

International Journal of Population Data Science

Journal Website: www.ijpds.org



The health informatics centre: a safe haven and trusted research environment enabling world-leading research

Laura M. Ward^{1,2}, Jenny Johnston¹, Keith R. Milburn¹, Christopher Hall¹, Claire Jones¹, Magalie Guignard-Duff¹, Susan Krueger¹, Gordon Milligan¹, Jill Anderson¹, Richard Walls¹, and Christian Cole^{1,2*}

Submission History

Submitted:	22/08/2025
Accepted:	17/12/2025
Published:	16/03/2026

¹Health Informatics Centre, School of Medicine, Willson House, University of Dundee, Dundee, DD2 1FD

²Division of Population Health and Genomics, School of Medicine, Ninewells Hospital, University of Dundee, DD1 9SY

Abstract

Introduction

The Health Informatics Centre (HIC) is a regional Scottish Safe Haven dedicated to secure data management, ensuring its integrity, confidentiality, and availability through a robust information governance framework.

Methods

As a data processor, HIC is responsible for the secure curation, storage, and provision of research data extracts. Research-ready data are made available to approved researchers via our customisable cloud Trusted Research Environment (TRE).

Results

The available granular data spans over 20 years, includes 2.1 million of the Scottish population, and HIC offers more than 170 datasets, with the most commonly used published with a digital object identifier. Data sources include clinical, hospital, laboratory, imaging, and research datasets which can be linked to new, and existing datasets. The data is quality-assured and released as project-specific extracts, ensuring robust privacy protection and research readiness. HIC's infrastructure is secure-by-design and supports high-performance computing, advanced data analytics, and is customisable to researcher's needs.

Conclusion

HIC has a long history of supporting a wide range of data-led research projects as a trusted and capable partner. At the time of publication 175 projects across academia, the NHS, and public sector organisations are active within HIC. Through adaptability, innovation and investment in people and infrastructure we have established a sustainable model which will continue to meet future needs and demands from world-leading research with sensitive data.

Keywords

Electronic Health Record; Trusted Research Environment; medical imaging; Data Governance; Common Data Model

Key features

- High quality granular regional data available on approximately 20% of the Scottish population with linkage to national data.
- Data sources available include medical data (cardiovascular, diabetes, emergency care, endocrinology, musculoskeletal, neurology, renal, respiratory), hospital admissions (Scottish Morbidity Records), laboratory results (biochemistry, haematology, immunology, microbiology, virology, pathology), clinical and diagnostic imaging, prescribing, demography, public health datasets, and research datasets (biobank, chronic pain).
- Integration across all HIC services; Research, Software, Governance, Infrastructure, and Data (including data standardisation to Observational Medical Outcomes Partnership (OMOP) common data model).
- High availability of customised cloud Trusted Research Environment (TRE) and next generation technologies including Artificial Intelligence/ Machine Learning (AI/ML).
- Collaborate with or provide support to researchers through the central contact email: HICSupport@dundee.ac.uk

*Corresponding Author:

Email Address: c.cole@dundee.ac.uk (Christian Cole)

Background

Research using routinely collected Electronic Health Record (EHR) data has led to significant advancements in health, improvements in healthcare, and service efficiency gains [1–3]. Leveraging large-scale, robust population-level data has become essential for research and innovation. The rapid access and use of patient data was crucial during the pandemic to understand, predict and respond to Covid-19 [4]. Improvements have been made in making patient data better suited to discovery and use across the entire population considering the deficiencies found during the pandemic [4–6]. Governments are recognising the importance of routinely collected EHR data for research and investing in how to maximise capacity securely [7–10].

The secondary use of routinely collected data has several advantages for researchers and funders. Compared to prospective patient study recruitment which may take years, it allows for faster data collection and is significantly more cost effective. For example, a multi-centre study on guideline adherence for antibiotic use analysed 10 years of data in a matter of weeks [11]. Additionally, routinely collected data provides large, population-level sample sizes that are less susceptible to the selection biases inherent in clinical trial or research recruitment. However, there are challenges; patient data reflects systemic and societal inequalities, where certain population groups do not access healthcare as often [12, 13] or receive suboptimal care [14]. Furthermore, as EHR data is primarily collected for the benefit of patient care and clinical use, the needs of research are rarely considered by system providers. This means that the data suffers from inconsistencies, errors, and gaps, which must not be underestimated when used for research [15, 16].

The Health Informatics Centre (HIC) at the University of Dundee is the East of Scotland regional Safe Haven, providing access to project-specific data extracts on behalf of NHS Tayside, NHS Fife, and NHS Forth Valley. Safe Havens operate on behalf of the Scottish Government and Chief Scientist Office to manage NHS Scotland data for researchers to access [17]. Access to pseudonymised patient data is provisioned via a Trusted Research Environment (TRE), sometimes called a Secure Data Environment (SDE), and operate within the “Five Safes” framework which ensures resilience in secure data management [18]. These safe settings support researchers by delivering secure computing environments for the analysis of EHR data whilst maintaining patient privacy. In Scotland, the Scottish Safe Haven Network (SSHNN) was established in 2014 comprising of four regional and one national Safe Haven, all operating under the Scottish Safe Haven Charter [19–21].

HIC originally focused on data linkage to support research [22]. In the early 2000s, patient records were predominantly paper based, which limited their potential research re-use. HIC’s initial purpose was to digitise records, such as community-dispensed prescriptions, and link these with secondary care data, such as hospital admission and prescribing data [23]. Today, HIC serves as a data processor for the data providers, managing data extracted from a wide range of clinical systems. This includes curating, linking, pseudonymising, and preparing the data for research use. Over the past twenty years, HIC has supported more than 1,000 projects generating over 400 research outputs. Here, we will

provide a description of HIC’s activities and data services, highlighting its independent, innovative and collaborative work with researchers using sensitive data.

