Population health analytics

from a data science perspective

CGI

Contents

This eBook provides an overview of the techniques used by data scientists when conducting population health analytics.

Introduction

Health and care needs have changed dramatically since most modern healthcare systems were conceived. In 2024, there were 2.6 million more attendances at type 1 A&E departments compared to 2010. This represents an increase of 19% demand when workforce and estates were already at breaking point.

People lead very different lifestyles, live considerably longer and increasingly with multiple long-term conditions such as diabetes, asthma, and dementia.

At the same time, populations have soared and become much more diverse whilst health inequalities within them have widened. This contributes to an increase in the volume and complexity of demand that our hospitals and other health and care providers are faced with, at a time when budgets across the NHS are being ever tightened.

Fundamentally, the way in which our health systems have been designed – to treat the patient once they become ill – is no longer fit for purpose or sustainable.

Addressing future health challenges starts with exploring:

- How do we prevent people from becoming ill in the first place?
- How do we identify individuals and groups at most risk of becoming ill, or at risk of being unable or unlikely to access services?
- How do we design targeted interventions that proactively support people to manage their health and care in the community and therefore protect secondary care services for those that need them most?
- How do we leverage real-time data from new technologies to make better decisions?

Population health management

This is the process of understanding the circumstances and health needs across a defined local, regional or national population with the aim of developing tailored models of care and targeted interventions that improve physical and mental health outcomes, reduce health inequalities, and critically make the best use of available resources by targeting prevention instead of treatment.

The ability to generate timely, equitable and actionable insights from data is essential but rigour is key – it's vital to understand how the data was generated, what the data represents and how to communicate the insights effectively.

Population health analytics produces the data-driven insights which inform population health management interventions.

It involves the analysis of a range of social, economic, and environmental data known as wider determinants of health, in addition to health and demographic data, to create a holistic view of a population's health and care needs. Population health analytics offers crucial insights into the overall health of a population, how we target specific groups and subsequently informs health and social care decision-making and funding.

By examining health-related data from various sources, population health analytics can:

- Detect patterns and trends in disease incidence and prevalence.
- Evaluate the efficacy of public health interventions, including patient care pathways.
- Enhance health outcomes by pinpointing and addressing health disparities and inequalities.
- Reduce the higher costs of treatment by targeting and supporting citizens with prevention measures.
- Identify groups at higher risk of adverse health outcomes and allow for specific interventions to be devised that target their needs.
- Detect areas of over-utilisation or under-utilisation to refine healthcare delivery and resource allocation, enabling the development of strategies to optimise healthcare delivery.

Population health analytics is essential, as it offers a data-driven approach to comprehending and improving the health of a population, ultimately informing policies and interventions that foster health and wellbeing for all.

Examples of population health analytics in action

Examples of population health analytics in action

Early detection of influenza outbreaks

The UK Health Security Agency (UKHSA) uses population health analytics to monitor and predict the spread of influenza. By analysing data from various sources, including general practitioners' reports, hospital admissions and social media, UKHSA can detect early signs of an outbreak and implement targeted interventions, such as promoting vaccination campaigns and increasing the availability of antiviral medications.

This proactive approach has helped reduce the impact of influenza on the UK population, minimising hospitalisations and saving lives.

Childhood obesity prevention

The National Child Measurement Programme (NCMP) in the UK collects data on the height and weight of children aged 4-5 and 10-11 years to monitor trends in childhood obesity. Population health analytics has been employed to identify areas with higher rates of childhood obesity and target interventions to promote healthy eating and physical activity in these communities.

As a result, local authorities and healthcare providers can develop tailored strategies to address childhood obesity and improve overall public health.

Case study: Protecting the vulnerable

Challenge

At the height of the COVID-19 pandemic, NHS England, supported by CGI and Oxford University, identified **1.7 million Clinically Extremely Vulnerable individuals** to be added to the national shielded patient list and prioritised for vaccination. The NHS England and CGI blended team made this possible by developing the COVID-19 Clinical Risk Assessment and Risk Stratification tools, which enabled millions of rows of demographic data to be fed through Oxford University's QCovid® risk-prediction algorithm.

