 were checked and released within 2 working days for existing data sources and 7 working days for new data sources. This increasingly allows projects to have access to the latest available data while at the same time maintaining the expected data quality levels.

The publicly available SAIL Data Catalogue has been updated, including more detail such as describing linkages between tables. The data catalogue is available via the SAIL website and also on the HDR UK Innovation Gateway, where SAIL currently leads the quality rankings for data descriptions.

**The graph that follows is that used by the analysts for their section on data quality review...**

## Number of data quality checks undertaken



## WORKSTREAM 3

**Increase research speed and efficiency by developing a set of re-usable tools (Concept Library) that will significantly reduce the time and effort required to conduct a complete analysis of a research question within the SAIL Databank.**

**The Concept library**, officially launched by SAIL in 2020, is a novel in-house developed tool for managing, publishing, and sharing definitions used in research. It is now fully operational and has been well-received by researchers being recognised as a leading tool in the Data Science community that both HDR UK and the Adolescent Mental Health Data Platform use for research. SAIL continues to work in to expand its functionality, with 40 new phenotypes being added since the end of 2021.

## Concept Library

The initial iteration of the Concept Library was adopted by HDR UK as the basis for their UK-wide Phenotype Library. As part of this collaboration SAIL has also developed a security model enabling the Phenotype Library to be available within the SAIL Gateway secure environment, so that researchers can access and immediately apply phenotype definitions to SAIL datasets.

**Several projects have been carried out to streamline and simplify analysis and use of data within SAIL:**

- Structural optimisation work has been carried out on the GP dataset. This has been well received by SAIL users who report that the dataset is easier to work with, and queries take less time to process.
- Research Ready versions of key datasets such as Welsh Results Reporting Service (WRRS), Welsh Longitudinal General Practice (WLGP) and Patient Episode Database for Wales (PEDW) have been produced which will help researchers navigate key datasets more easily: and we continue to look at developing a simplified view of Welsh Demographic Service Dataset (WSDS).
- Work is progressing to develop a dataset green-listing tool which will check datasets for disclosive information and redact such information at a granular level. Currently whole sections of datasets are withheld from researchers where the effort required for human checking for disclosive information is too great to be practicable. This automated tool will be quicker and more reliable than human checking and will enable far more data to be safely made available within specific datasets.
- Development has started on a tool that will help researchers check their file out requests for compliance with SAIL governance rules (e.g., no results less than 5 and the presence of sensitive codes) before they submit them for review.
- Provision of a new service to researchers to help them write more effective, efficient code when analysing data, which reduces processing time.
- Efforts to automate resource and time intensive processes such as data quality checks are ongoing.

## DEVELOPING THE NEXT GENERATION

In addition to supporting the current research community, SAIL Databank is committed to supporting and training the next generation of researchers and data scientists. During 2021-22 the SAIL Databank team developed a streamlined programme in coordination with the Faculty of Medicine, Health & Life Science at Swansea University to provide secure and supervised access to a set of approved data for students and their supervisors. This provides an invaluable opportunity to gain experience of using real-world anonymised individual-level population-scale data as part of undergraduate studies and projects. Early indications are that the initiative has been valuable and well received, and the SAIL team are investigating options to broaden the scope of the programme further during coming years.

Further information on the researchers and their projects can be accessed at: <https://popdatasci.swan.ac.uk/sail-databank-announces-an-inaugural-cohort-of-budding-student-researchers/>

# PUBLIC INVOLVEMENT & ENGAGEMENT



SAIL Databank's long-running Consumer Panel is now into its 11th year and continues to be at the heart of our commitment to public engagement and involvement in our work. The panel is comprised of 18 people, all of whom are members of the public. Panel members are invited to assist in many areas of the SAIL Databank process. They provide help and advice to researchers in developing their project ideas, shape funding applications by ensuring that all research includes a clear plan for maximising public benefit and advise researchers on good practice in the dissemination of their research findings. Some members of the Consumer Panel are also members of the Information Governance Review Panel, the independent body which makes decisions on applications to use SAIL data.

Despite the restrictions in place during 2021-22 due to COVID-19, SAIL was able to continue quarterly Consumer Panel meetings which took place virtually and in line with government guidelines. During this timeframe SAIL recruited new members to join the panel and will continue to recruit into 2022 due to the increasing demand for public and patient involvement and engagement in research.

**In addition to the ad hoc support to researchers described above, during the April 2021-March 2022 period the Consumer Panel provided support to several key initiatives, including:**

**Updating the SAIL website** (April 2021) - as the number of SAIL users, datasets and influence on public policy increased substantially during the pandemic, a need to improve the SAIL website was identified, to ensure it continued to communicate its activities accurately and effectively to researchers, policymakers and the public. SAIL Marketing and Communications Officer Chris Roberts presented a proposed revised website to the panel, who provided feedback on transparency, suggested additional information they thought should be accessible to the public via the website, made recommendations on the frequency with which they would recommend information to be refreshed and provided some exemplars of good website design from which SAIL could draw inspiration. This feedback was analysed and has been implemented in the new SAIL website.

**Information Governance Draft Review** (September 2021) - SAIL co-director Professor Kerina Jones presented a set of Information Governance good practice principles being developed for use by SAIL projects and the wider teams within the Swansea University Population Data Science group. The panel was asked to review and revise these draft information principles and provided feedback for further consideration and development.

**Networked Data Lab Wales** The Networked Data Lab, created by the independent charity the Health Foundation, is the first network of its kind, bringing together analytical teams from across the country to develop a deeper understanding of the factors affecting people's health in the UK. It is a collaboration with SAIL Databank, Public Health Wales, DHCW and Social Care Wales. On numerous occasions between April 2021 and March 2022, the project team met with the SAIL Consumer Panel to gain advice and perspective on the scope and direction of the project and the design and content of key outputs such as publications on the work of the project to be released to the public.

# Secure Anonymised Linkable

Containing billions of person-based records, SAIL Databank is a rich and trusted population databank. It improves lives by providing researchers with secure, linkable and anonymised data that can be accessed and analysed from anywhere in the world.

## Funded by



## Supported by



## Certified & accredited



## In partnership with



[www.SAILDatabank.com](http://www.SAILDatabank.com)

## Powered by