Operational model

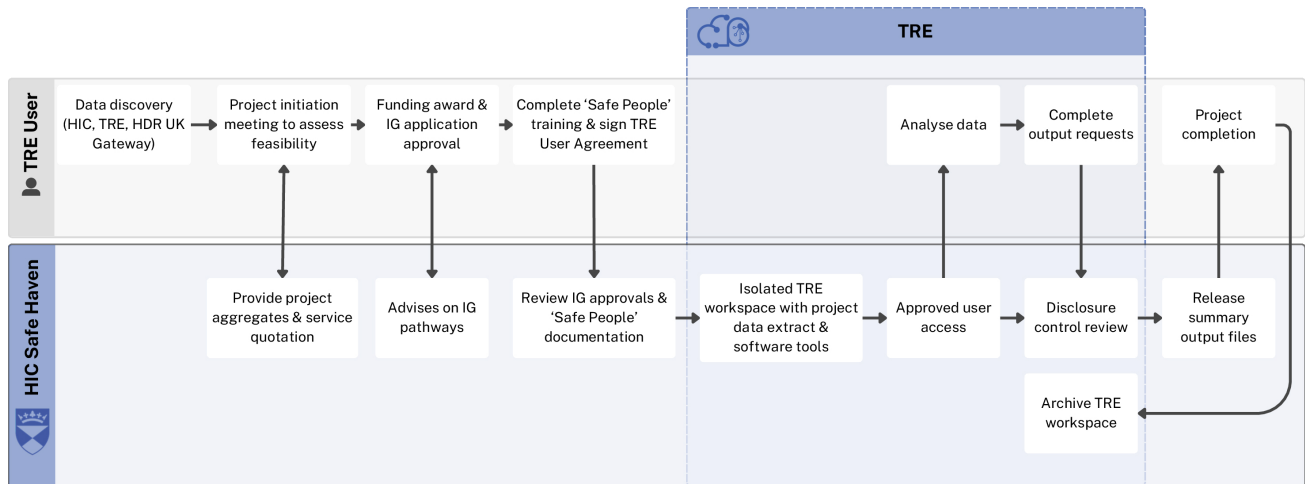
Information governance

Information Governance (IG) is the foundation of a Safe Haven and TRE. HIC’s IG framework encompasses people, processes, IT systems, and risk management. It ensures the integrity, confidentiality, and availability of data in compliance with regulatory and legal requirements, including the UK General Data Protection Regulation (UK GDPR), UK Data Protection Act 2018 and the Common Law Duty of Confidentiality. To support robust information security, HIC maintains an ISO 27001 accreditation, the leading international independent audit standard for managing information security risks, offering a structured approach to the protection of sensitive data. Continuous monitoring and auditing processes are implemented to assess and improve the effectiveness of our information security, processes and risk management measures. HIC also holds NHS England’s Data Security and Protection Toolkit security standard and is in the process of obtaining Digital Economy Act (DEA) accreditation from the UK Statistics Authority.

As a designated Safe Haven, HIC ensures the legal basis for processing patient data, operating under delegated authority from NHS Health Boards to enable researcher access to patient data within well-defined governance boundaries. To support timely access (within a few weeks), HIC has established a streamlined regional governance framework allowing projects to proceed quickly while maintaining robust security controls. For projects outside this delegated authority, alternative approval routes apply. HIC does not make research approval decisions but functions as a data processor on behalf of the data controller, relying on the appropriate project-specific IG pathway. These may include local Caldicott Guardian approvals (per NHS health board), Research Ethics Committees, NHS Research and Development approvals, or Public Benefit and Privacy Panel approvals. While these processes vary depending on the research project, they all ensure the secondary use of patient data is appropriate and proportionate. The ‘Safe People’ principle ensures that individuals accessing sensitive data are authorised, appropriately trained, and trusted. Users must provide evidence of completing and passing mandatory IG training, along with a signed TRE User Agreement that outlines roles and responsibilities for TRE Users, their organisation, and HIC, as the TRE provider. All documentation is maintained within an information security management system. Researchers are encouraged to engage with HIC at pre-award stage where we can consult on IG processes, project support, and feasibility. An end-to-end data access pathway can be found in Figure 1.

Securely managing and maintaining sensitive data for research, such as EHRs, is labour and resource intensive. It requires expertise in data engineering, data analysis, data management, data security, IG including disclosure control, and infrastructure maintenance. As part of the University of

Figure 1: The HIC data project life cycle with separate channels for the researcher (TRE User) and Safe Haven (HIC as the TRE provider)



Dundee, HIC are a Scottish Registered Charity (SCO15096), and *do not own nor sell data*, but operate on a cost-recovery basis to meet staff, TRE infrastructure and data management costs. All HIC's operational costs are recovered through grant or project funding with no core support or subsidy.

Transparency is crucial to maintain trust and is something that the public often demand when data about them is used for research [9, 24]. Public involvement was integral to the *Standardised Architecture of Trusted Research Environments* (SATRE) research project, with the version 1 specification co-designed with public leads [25]. SATRE resulted in the establishment of a data use register that documents the projects supported by HIC [26]. Additional operational changes within HIC includes consolidating the Executive and Information Governance Committee's into a single Strategic Board. This reflects the maturity of our digital and IG processes and ensures proportionate, efficient oversight. HIC contributes to open science by making internal developments available through our open-source GitHub repository [27]. Standard operating procedures and documentation are freely accessible online at <https://www.dundee.ac.uk/hic>, reinforcing HIC's commitment to responsible data management.

Architecture and information technology

Data security

As a data processor, HIC must manage sensitive data safely and securely, operating under the Five Safes, ISO 27001 accreditation and ensuring legal basis are met. Figure 2 shows a schematic view of the HIC infrastructure, including data flows and their approvals. This demonstrates the separation between the NHS, Cloud and University of Dundee networks and how the Five Safes are implemented in practice. Individual-level identifiable NHS data is received via secure encrypted transmission, which is then stored, quality assured, and processed on secure servers with secure work areas by only specifically approved HIC staff. Data is accessed as a project-specific extract provisioned in isolated TRE workspaces when all requirements are in place. Approved

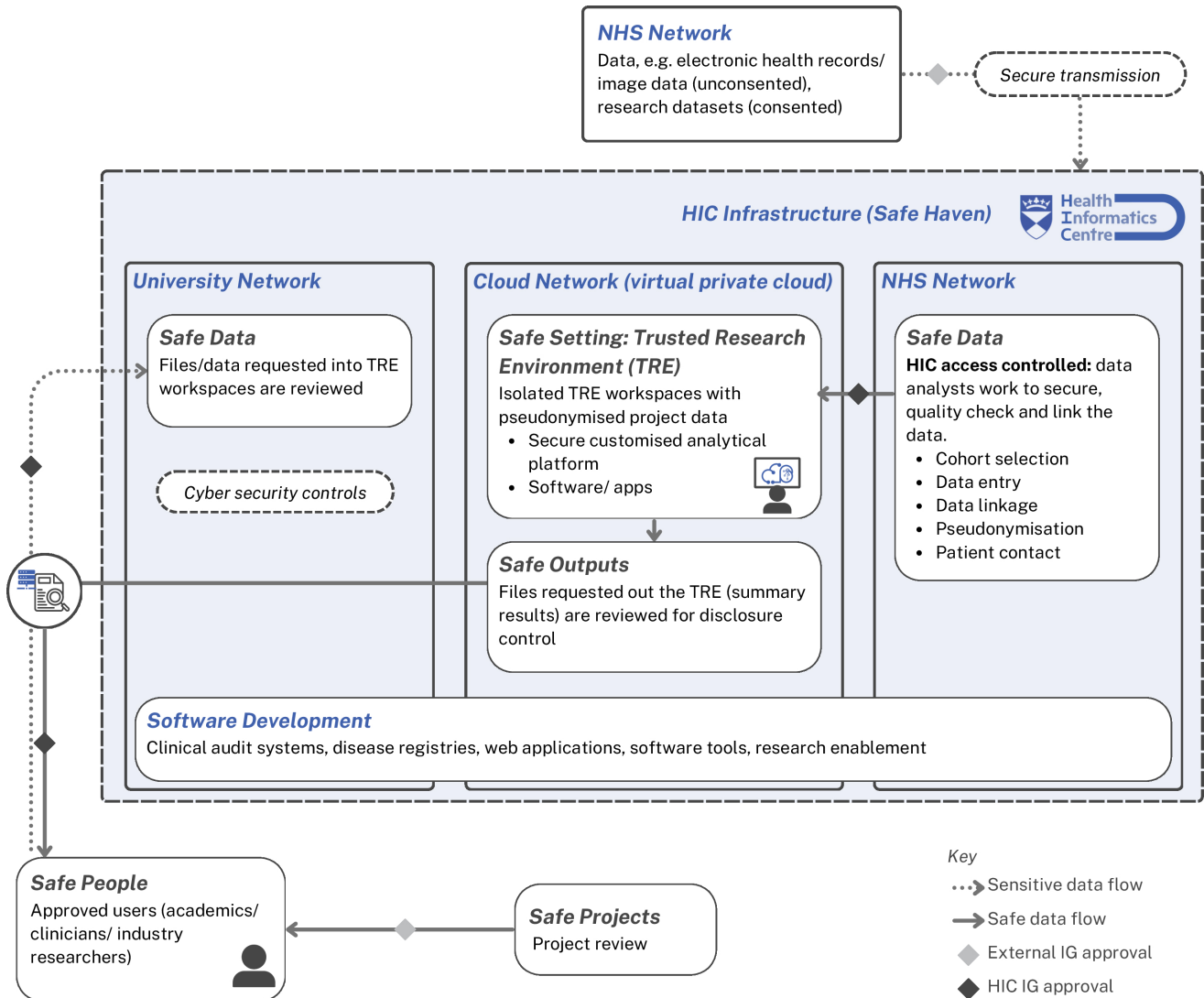
researchers interact with project data solely within the TRE and can request data summaries or plots as outputs which are manually reviewed to prevent disclosure of individual-level information using best practice methodologies [28]. For software code or scripts requested either in or out of the TRE, the same process is applied wherein our technical teams review files in an 'Information Governance' workspace within the TRE. This allows HIC to complete a risk and security assessment (covering the risk to data, governance, and environment), while reviewing licensing documentation and approvals required for release. This process applies to both analytical code and pre-trained AI/ML models. HIC's infrastructure and processes safeguard data security at every stage, enabling high quality research while upholding public trust and supporting researchers in their duty of care.