Research commissioned by England's Chief Medical Officer uncovered several risk factors that could increase the likelihood of death from COVID-19.

Using this insight, the University of Oxford developed QCovid® – a COVID-19 risk-prediction model. With a cutting-edge risk-prediction model in place, NHS England and CGI sought to create a data solution capable of processing millions of rows of NHS patient data* through QCovid®.

Solution

CGI and NHS England were then tasked with running England's entire adult population (aged 19-100) through QCovid®. The team processed **46 million lines of patient data** using NHS England's Databricks platform, data was filtered down using a minimisation process, removing people with no recorded high-risk factors.

Approximately **15.5 million people were identified** to be processed through QCovid®, culminating in an additional **1.7 million individuals being added to the shielding list**. These individuals were yet to be selected for vaccination and were therefore urgently prioritised by their local GPs.

*CGI did not store or access any NHS patient data throughout this delivery

Vulnerable individuals prioritised for shielding and vaccine

1.7M

Delivered in just

7 months

Recognised with

Florence Nightingale Award for Excellence in Healthcare Data Analytics

The project referenced in this case study was delivered by BJSS, which CGI acquired in February 2025.

Outcome

CGI and NHS England developed a secure online web viewer to allow GPs to view details of the 15.5 million patients who were processed through QCovid®. This data set contained patients who had some risk factors but were not considered to be Clinically Extremely Vulnerable. These patients were therefore not added to the Shielded Patients List.

This secure online portal solution was **delivered in just four weeks**, utilising NHS Mail for multi-factor authentication, a DynamoDB authorisation solution, NHS England Databricks, AWS S3 buckets, and Nextjs server-side pages.

> "I wanted to say a heartfelt thank you for the incredible hard work that has been done by the whole team. Every single person has remained positive, focused and good-humoured, and together we have achieved great things."

Jonathan Benger, Interim Chief Medical Officer, NHS England

The project referenced in this case study was delivered by BJSS, which CGI acquired in February 2025.

Exploratory data analysis

Exploratory data analysis (EDA)

This is a critical component of population health analytics, as it allows researchers to explore and understand the patterns and trends in health data. In this chapter, we will discuss some of the key techniques associated with EDA.

Data cleaning

The process of identifying and correcting errors and inconsistencies in the data. This may involve removing outliers, correcting data entry errors, or identifying and addressing data inconsistencies across different variables. Data cleaning techniques are important in population health analytics to ensure that data is accurate and reliable for analysis.

Missing value handling

Missing values are a common issue in health data and can occur due to a variety of reasons, such as data entry errors or missing responses in surveys. Missing values can affect the accuracy and validity of data analysis, and it is therefore important to address missing values before the analysis stage. Missing value imputation techniques, such as mean imputation and multiple imputation, are commonly used in population health analytics to replace missing values with estimated ones.

'Missingness' can be due to different reasons and may in itself be informative, hence it is important to establish underlying data generating mechanisms to determine the appropriate method for handling missing values. The volume of missingness is also a key consideration when deciding how to handle missing values and is a key metric to report during EDA.

Transformation and normalisation

Processes that involve converting raw data into a format suitable for analysis. This may include converting data types, scaling data to a common range, and converting categorical data to numerical data. These techniques are important in population health analytics to ensure that data is in a format that can be easily analysed and compared across different variables.