Trusted Research Environment (TRE)

The HIC TRE platform is the primary mechanism by which approved users can access pseudonymised data for research purposes. By being deployed on a secure public cloud infrastructure, the HIC TRE offers flexibility, scalability, and rapid access to advanced technologies. The architecture uses a 'secure-by-design' approach, and applies roles-based access controls, to maintain the integrity of the infrastructure which is managed exclusively by HIC. The cloud provider does not and cannot have access to any data provisioned within the TRE. Approved users are provisioned with project-specific, isolated and time-limited workspaces with access to their project data extract as per the project governance requirements. The flexibility of the TRE allows the provisioning of environments such as Windows or Linux operating systems, along with open-source software (i.e. R/RStudio) and proprietary software (licence dependent). To support reproducible research and version control, users have command-line access to Git within their workspace, enabling local code management without connecting to external repositories. We also offer researchers the use of software containerisation for tailored workspaces within projects.

The cloud-based TRE supports isolated high-performance computing, including batch job submissions via a Slurm

Figure 2: High level overview of HIC infrastructure and flow of sensitive data



Secure data management processes are labelled within the Five Safes, and different Information Governance (IG) responsibilities are identified as internal or external to HIC.

interface, significantly enhancing processing capabilities. This supports computationally intensive methodologies such as Artificial Intelligence/ Machine Learning (AI/ML), Natural Language Processing and 'omics' analyses. The TRE offers appropriate technical support for multi-omics research, enabling the integrated analysis of large-scale multimodal data, including genomics, phenomics, clinical imaging, and EHRs. Researchers can develop models on patient-level data within the TRE, with reviewed outputs supporting precision medicine [29]. Lastly, using a cloud solution comes with industry standard security and resilience, minimising downtime and reducing potential attack vectors.

Software as a service

HIC has adapted to meet the needs of clinical research which does not always revolve around data and has established a range of software services and solutions for researchers across the UK and internationally. These include electronic case report forms and data collection tools, project administration

software [30], web applications, reporting systems [31–33], clinical audit tools [34], disease registries [35, 36], programming interfaces and patient contact solutions such as text messaging [37, 38].

Data catalogue

A key strength of HIC lies in the deep granular regional electronic healthcare data, which can be linked to national datasets and generate transformative insights. HIC serves a population of approximately 1.1 million people across NHS Tayside, NHS Fife, and NHS Forth Valley, representing 20% of the Scottish population [39]. The available EHR data is sourced primarily from routine administrative hospital and medical systems, recording detailed information on patients, their care, medical facilities, and clinicians. As a Safe Haven, HIC provides secure access to these resources, subject to the appropriate ethics and IG approvals. HIC can provide access to over 170 datasets, including retrospective sources of health data spanning over three decades. The current commonly

used datasets are shown in Table 1. HIC provides access to Information Service Division (ISD) nationally collected data, specific to the regional health boards, which can be linked to other regional datasets. HIC supports the FAIR principles of Findability, Accessibility, Interoperability and Reusability [40] with Digital Object Identifiers (DOIs) published on the University of Dundee's Discovery platform [41] and are expected to be automatically populated on the Health Data Research (HDR) UK Gateway [42] in the future. Additionally, data has been mapped to the Observational Medical Outcomes Partnership (OMOP) common data model where feasible [43], facilitating standardisation across datasets. Many of these mapped datasets are from the Alleviate Pain Data Hub (see Data Hubs) shown in Table 2.

Data linkage

Administrative and health care data is taken from systems in raw format and HIC curate this with minimal processing, for example reformatting date or string variables. EHR data is collected through a multitude of different technical healthcare systems and stored via local and national mechanisms. The only way to make sense of these disparate datasets is to link them together through personal identifiable information such as the CHI number in Scotland. For data privacy reasons this must be carried out by a data linkage team outside of any research project and represents one of the core roles of HIC in any research project. Data minimisation and project-level pseudonymisation is applied, with personal data fully removed by default. Where appropriate IG approvals are in place, we can also support the acquisition and linkage of new datasets to support research projects. HIC's linkage and data processes are reproducibly managed by our open-source Research Data Management Platform described elsewhere [44] and more generally across the SSHN [19]. We maintain reproducible processes enabling 'refreshes' of data extracts into researcher SQL databases for consistency when required. HIC do offer data linkage to consented research cohorts as a 'bring your own' data model, which can be made available within the TRE.

Data hubs

Alleviate pain data hub

The UK-wide Alleviate Pain Data Hub is the data hub for the Advanced Pain Discovery Platform (APDP) [45], a large-scale initiative addressing the unmet need in chronic pain research. Alleviate includes genetic, medical, lifestyle and health information with a common focus on chronic pain, with over 387,000 records as detailed in Table 2. Hosted at HIC, Alleviate allows secure discovery and access to pain data through two models: metadata captured under the Alleviate collection on the HDR UK Gateway [46], or via data standardisation to the OMOP common data model [47]. Alleviate evolved from The *COVID – Curated and Open aNalysis aNd rEsearCh plaTform* (CO-CONNECT) project, which designed a federated data platform enabling researchers to discover research-ready datasets from across the UK via the HDR UK Cohort Discovery Tool [5]. Through the CO-CONNECT and Alleviate projects, we have established

an expertise of the OMOP common data model, having mapped over 10 million records from 30 datasets, including birth cohorts, national datasets, chronic pain and COVID-19 research datasets. We achieved this using the streamlined and privacy-preserving CaRROT Tools [48, 49]. This capability is being applied to the HDR UK Dementia Trials Platform and NHS England SDE network.