Descriptive statistics

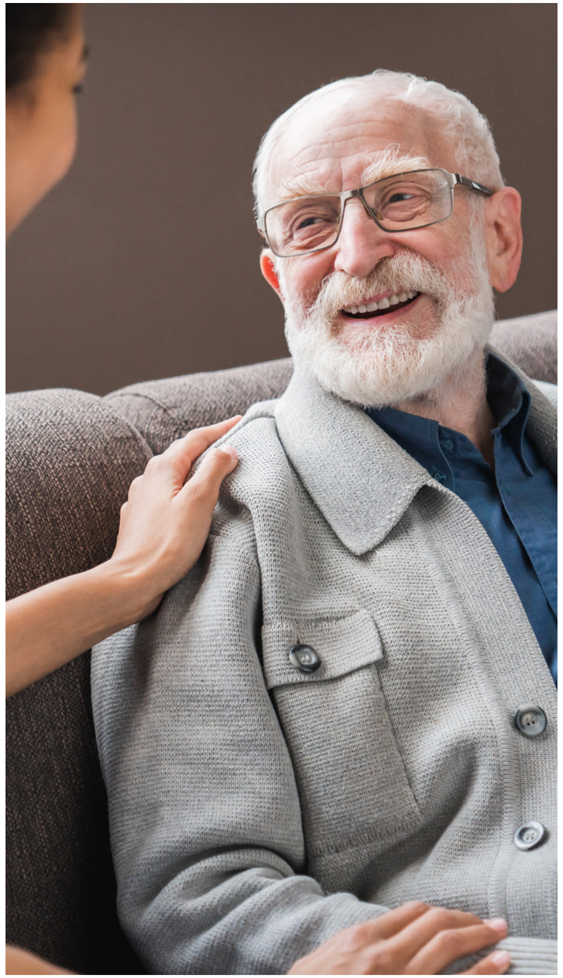
Provides a summary of the data using measures such as mean, median, mode, range, and standard deviation to describe the central tendency, dispersion, and distribution of the data. These techniques help researchers to understand the main features of a dataset, identify patterns and trends, and communicate complex health data to a wide range of stakeholders.

Data visualisation

Techniques essential for communicating the results of population health analytics in a clear and accessible way. Visualisations such as bar charts, line charts, scatter plots, and maps can help to convey patterns and trends in the data, while interactive visualisations can enable stakeholders to explore the data in more depth.

Inferential statistics

Enables researchers to draw conclusions about a population based on a sample of data. Techniques such as hypothesis testing, confidence intervals, and regression analysis are used to make inferences about population parameters and to test relationships between variables.



Important considerations

Data quality

This is a major concern in population health analytics, as inaccurate or incomplete data can lead to incorrect conclusions and flawed policy decisions. Ensuring data quality requires careful attention to data collection, storage, and cleaning processes. In the UK, organisations like NHS Digital are responsible for managing health data and ensuring data quality standards are met.

Data sharing and collaboration

Important components of population health analytics that enable researchers and policymakers to access and analyse large and complex datasets. In the UK, initiatives like the UK Biobank and the UK Data Service are working to promote data sharing and collaboration in the field of population health analytics.

Machine learning and predictive modelling

Machine learning (ML) and predictive modelling

These are powerful tools in population health analytics, as they allow researchers to build models that can predict health outcomes and identify risk factors for disease. In this chapter, we will discuss some of the key techniques associated with machine learning and predictive modelling.

Regression and classification

Used to create predictive models, classify individuals into risk groups, and identify patterns in the data. Common machine learning algorithms used in population health analytics include decision trees, random forests, logistic regression, support vector machines, and neural networks.

Clustering

Used to group similar data points based on shared characteristics. This can help to identify subpopulations within the larger population that may have distinct health needs or risk factors. Common clustering algorithms include K-means, hierarchical clustering, and DBSCAN.

Time series analysis

Used to analyse data collected over time, such as disease incidence rates or healthcare utilisation patterns. Techniques such as autoregressive integrated moving average (ARIMA) models, exponential smoothing, and seasonal decomposition can help to identify trends, seasonal patterns, and forecast future values.

Natural language processing (NLP)

Used to analyse unstructured textual data, such as electronic health records or social media posts, to gain insights into population health. Common NLP techniques include sentiment analysis, topic modelling, and named entity recognition.

Network analysis

Used to study relationships between individuals, organisations, or other entities. This can help identify social networks or patterns of communication that may influence health behaviours or outcomes. Techniques used include graph theory, centrality measures, and community detection algorithms.

Geospatial analysis

Used to study the geographic distribution of health outcomes and risk factors, as well as the impact of environmental factors on population health. Techniques such as spatial autocorrelation, hot spot analysis, and spatial regression can help to identify areas of high or low health risk and inform targeted interventions.

Causal modelling

Used to determine and quantify the underlying cause-effect relationship between risk factors and health outcomes, often to determine whether an intervention is effective. Techniques such as structural equation modelling (SEM), directed acyclic graphs (DAGs) and counterfactual-based methods are commonly used.

Image analysis

Deep learning models, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), can automatically extract features and detect patterns from images. They can be used to support the detection and monitoring of health complications.

Feature analysis and dimensionality reduction

Important for understanding variables in the dataset while removing redundant or irrelevant features. This process can improve the efficiency and accuracy of the analysis, particularly when working with high-dimensional data. Techniques such as Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) are commonly used for dimensionality reduction and feature selection, respectively.

Modelling approaches

In the field of population health analytics, it is common for many people to perform de novo prediction model development for the same purpose, often using slightly different variables and different datasets. It is important to consider whether it is possible to utilise existing information from previously developed models and target generalisability.

Building a new model

If no similar models exist or the intention is to utilise new information (predictors or outcomes), a de novo approach is appropriate. It offers maximum flexibility but requires sufficiently large, high quality data to support model training, tuning and validation.

Transfer learning

This approach leverages knowledge gained from one task or dataset to improve model performance on a different but related task or dataset. Instead of training a model from scratch, transfer learning begins with a pretrained model, which is then fine-tuned on a new dataset with similar structure or features. It can help with generalisability and alleviate statistical power constraints when limited data is available.

Ensemble methods

Rather than using a single model, multiple models are combined. This approach leverages the diversity of models, either through different algorithms, subsets of data, or training strategies, and aggregates their outputs through techniques such as voting, averaging, or stacking. It has the potential to improve overall performance and reduce overfitting.

Federated learning

Uses data from multiple sources that cannot be centralised or hosted in the same location due to privacy, regulatory, or logistical constraints. It involves an iterative process where models are trained locally at each data source, and only the model parameters (not the raw data) are shared with a central aggregator. These updates are then combined to refine a global model, which is redistributed for further training.

Public health surveillance

Public health surveillance

Public health surveillance is a key component of population health analytics, as it involves the use of data science methods and techniques to identify patterns and trends in disease incidence and prevalence, and to inform evidence-based public health policies and interventions.

Public health surveillance can involve the collection of data from a variety of sources, such as electronic health records, national health surveys, and administrative datasets.

The data collected can include information on disease incidence and prevalence, risk factors for disease, healthcare utilisation and costs, as well as demographic and socioeconomic characteristics of populations.

The use of data science techniques, such as machine learning and statistical analysis, can help identify patterns and trends in health-related data, and to develop predictive models for disease outcomes. Public health surveillance can also involve the use of data visualisation and communication tools, such as dashboards and infographics, to make health-related data more accessible and understandable to policymakers, healthcare providers, and the public.

Types of surveillance

There are several types of public health surveillance, including passive, active, syndromic, behavioural, and environmental:

Passive surveillance

The routine reporting of health data from healthcare providers to public health agencies.

Active surveillance

Targeted data collection to identify specific health outcomes or risk factors.

Syndromic surveillance

The real-time monitoring of health data, such as emergency department visits or internet searches, to identify potential outbreaks or epidemics.

Behavioural surveillance

Monitoring behaviours that may be associated with increased risk for a particular disease or condition, e.g., alcohol consumption or diet.

Environmental surveillance

Monitoring environmental factors that may be associated with increased risk for a particular disease or condition, such as air pollution or access to healthcare.

Each type of surveillance has its own strengths and weaknesses and may be more appropriate for certain types of diseases or conditions. The choice ultimately depends on the public health question being asked and the available data sources. Public health surveillance is a critical component of population health analytics, as it involves the use of data science methods and techniques to identify patterns and trends in disease incidence and prevalence, and to inform evidence-based public health policies and interventions that promote health and well-being for all.

Data sources

Population health analytics and public health surveillance both rely on a wide range of data sources in the UK:

Electronic health records (EHRs)

Digital records of a patient's health information that are widely used in the UK and are maintained by healthcare providers such as hospitals and general practitioners. EHRs can provide rich and detailed data on individual patients and can be used to track health outcomes over time.

National health surveys

Large-scale surveys conducted by government agencies and research organisations to collect information on the health status, behaviours, and attitudes of the UK population. National health surveys can provide valuable data on population health and can be used to identify health disparities and inform policies and interventions.