Clinical imaging

HIC hosts a substantial collection of routine clinical images, with access to approximately 200,000 adults from NHS Tayside adults from 2022 onwards, updated nightly. Earlier data from 2012 can be extracted from the national Picture Archiving and Communication System on a project-by-project basis. All clinical images are stored in DICOM format with a four-level hierarchy of person (individual undergoing clinical imaging), study (the reason for examination, which may entail multiple scans or modalities), series (information about each scan), and image (individual image during the scan). The image catalogue and hierarchical breakdown is shown in Table 3. This research-ready imaging data resource is provided with detailed metadata, supporting secure data linkage to other EHRs for enabling large-scale research. This capability was established through the *Interdisciplinary collaboration for efficient and effective use of clinical images in health care research* (PICTURES) project, a transformative 5-year programme led by HIC in collaboration with the University of Edinburgh, Abertay University, NHS Scotland, and industry collaborators [50]. The project produced the Scottish Medical Imaging (SMI) platform, which was developed to host, extract, and link clinical imaging data with other health datasets, unlocking large-scale research on two petabytes of data covering 57.3 million radiological studies. SMI is in production at Public Health Scotland and HIC, with the source code openly available [51]. This represents a world first in securely enabling population-scale research using routine, multi-modality clinical imaging data.

SHARE and Biorepository

The Scottish Health Research Register and Biobank (SHARE) is a significant and valuable resource for clinical trials and research with over 305,000 registered individuals, providing researchers with volunteers, biological samples, health data and genomic data [52]. SHARE is a growing research register with the unique addition of consented 'spare blood' collected during routine testing which is stored for future research including 'omics' studies or biomarker discovery. SHARE includes volunteers from across Scotland and can include general practice data as well as genomics. While SHARE operates under independent governance pathways separately from HIC, once approvals are in place, HIC enables study feasibility, identification, and creation of project cohorts to be contacted for recruitment, and completes any project which requires linkage and extraction, with data provision in the HIC TRE, on behalf of SHARE.

Relatedly, the NHS Research Scotland Biorepository [53] in Tayside provides consented tissue samples for use in approved medical research projects which can be linked to EHRs by HIC.

Table 1: Commonly used datasets available for access from the Health Informatics Centre

Sensitive data area	Dataset	Coverage	Time period	Approximate sample size	% F	DOI/Source
Cardiovascular	Echocardiogram (ECHO)	T, F	1994 – present	279,520	49.9%	https://doi.org/10.15132/10000214
Diabetes	Diabetes	T, F	2006 – present	106,000	47.3%	https://doi.org/10.15132/10000203
	Scottish Care Information - Diabetes (SCI-D)	T, F	1996 – present	112,500	47.3%	https://doi.org/10.15132/10000217
Emergency care	Accident & Emergency (A&E)	T, F	2004 – present	3,144,190	49.4%	https://doi.org/10.15132/10000209
	A&E Diagnosis	T	2003 – 2017	1,019,250	48.3%	https://doi.org/10.15132/10000222
	A&E Prescribing	T	2003 – 2017	166,380	45.7%	https://doi.org/10.15132/10000213
	National Scottish Ambulance Service (SAS)	T, F	2008 – 2021	22,300	44.3%	https://doi.org/10.15132/10000219
	Hypoglycaemic Events					
Endocrinology	Renal Register	T, F	2002 – present	2,800	42.6%	https://doi.org/10.15132/10000232
	SMR00 (Outpatient hospital admissions)	T, F	1997 – present	1.2 million	51.9%	https://publichealthscotland.scot/resources-and-tools/health-intelligence-and-data-management/national-data-catalogue/smr-data-manual/
	SMR01 (Acute stay hospital admissions)	T, F	1981 – present	1.1 million	50.8%	
	SMR02 (Maternity)	T, F	1975 – present	187,000	100.0%	
	SMR04 (Psychiatric hospital admissions)	T, F	1981 – present	41,000	50.1%	
	SMR06 (Cancer)	T, F	1980 – present	225,000	54.4%	
Laboratory results	Biochemistry	T, F	1992 – present	1.1 million	54.0%	https://doi.org/10.15132/10000215
	Haematology	T, F	1992 – present	981,000	54.1%	https://doi.org/10.15132/10000218
	Immunology	T, F	1993 – present	385,000	59.4%	https://doi.org/10.15132/10000230
	Microbiology	T, F	1998 – present	514,000	57.7%	T: https://doi.org/10.15132/10000223 F: https://doi.org/10.15132/10000228
				129,000		
	Pathology	T	1992 – present	330,000	57.1%	https://doi.org/10.15132/10000212
	Virology	T	1995 – 2013	341,000	57.8%	https://doi.org/10.15132/10000208
Pharmacy	Lighthouse	T, F	2019 – 2021	110,000	52.0%	https://doi.org/10.15132/10000211
	Community Dispensed Prescribing	T, F	1989 – present	1.4 million	52.7%	https://doi.org/10.15132/10000210
Population	Community Health Index (CHI)	N	1989 – present	Dynamic		
Health	Death certificate data	N	1988 – present	734,000	51.4%	https://doi.org/10.15132/10000226
	Death data	T, F	1980 – present	2.5 million	51.1%	https://doi.org/10.15132/10000220
	Demography	N	2009 – present	11.7 million	51.5%	https://doi.org/10.15132/10000221
	Scottish Population Health Research Register (SHARE) and BioBank	N	2013 – present	305,000	60.1%	https://doi.org/10.15132/10000216
Public Health	Bowel Screening	T	2000 – 2019	158,000	50.9%	https://doi.org/10.15132/10000205
	Scottish Index of Multiple Deprivation (SIMD)	N	2004 – 2016	Dynamic		https://doi.org/10.15132/10000237

This includes regional datasets covering NHS Tayside (T) or NHS Fife (F), or nationwide (N) across Scotland.

Table 2: Alleviate Pain Data Hub collection for the Advanced Pain Discovery Platform (ADPD) datasets

Pain Dataset	Time period	Approximate sample size	Mapped to OMOP
Generation Scotland: Scottish Family Health Study	2006 – 2011	24,000	Yes
Knee Pain In the Community (KPIC)	2014	9,000	No
Omega 3 Cohort	2021	75	Yes
Genetics of Osteoarthritis and Lifestyle (GOAL)	2017	3,200	Yes
Webex Cohort	2018	150	Yes
TwinsUK	1992 – 2021	15,370	Yes
GoDARTS (Diabetes)	2005 – 2008	16,800	No
1985 National Child Development Study	1985 - present	8,500	No
Oxford Carpal Tunnel Syndrome (CTS) Cohort	2021	150	Yes
BathTAP study cognition in pain	2015 – 2017	3,200	No
Decode-ME (Myalgic Encephalomyelitis)	2021	25,000	No
Neuropathic Pain Profiling in Retroviral Infection Study (NIPPR)	2021	40	No
HIV-related Peripheral Neuropathy (PINs)	2014	100	No
DOLORisk (Neuropathic pain, diabetes and genetics)	2015		
<i>DOLORisk University of Dundee</i>		9,160	Yes
<i>DOLORisk University of Oxford</i>		750	Yes
<i>DOLORisk Imperial College London</i>		190	Yes
Carpal Tunnel Syndrome (CAPS)	2021	70	No
HIV-POGO (Neuropathic Pain)	2021	150	No
Intellectual Disability & Pain	2012 – 2022	10,700	No
Scottish Population Health Research Register (SHARE) and BioBank	2013 - present	261,200	Yes
Birmingham Inflammation and joint pain study	2017 – 2024	70	Yes
FORECAST		280	No

A UK-wide initiative with growing number of datasets findable on the HDR UK Gateway (46). Data standardisation to the OMOP common data model which are visible, standardised, and accessible.

Table 3: NHS Tayside clinical imaging data held by the Health Informatics Centre (HIC) with approximate counts for individuals, images, series, and studies as of August 2025

Modality	Individuals (n)	Images (n)	Series (n)	Studies (n)	Females (%)
Computed Radiography	137,690	190,476,270	1,285,180	309,880	52.4%
Ultrasound	141,270	4,689,230	275,660	273,750	63.7%
Computed Tomography	137,700	190,484,150	1,285,140	309,900	52.4%
Magnet Resonance	76,180	49,376,220	1,381,940	128,090	54.5%
X-ray Angiography	34,330	812,550	187,500	52,900	45.9%
Mammography	30,370	293,880	190,040	62,870	97.7%
Other	20,920	1,424,420	34,740	28,480	63.5%
Presentation State	16,890	128,150	43,780	21,330	53.6%
Digital Radiography	156,850	572,590	471,680	385,390	52.2%
Radio Fluoroscopy	14,590	590,130	120,450	17,810	57.3%
Intra-oral Radiography	10,720	26,690	20,550	16,490	49.7%
Position Emission Tomography	5,430	3,709,480	14,880	6,980	48.6%
Secondary Capture	3,980	7,030	4,560	4,510	55.5%
External-camera Photography	3,230	7,290	4,030	4,030	58.5%

Noteworthy outputs

The research-led activities within HIC provide a unique perspective in health data research as both a service provider and research user. This symbiotic relationship between teams allows HIC to actively develop and incorporate research-driven insights into operational processes. This continuously improves

service delivery to meet evolving research requirements, and ultimately benefits the UK TRE community. This section showcases recent research advancements and the ways HIC has made a meaningful difference in health informatics research.

While disclosure control for “Safe Outputs” (Figure 2) requires checking all files, the introduction of AI/ML methods is a significant challenge as models cannot be

manually inspected [54]. The *Guidelines and Resources for Artificial Intelligence Model Access from Trusted Research Environments* (GRAIMatter) Green Paper provided a set of recommendations and mitigations to help all stakeholders securely support AI/ML model development and release [55, 56]. The real-world complexity of these evolving methodologies presents ongoing challenges for output checking processes, and work is continuing on semi-automated disclosure control of models from TREs through the SACRO project [57].

TREs across the UK have developed organically and independently, making it harder for researchers to transfer experience, and for data controllers to assess their suitability for processing sensitive data. The *Standardised Architecture for Trusted Research Environments* (SATRE) specification, first published in October 2023, defined a set of capability requirements to help standardise TREs, lays out a framework for rationalising what a TRE can, should, or must do [25]. SATRE was developed openly and collaboratively, with contributions from over 60 organisations and feedback gathered from 14 open Collaboration Cafés: regular, inclusive spaces for all contributors including the public. This has resulted in rapid and wide adoption across the UK. By establishing best practices, SATRE ensures TREs meet a set of capabilities and requirements to operate effectively within the Five Safes framework and potentially leading to formal accreditation such as ISO 27001. SATRE was identified as having a role in enabling four nations health data research capability across the UK in the 2024 Sudlow Review: *Uniting the UK's Health Data* [9]. The SATRE project is still active and is being used as a foundation for developing the blueprint for a European federation of TREs, EOSC-ENTRUST [58], and, through the TREvolution Data and Analytics Research Environments (DARE) UK Programme, across the UK to facilitate interoperability and federation.

Clinical trials are an important and heavily regulated area of research and HIC has two tools to support trials: the Tayside Randomisation System (TRuST) [30] and Pharmacovigilance Tracker (PV Tracker) [59]. TRuST has undergone multiple Medicines and Healthcare Regulatory Agency (MHRA) inspections and passed the regulatory and quality standards, with approved usage by the Tayside Clinical Trials Unit. Since 2012, TRuST has supported over 60 trials (e.g., 30–32), randomised over 18,000 participants and managed over 1,400 users and has passed two inspections. HIC released the PV Tracker in 2023 to remotely support the tracking and reporting of Significant Adverse Events, improving trial efficiency without a risk to patient safety in line with MHRA guidance. PV Tracker has been used for eight clinical trials so far and supported over 280 users. These HIC-developed applications offer professional solutions for conducting and auditing clinical trials, whether they focus on drug or medical technology safety.

Discussion

Since the establishment of HIC in 2004, healthcare research has undergone an electronic revolution. It has progressed from digitising paper records to implementing cloud computing and AI workflows. HIC has not only adapted to these innovations but has demonstrated leadership across Scotland and the

UK by developing research programmes that make EHR and patient data more readily available for research, while improving the secure environment in which this work occurs. We have established ourselves as a key partner for healthcare data projects and collaborative research programmes. HIC have an extensive catalogue of rich data resources available in different formats to suit research needs, a flexible and scalable secure computing environment, an active team of researchers, data analysts, software developers and governance professionals ready to engage with new projects. HIC has focused on process delivery to best support research TRE projects as effectively as possible, and promote best practice in secure data management. For regionally available datasets, we have an efficient data access process for approved projects within a month, and output requests are typically completed within one working day (for non-AI/ML files).

Safe Havens, TREs, SDEs, or other secure environments for working with sensitive data share a common challenge; balancing data privacy with enabling research access. Legal requirements such as (UK) GDPR, the Data Protection Act (2018), and the Scottish Safe Haven Charter (2025) place obligations on organisations regarding appropriate use of personal data. On the other side, organisations aim to maximise researcher access and infrastructure to deliver their research for public benefit. This is a challenging dichotomy in what is often a risk averse domain where the simplest option is to deny access. In HIC's experience, the solution and reality lies in building trusted relationships with stakeholders demonstrating professionalism, experience, and transparency. Once trust is established, decision making becomes much easier. A further common challenge is the navigation of somewhat complicated governance pathways. Over the last two years, HIC have implemented a refreshed IG culture. The team moved from a reactive, audit-driven periodic IG review, to a continuous organisation-wide IG culture which is inclusive, engaged, and effective. All processes and resources are fully digitised with clear responsibilities meaning that HIC are always audit-ready. This work is the focus of a future publication.

Through the Scottish Safe Haven Network (SSHN) sitting under the Chief Scientist Office, we have access to whole-population data in Scotland. Via Research Data Scotland and HDR UK, this extends to the wider UK data ecosystem. Data remains a core element of UK Industrial Strategy, with Scotland recognised as a key contributor. To support data access efficiency, the SSHN have established a federated governance framework together with the Caldicott Guardians to improve access to multi-regional data across Scotland. The SSHN pre-dates the implementation of (UK) GDPR and DEA accreditation and has, by design, created an ecosystem well adapted to work with sensitive data for the benefit of researchers, data controllers and the public. By being informed by best practice from other sectors and continuing to do so, HIC can meet all stakeholders' demands, such as increased scrutiny of the privacy risks of AI/ML, or concerns around data sovereignty.