Administrative datasets

Records of activities and transactions within a healthcare system, such as hospital admissions, emergency department visits, and prescription drug dispensing. These datasets are typically collected and maintained by healthcare providers and payers, such as hospitals and insurance companies. Administrative datasets can provide detailed data on healthcare utilisation and costs and can be used in population health analytics to identify patterns and trends in healthcare utilisation, as well as to evaluate the effectiveness of healthcare interventions.

Audit and registry data

Structured collections of data focused on specific diseases, conditions, procedures, or populations. These datasets are often developed through clinical audits or national registries, such as cancer registries or joint replacement registries. They capture detailed and high-quality clinical information, including diagnostic data, treatment pathways, and outcomes. Audit and registry datasets are valuable for benchmarking clinical performance, monitoring disease trends, and supporting epidemiological research and quality improvement initiatives in population health.

Population research data

Large-scale, research-focused repositories that collect and store biological samples (such as blood, saliva, or tissue) alongside extensive health, lifestyle, genetic, and sometimes imaging data from participants. UK Biobank is a prominent example. Biobank datasets enable the study of genetic, environmental, and lifestyle determinants of health and disease. They are especially valuable for longitudinal studies, gene-environment interaction research, and the development of precision medicine approaches in population health.

Digital trace data

Data generated passively or actively through digital technologies such as smartphones, wearables, and connected devices, often collected outside traditional healthcare settings. A core example is the NHS COVID-19 app that was used to monitor geolocation, identifying proximity contact via Bluetooth and supporting symptom reporting.

Other data sources that can be used in population health analytics in the UK include data on social determinants of health, such as income, education, and housing, as well as environmental data, such as air quality and water quality. Data from wearable devices and other personal health technologies can also be used. These data sources can provide important insights into the factors that influence health outcomes.

In the UK, organisations like NHS England and the UK Health Security Agency are responsible for managing and analysing health data for public health surveillance purposes. These organisations ensure that data is collected, stored, and used in accordance with ethical and legal standards, and that appropriate data governance policies and procedures are in place. By using a variety of data sources and data analysis methods, population health analytics and public health surveillance can provide valuable insights to inform evidence-based policies and interventions.

The challenges of population health analytics

The challenges of population health analytics

There are several challenges associated with population health analytics, which include:

Data quality

To realise the potential benefits of population health analytics requires high quality, accurate and consistent data. Health data can be complex and heterogeneous, and may be incomplete or inconsistent, making it difficult to draw meaningful conclusions. Data pre-processing and cleaning can also be complicated and time consuming, particularly when working with large and complex datasets. It is important to carefully consider the appropriate techniques for each specific dataset and analysis, as well as ensure that data cleaning and pre-processing do not introduce biases or inaccuracies.

Data integration

Health data is often fragmented and stored in multiple locations, making it difficult to integrate data from different sources for analysis. This requires innovative solutions for integration and interoperability, including the use of data standards and common data models. Integrated care records have been a positive advancement in recent years, allowing a more holistic understanding of a patient's health. However, these are often limited to smaller geographic regions.

Resource constraints

Population health analytics requires significant resources, including skilled personnel, technology, and infrastructure. Many healthcare organisations and public health agencies in the UK may face resource constraints, making it difficult to invest in population health analytics initiatives.

Data privacy and confidentiality

These are critical considerations in population health analytics, particularly given the sensitive nature of health information. Health data must be collected and used in a way that respects individuals' rights and interests, with appropriate safeguards in place to protect the privacy and confidentiality of the individual. This requires adherence to relevant data protection laws and regulations, such as the GDPR, as well as taking measures such as de-identification or anonymisation of personal information, implementation of secure data storage and transfer protocols, and establishing data sharing agreements and access controls.

Data bias and fairness

The dangers of bias and fairness in population health analytics are significant and can have serious consequences for health outcomes and equity. Bias in data analysis can occur in various forms, such as sampling bias, measurement bias, or confounding variables. When bias is present in the data used for analysis, it can lead to inaccurate or incomplete conclusions, such as underestimation or overestimation of disease prevalence or incidence, which can have negative impacts on public health policies, interventions, and resource allocation.