In the early 2000s, when HIC was established, only a handful of TREs existed. Since then, the UK TRE landscape has changed dramatically with approximately 80 TREs now in operation and more in development. HIC as a Safe Haven has maintained a software development team since its inception,

providing a diversity of solutions and capabilities. More recently, the demands made of data providers have changed to include more complex datasets, advanced methodologies, and greater requirements for digital support and infrastructure. In 2023, we responded to these growing needs by migrating our TRE to a cloud-based TRE achieving our aim of enabling greater flexibility in the services we offer (e.g. large GPU compute, high-performance clusters, large data analysis), scalability, and resilience. While we have found that there is more effort in managing costs, it has offloaded the need to continually purchase new hardware when it reaches end of life, and has let us benefit from the commitments made by our cloud provider to reducing their carbon footprint. Many other mature TRE providers in the UK are operating solely on-premises infrastructure, limiting the range of services they can offer, and either restricting the number of projects they can enable, or hosting expensive and unused hardware. Longer term, HIC will be investigating a hybrid approach, where a mix of on-premise and flexible cloud infrastructure allows us to benefit from the advantages of both models. Other TRE providers are now starting to investigate cloud, and many new entrants (i.e. NHS England SDEs) are cloud first. Our experience in sustainability has helped us focus building capacity that is both useful and affordable within the UK landscape. In the coming years, this is likely to be an issue where rapidly built solutions may need to be rationalised to match costs and resource capacity with long term funding constraints.

HIC actively collaborates with new and existing organisations to share our outputs, tools, and experience. Our commitment to open science and reproducibility is reflected in the release of open-source tools [25, 44, 50, 60] and code on our GitHub repository [27], and our leadership of community initiatives, including the UK TRE Community and the Scottish Safe Haven Charter [21]. The HIC model is adaptable to work in most national and international data environments, and as such we have engaged at a high level in several countries. However, transitions at national level take time to be adopted. It is our opinion that continued coordination and shared knowledge across the UK TRE community will enhance collective capability making the UK a safe setting for sensitive data research.

Challenges remain as data demands rapidly grow; to build ever more informative research outputs, greater access to data and computational power is required. The UK government's funding strategy to build AI capacity through AI Research Resource (AIRR), and improving data access via the Health Data Research Service, addresses some aspects but not the disparate and disjointed IG landscape across the UK. For example, access to primary care data remains limited to the few projects that, through fortunate timing or local approaches, were able to secure the governance. Uneven funding across the four nations that prioritises new, potentially duplicative capacity over adequately supporting existing infrastructure, is also slowing the UK health data ecosystem from developing as quickly, efficiently and sustainably as possible. DARE UK funded programmes, are a valuable initiative to bring together fundamental capacity working on sensitive data into common and reusable components. HIC's involvement in the current TREvolution programme will help new and existing organisations manage costs,

improve sustainability, avoid duplication and provide federation capacity nationally and internationally.

Looking forward, there are many future improvements; better integration across the four nations for inclusive UK-wide research, better training and user experience, and responsible access to valuable datasets such as primary care data. The very rapid growth of the UK TRE ecosystem following a significant investment phase, needs to transition to sustainable models, which HIC and other mature TREs have shown is possible.

Data access

Researchers can discuss feasibility and request access to data held by HIC by contacting HICSupport@dundee.ac.uk. Data cannot be shared openly due to being special category data and must remain secured under HIC's standard operating procedures and ISO 27001 accreditation. As part of the University of Dundee, HIC have a webpage and outputs (data, code, publications) published on the University of Dundee Discovery webpage [41] and a live Project Registry updated every 8 hours [26]. Additionally, we continue to work with HDR UK to populate the Gateway [42] and researchers can explore the Alleviate Hub collection using the cohort discovery tool [46].

Conclusion

We live in a world overflowing with data generated from diverse sources, and extracting meaningful insights is crucial. Data drives research and innovation, informs decision making, facilitates progress and ultimately saves lives. The value of this UK data ecosystem was reinforced during the Covid-19 global pandemic, through the rapid and coordinated response of HIC and our UK TRE data community. HIC will continue to lead advancements with the Scottish Safe Haven Network and UK TRE community in the use of sensitive data, focusing on sustainability and the development of secure, flexible, scalable and efficient infrastructure and methods, supported by strategic investment. Our agility and commitment to continual improvement will be key to advancing data re-use and open science, ensuring long term impact in research and innovation.

Statements

Acknowledgements

HIC would like to thank the hundreds of researchers that we have had the pleasure to work and collaborate with over the last 20 years and acknowledge the University of Dundee's School of Medicine for supporting us to grow into a centre of excellence and innovation.

We would also like to thank NHS Tayside and NHS Fife for their continued support and engagement in enabling research on patient data.

CRediT: Conceptualisation: LW, CC; Data curation: CH, MG; Funding acquisition: RW, CC; Methodology: JJ, KM, CC; Project administration: JJ, SK, GM, JH; Resources: KM, CH, MG; Software: KM, CJ; Supervision: CH, MG, CC; Visualisation: LW; Writing – original draft: LW, JJ, CH, CJ,

SK, GM, RW, CC; Writing – review & editing: LW, SK, JH, RW, CC.

Ethics statement

This manuscript is not a study and does not require ethical approval.

Conflict of interest statement

The authors declare that they have no personal, commercial, political, academic or financial conflicts of interest.

Publication consent

We have consent to publish and openly share the data within this manuscript.

Funding statement

HIC is self-funded via successful grant applications, services and a limited amount of direct Scottish Government funding to enable our activities in supporting research projects. The HIC development activities are constantly evolving to ensure sustainability without an ability to utilise core underpinning support for operations and development from the institution or Scottish Government. Research at HIC has been funded by the UKRI, the Chief Scientist Office of Scotland, Research Data Scotland, DARE UK, HDR UK, EU Horizon 2020, industry collaborations with Astra Zeneca, Roche, Lilly, Novo Nordisk and many charities. The Alleviate Pain Data Hub was funded under MRC SPF grant MR/W014335/1 and the DARE UK Projects were funded under grants (MC_PC_21033, MC_PC_23008, MC_PC_23006 and MC_PC_21032).