Fairness is another important consideration, referring to the ethical principle of treating all individuals or groups equitably, without discrimination or bias. When fairness is not addressed in population health analytics, it can lead to discrimination, stigmatisation, or exclusion of certain groups. For example, if a particular group is systematically excluded from a health survey or study, their health needs and risks may not be adequately addressed in public health policies and interventions, leading to disparities in health outcomes.

Data governance and ownership

These are important legal considerations in population health analytics, particularly when working with large and complex datasets. It is crucial to ensure that data is collected, stored, and used in accordance with ethical and legal standards, and that data ownership and access rights are clearly defined and transparent. In the UK, organisations like NHS Digital and Public Health England are responsible for developing and enforcing data governance policies and procedures.

Effective data governance and ownership require collaboration and communication among stakeholder groups, including researchers, healthcare providers, policymakers, and the public. Transparent and accountable mechanisms for data governance and ownership must be established to ensure that individuals' rights and interests are respected throughout the data lifecycle, from collection to disposal. Adhering to data governance and ownership standards is essential to ensure that the use of health data in population health analytics is conducted in a way that is legal, ethical, and transparent.

Informed consent

In the UK, research involving human subjects requires researchers to obtain informed consent from participants and to ensure that participants understand the purpose, risks, and benefits of the research. However, obtaining informed consent in population health analytics can be challenging, particularly when working with large datasets or secondary data sources, where obtaining individual consent may not be practical or feasible. Several approaches to obtain informed consent in population health analytics, such as opt-out consent or broad consent, are used to address this, but there are ongoing debates about their effectiveness and ethical implications.

Communicating with patients

An additional complexity is added when a Population Health programme needs to contact a patient after identifying them. This requires them to understand the channels the patient can be contacted on and ensure they have the right details and accessibility requirements. As with other health data, these can often be outdated or missing. Therefore programmes, need procure or build messaging fulfilment solutions to enable this, which can be both costly and time-consuming.

Case study: Communicating with patients at scale

Challenge

600 million is a massive amount. Yet this is the number of patient contacts that NHS England (NHSE) recorded with GPs, community clinics, hospitals, NHS 111 and ambulance services across 2023-24 – **the equivalent of 1.7 million interactions with patients every day.**

On top of engagement around appointments, much of this communication involves calling patients for health screenings, vaccinations or healthcare referrals. But the rising costs associated with patient messaging – particularly the distribution of physical letters – is unsustainable.

To mitigate this, NHSE actively sought to create a service that would simplify message delivery and reduce costs but could also provide a fast and reliable patient communications method that could be easily extended as the need arose – for example around urgent vaccination programmes.

What resulted was NHS Notify – a platform which has unified patient messaging and allowed a broad range of NHSE organisations and services to easily adopt the same process using a single integration.

By enabling a more effective communication process, to date **over 257 million patient messages have been processed** from various services using NHS Notify including **88 million to the NHS App, 87 million via SMS messages, 59 million emails, and 22 million letters**.

Solution

NHSE, CGI and Hippo (CGI' digital partner) formed a specialist team across strategy, customer experience, product design, and engineering. From this, NHS Notify was conceived as a scalable and reusable service with an Application Programming Interface (API), which would streamline the distribution of NHS App messages, emails, texts, and letters to patients and the public, enabling adoption across all NHSE organisations.

Interactions with patients every day

1.7M

Patient messages processed

257M

Delivered in just

6 months

Data quality

CGI established data guardrails to ensure message consistency. NHS strategy mandates that only data from its Personal Demographics Service (PDS) – the national electronic database of NHS patient demographic data – should feed into the messaging platform as the single source of truth. CGI worked with colleagues in the NHS PDS team to increase data quality by analysing failed communications with a view to resolving these.

Scaling in the cloud

CGI built a unified messaging system on an Amazon Web Services (AWS) platform, utilising services like Lambda, SQS, DynamoDB, S3, and others. These were selected for their scalability, ability to meet NHS delivery needs, and ease of maintenance for future adopters.

The AWS solution also addressed cost and consistency goals while providing users with detailed performance insights, including message volumes per channel and the shift from physical to digital communication.