References

- Franklin JM, Glynn RJ, Martin D, Schneeweiss S. Evaluating the Use of Nonrandomized Real-World Data Analyses for Regulatory Decision Making. *Clin Pharma and Therapeutics*. 2019 Apr;105(4):867–77. <https://doi.org/10.1002/cpt.1351>
- Oo MM, Gao C, Cole C, Hummel Y, Guignard-Duff M, Jefferson E, et al. Artificial intelligence-assisted automated heart failure detection and classification from electronic health records. *ESC Heart Failure*. 2024 Oct;11(5):2769–77. <https://doi.org/10.1002/ehf2.14828>
- Weir A, Bishop J, McGurnaghan SJ, McAllister D, Robertson C, Wood R, et al. Relation of severe COVID-19 to polypharmacy and prescribing of psychotropic drugs: the REACT-SCOT case-control study. *BMC Med*. 2021 Dec;19(1):51. <https://doi.org/10.1186/s12916-021-01907-8>
- Vasileiou E, Simpson CR, Shi T, Kerr S, Agrawal U, Akbari A, et al. Interim findings from first-dose mass COVID-19 vaccination roll-out and COVID-19 hospital admissions in Scotland: a national prospective cohort study. *The Lancet*. 2021 May;397(10285):1646–57. [https://doi.org/10.1016/s0140-6736\(21\)00677-2](https://doi.org/10.1016/s0140-6736(21)00677-2)
- Jefferson E, Cole C, Mumtaz S, Cox S, Giles TC, Adejumo S, et al. A Hybrid Architecture (CO-CONNECT) to Facilitate Rapid Discovery and Access to Data Across the United Kingdom in Response to the COVID-19 Pandemic: Development Study. *J Med Internet Res*. 2022 Dec 27;24(12):e40035. <https://doi.org/10.2196/40035>
- Orton C, Ford D, Sheikh A, Quint J, Tobin M, Hall I, et al. BREATHE: The Health Data Research Hub for Respiratory Health. *IJPDS*. 2022 Aug 25;7(3). <https://doi.org/10.23889/ijpds.v7i3.1994>
- Department of Health and Social Care. Data saves lives: reshaping health and social care with data [Internet]. London: UK Government; 2022 Jun. Available from: <https://www.gov.uk/government/publications/data-saves-lives-reshaping-health-and-social-care-with-data/data-saves-lives-reshaping-health-and-social-care-with-data>
- DARE UK. Paving the way for a coordinated national infrastructure for sensitive data research [Internet]. 2022 Aug [cited 2025 Aug 20]. Available from: <https://doi.org/10.5281/zenodo.7022439>
- Sudlow C. Uniting the UK's Health Data: A Huge Opportunity for Society [Internet]. Zenodo; 2024 Nov [cited 2025 Aug 20]. Available from: <https://doi.org/10.5281/zenodo.13353746>
- Scottish Government. Health and social care: data strategy [Internet]. Scottish Government; 2023 [cited 2025 Aug 20]. Available from: <https://www.gov.scot/publications/data-strategy-health-social-care-2/>
- Guo Y, Raventós B, Català M, Elhussein L, López-Güell K, Tan EH, et al. Time Series Methods to Assess the Impact of Regulatory Action: A Study of UK Primary Care and Hospital Data on the Use of Fluoroquinolones. *Pharmacoepidemiology and Drug*. 2024 Oct;33(10):e70022. <https://doi.org/10.1002/pds.70022>
- Agniel D, Kohane I, Weber G. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. *BMJ*. 2018 Oct 18;k4416. <https://doi.org/10.1136/bmj.k1479>
- Rassen JA, Murk W, Schneeweiss S. Real-world evidence of bariatric surgery and cardiovascular benefits using electronic health records data: A lesson in bias. *Diabetes Obesity Metabolism*. 2021 Jul;23(7):1453–62. <https://doi.org/10.1111/dom.14338>
- Rydzewska E, Nijhof D, Hughes L, Melville C, Fleming M, Mackay D, et al. Rates, causes and predictors of all-cause and avoidable mortality in 514 878 adults with and without intellectual disabilities in Scotland: a record linkage national cohort study. *BMJ Open*.

- 2025 Feb;15(2):e089962. <https://doi.org/10.1136/bmjopen-2024-089962>
15. Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: A call for open science. *Patterns*. 2021 Oct;2(10):100347. <https://doi.org/10.1016/j.patter.2021.100347>
16. Harron K, Dibben C, Boyd J, Hjern A, Azimae M, Barreto ML, et al. Challenges in administrative data linkage for research. *Big Data & Society*. 2017 Dec;4(2):205395171774567. <https://doi.org/10.1177/2053951717745678>
17. A charter for safe havens in Scotland: handling unconsented data from National Health Service patient records to support research and statistics. Edinburgh: Scottish Government; 2015.
18. Ritchie F. Secure access to confidential microdata: four years of the Virtual Microdata Laboratory. *Econ Lab Market Rev*. 2008 May;2(5):29–34. <https://doi.org/10.1057/elmr.2008.73>
19. Gao C, McGilchrist M, Mumtaz S, Hall C, Anderson LA, Zurowski J, et al. A National Network of Safe Havens: Scottish Perspective. *J Med Internet Res*. 2022 Mar 9;24(3):e31684. <https://doi.org/10.2196/31684>
20. O'Sullivan K, Wilde K. A profile of the Grampian Data Safe Haven, a regional Scottish safe haven for health and population data research. *IJPDS [Internet]*. 2023 Mar 16 [cited 2025 Aug 20];4(2). Available from: <https://doi.org/10.23889/ijpds.v4i2.1817>
21. Scottish Government. Charter for Safe Havens in Scotland. Principles and Standards for the Scottish Safe Haven Network to Support the Use of Data to Enable Research and Innovation in Scotland. 2025.
22. Evans JMM, McDevitt DG, McDonald TM. The Tayside Medicines Monitoring Unit (MEMO): A record-linkage system for pharmacovigilance. *Pharmaceutical Medicine*. 1995 Dec 1;9(3):177–84. <https://doi.org/10.1046/j.1365-2125.1999.00853.x>
23. Evans JMM, MacDonald TM. Record-linkage for pharmacovigilance in Scotland. *Brit J Clinical Pharma*. 1999 Jan;47(1):105–10. <https://doi.org/10.1046/j.1365-2125.1999.00853.x>
24. DARE UK. DARE UK Phase 1 Driver Projects Public Involvement and Engagement Report [Internet]. Zenodo; 2024 Jun [cited 2025 Aug 20]. Available from: <https://doi.org/10.5281/zenodo.11508647>
25. Cole C, Machin T, Chuter A, Oldfield K, Beggs J. Standardised Architecture for Trusted Research Environments (SATRE) [Internet]. Available from: <https://satre-specification.readthedocs.io/>
26. Health Informatics Centre. Health Informatics Centre Trusted Research Environment Project Registry [Internet]. 2025. Available from: <https://hicservices.atlassian.net/wiki/spaces/HKB/pages/335544328/Project+Registry>
27. Health Informatics Centre Services GitHub [Internet]. Health Informatics Centre: University of Dundee; 2025. Available from: <https://github.com/HicServices>
28. Griffiths E; Greci C; Kotrotsios Y; Parker S; Scott J; Welpton R; Wolters A; Woods C; Safe Data Access Professionals Working Groups. Handbook on Statistical Disclosure Control for Outputs. 2nd ed. Colchester: UK Data Archive; 2024.
29. Syed MG, Trucco E, Mookiah MRK, Lang CC, McCrimmon RJ, Palmer CNA, et al. Deep-learning prediction of cardiovascular outcomes from routine retinal images in individuals with type 2 diabetes. *Cardiovasc Diabetol*. 2025 Jan 2;24(1):3. <https://doi.org/10.5281/zenodo.11508647>
30. Jones C, Jefferson E, Hogarth F, Littleford R, Band M. Supporting clinical trials through healthcare informatics. *Trials*. 2015 Dec;16(S2):O67. <https://doi.org/10.1186/1745-6215-16-S2-O67>
31. Robles-Zurita JA, McMeekin N, Sullivan F, Mair FS, Briggs A. Health Economic Evaluation of Lung Cancer Screening Using a Diagnostic Blood Test: The Early Detection of Cancer of the Lung Scotland (ECLS). *Current Oncology*. 2024 Jun 18;31(6):3546–62. <https://doi.org/10.3390/curroncol31060261>
32. Sullivan FM, Farmer E, Mair FS, Treweek S, Kendrick D, Jackson C, et al. Detection in blood of autoantibodies to tumour antigens as a case-finding method in lung cancer using the EarlyCDT®-Lung Test (ECLS): study protocol for a randomized controlled trial. *BMC Cancer*. 2017 Dec;17(1):187. <https://doi.org/10.1186/s12885-017-3175-y>
33. Sullivan FM, Mair FS, Anderson W, Armory P, Briggs A, Chew C, et al. Earlier diagnosis of lung cancer in a randomised trial of an autoantibody blood test followed by imaging. *Eur Respir J*. 2020 Jul 30;2000670. <https://doi.org/10.1183/13993003.00670-2020>
34. Witham MD, Band MM, Littleford RC, Avenell A, Soiza RL, McMurdo MET, et al. Does oral sodium bicarbonate therapy improve function and quality of life in older patients with chronic kidney disease and low-grade acidosis (the BiCARB trial)? Study protocol for a randomized controlled trial. *Trials*. 2015;16(1):326. <https://doi.org/10.1186/s13063-015-0843-6>
35. Cerebral Palsy Integrated Pathway [Internet]. 2025. Available from: <https://www.cpipuk.org/>.
36. EMBARC The European Bronchioectasis Registry [Internet]. 2025. Available from: <https://bronchiectasis.hicservices.dundee.ac.uk/>.

37. Dombrowski SU, McDonald M, Van Der Pol M, Grindle M, Avenell A, Carroll P, et al. Game of Stones: feasibility randomised controlled trial of how to engage men with obesity in text message and incentive interventions for weight loss. *BMJ Open*. 2020 Feb;10(2):e032653. <https://doi.org/10.1136/bmjopen-2019-032653>
38. Innes N, Fairhurst C, Whiteside K, Ainsworth H, Sykes D, El Yousfi S, et al. Behaviour change intervention for toothbrushing (lesson and text messages) to prevent dental caries in secondary school pupils: The BRIGHT randomized control trial. *Comm Dent Oral Epid*. 2024 Aug;52(4):469–78. <https://doi.org/10.1111/cdoe.12940>
39. National Records of Scotland. Mid-2024 population estimates [Internet]. 2025. Available from: <https://www.nrscotland.gov.uk/publications/mid-2024-population-estimates/>.
40. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016 Mar 15;3(1):160018. <https://doi.org/10.1038/sdata.2016.18>
41. Health Informatics Centre [Internet]. Available from: <https://discovery.dundee.ac.uk/en/equipments/health-informatics-centre>.
42. HDR UK. HDR UK Gateway. Health Data Research UK Gateway Health Informatics Centre Collection. Available from: <https://healthdatagateway.org/en/collection/20>.
43. Milligan G, Masood E, Quinlan P, Cox S, Giles T, Mendez Villalon A, et al. Mapping UK Pain Datasets to the OMOP Common Data Model: Lessons Learned. *IJPDS*. 2024 Sep 10;9(5). <https://doi.org/10.23889/ijpds.v9i5.2613>
44. Nind T, Galloway J, McAllister G, Scobbie D, Bonney W, Hall C, et al. The research data management platform (RDMP): A novel, process driven, open-source tool for the management of longitudinal cohorts of clinical data. *GigaScience*. 2018 Jul 1;7(7):giy060. <https://doi.org/10.1093/gigascience/giy060>
45. Advanced Pain Discovery Platform [Internet]. Available from: <https://apdp.community/>.
46. HDR UK Gateway [Internet]. Health Data Research UK Alleviate Pain Data Hub. Available from: <https://healthdatagateway.org/en/collection/75>.
47. OHDSI: Observational Health Data Sciences and Informatics. Standardized Data: The OMOP Common Data Model [Internet]. Available from: <https://www.ohdsi.org/data-standardization/the-common-data-model/>.
48. Cox S, Masood E, Panagi V, Macdonald C, Milligan G, Horban S, et al. Conversion of Sensitive Data to the Observational Medical Outcomes Partnership Common Data Model: Protocol for the Development and Use of Carrot. *JMIR Res Protoc*. 2025 Apr 2;14:e60917. <https://doi.org/10.2196/60917>
49. Macdonald C, Panagi V, Appleby P, Simon, David, Santos R, et al. HDRUK/CarROT-CDM: Chunking correction [Internet]. Zenodo; 2022 [cited 2025 Aug 20]. Available from: <https://doi.org/10.5281/zenodo.5529833>.
50. PICTURES: Supporting the use of data for health care research - Image on a Mission. [Internet]. Available from: <https://imageonamission.ac.uk>.
51. Health Informatics Centre. Software for Medical Imaging [Internet]. Available from: <https://smi.readthedocs.io/en/latest/>.
52. McKinsty B, Sullivan FM, Vasishta S, Armstrong R, Hanley J, Haughney J, et al. Cohort profile: the Scottish Research register SHARE. A register of people interested in research participation linked to NHS data sets. *BMJ Open*. 2017 Feb;7(2):e013351. <https://doi.org/10.1136/bmjopen-2016-013351>
53. NRS. NHS Research Scotland Biorepository [Internet]. Available from: <https://nrsbiorepository.hicservices.dundee.ac.uk/>.
54. Ritchie F, Tilbrook A, Cole C, Jefferson E, Krueger S, Mansouri-Bensassi E, et al. Machine learning models in trusted research environments – understanding operational risks. *IJPDS*. 2023 Dec 14;8(1). <https://doi.org/10.23889/ijpds.v8i1.2165>
55. Jefferson E, Cole C, Boixader AC, Rogers S, Malone M, Ritchie F, et al. GRAIMatter: Guidelines and Resources for AI Model Access from Trusted Research environments (GRAIMatter). *IJPDS* [Internet]. 2022 Aug 25 [cited 2025 Aug 20];7(3). Available from: <https://doi.org/10.23889/ijpds.v7i3.2005>.
56. Jefferson E, Liley J, Malone M, Reel S, Crespi-Boixader A, Kerasidou X, et al. GRAIMATTER Green Paper: Recommendations for disclosure control of trained Machine Learning (ML) models from Trusted Research Environments (TREs). Zenodo; 2022 Sep. <https://doi.org/10.5281/zenodo.7089490>
57. Smith J, Albashir M, Seb Bacon, Ben Butler Cole, Jackie Caldwell, Christian Cole, et al. SACRO: Semi-Automated Checking of Research Outputs. *DARE UK*; 2023 Nov p. 1–12.
58. European Open Science Cloud - A European Network of Trusted Research Environments [Internet]. Available from: <https://eosc-entrust.eu/>.
59. Health Informatics Centre. Transforming SAE reporting in clinical trials with Pharmacovigilance Tracker (PV Tracker) [Internet]. University of Dundee; 2016. Available from: <https://www.dundee.ac.uk/projects/transforming-sae-reporting-clinical-trials-pharmacovigilance-tracker-pv-tracker>.

60. Li S, Taylor S, Shah K, Gao C, Sutherland J, Jefferson E, et al. TREEHOOSE v1.0.0-beta1 [Internet]. Zenodo; 2022 [cited 2025 Nov 20]. Available from: <https://doi.org/10.5281/zenodo.6908252>

Abbreviations

ADPD:	Advanced Pain Discovery Platform
AI/ML:	Artificial Intelligence/ Machine Learning
AIRR:	AI Research Resource
CHI:	Community Health Index
DOI:	Digital Object Identifier
EHR:	Electronic Health Record
FAIR:	Findable, Accessible, Interoperable, Reusable
GDPR:	General Data Protection Regulation
HDR:	Health Data Research UK
HIC:	Health Informatics Centre
IG:	Information Governance
ISD:	Information Service Division
MHRA:	Medicines and Healthcare Regulatory Agency
OMOP:	Observational Medical Outcomes Partnership
PV:	Pharmacovigilance Tracker
SDE:	Secure Data Environment
SHARE:	Scottish Health Research Register and Biobank
SSHN:	Scottish Safe Haven Network
TRE:	Trusted Research Environment
TRuST:	Tayside Randomisation System