Additionally, CGI integrated with NHS England's Apigee API Management tool, creating an internet-facing API which allows integrators like the national vaccination and screening programmes, e-Referral Service, external GP Suppliers, and more the ability to send messages through NHS App, SMS, Email and Letter formats to recipients.

Outcome

Initially engaged in January 2023, the first service to use NHS Notify started six months later in July 2023 supporting GP registrations. There quickly followed a succession of other programmes going live using the product including the Autumn/Winter Flu and COVID Vaccination Programme in late 2023, NHS e-Referral Service in April 2024 and the National Breast Screening Programme in December 2024.

Communications
streamlined

Messaging costs
reduced

Sustainability
boosted

The project referenced in this case study was delivered by BJSS, which CGI acquired in February 2025.

How CGI can help

How CGI can help

Partnering with health and care organisations CGI delivers transformative digital solutions that enhance patient experiences, streamline operations, and improve outcomes.

Combining global expertise with local insight, we co-create strategies that place people at the centre of care.

Our services encompass digital transformation, cybersecurity, intelligent automation, and advanced analytics, tailored to meet the unique challenges of the health sector.

With a commitment to collaboration and innovation, CGI is dedicated to supporting the evolution of health and care services.

Go to health page

References

Alpaydin, E. (2014). Introduction to machine learning. MIT Press.

Barr, B., Kinderman, P., & Whitehead, M. (2015). Trends in mental health inequalities in England during a period of recession, austerity and welfare reform 2004 to 2013. Social Science & Medicine, 147, 324-331.

Bellazzi, R., & Zupan, B. (2008). Predictive data mining in clinical medicine: current issues and guidelines. International Journal of Medical Informatics, 77(2), 81-97.

Birkhead, G. S., Klompas, M., Shah, N. R., Uses of electronic health records for public health surveillance to advance public health. Annual Review of Public Health, 36, 345-359.

Centers for Disease Control and Prevention (CDC). (2021). Types of Surveillance. cdc.gov/surveillance/

Department of Health and Social Care. (2019). Prevention is better than cure: Our vision to help you live well for longer. gov.uk/government/publications/prevention-is-better-than-cure-our-vision-to-help-you-live-well-for-longer

Elliot, A. J., Kara, E. O., Loveridge, P., Bawa, Z., Morbey, R. A., Moth, M., ... & Smith, G. E. (2015). Internet-based remote health self-checker symptom data as an adjuvant to a national syndromic surveillance system.

Epidemiology & Infection, 143(16), 3416-3422.

Evans-Lacko, S., Knapp, M., McCrone, P., Thornicroft, G., & Mojtabai, R. (2013). The mental health consequences of the recession: economic hardship and employment of people with mental health problems in 27 European countries. PloS one, 8(7), e69792.

Health Data Research UK. (n.d.). Federated analytics. hdruk.ac.uk/research/research-data-infrastructure/federated-analytics/

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.

Information Commissioner's Office. (2021). Data protection and coronavirus. ico.org.uk/global

Kelly, F. J., & Fussell, J. C. (2015). Air pollution and public health: emerging hazards and improved understanding of risk. Environmental Geochemistry and Health, 37(4), 631-649.

Kindig, D., & Stoddart, G. (2003). What is population health?. American Journal of Public Health, 93(3), 380-383.

Kings Fund. kingsfund.org.uk/insight-and-analysis/data-and-charts/key-facts-figures-nhs#:~:text=For%20example%2C%20in%20emergency%20care,million%20admissions%20in%202023%2F24

Little, R. J., & Rubin, D. B. (2019). Statistical analysis with missing data. John Wiley & Sons.

Miksad, R. A., Abernethy, A. P., Voelker, J., Fu, A., Park, J. S., Gillett, B., & Schilsky, R. L. (2023). Integrating real-world evidence into regulatory decision making for oncology: A call to action. JCO Clinical Cancer Informatics, 7, Article e2300035. pmc.ncbi.nlm.nih.gov/articles/PMC10196539/

NHS Digital. (2021). digital.nhs.uk

NHS England. (2025, February). Waiting list falls as NHS staff treated record numbers last year. england.nhs.uk/2025/02/waiting-list-falls-as-nhs-staff-treated-record-numbers-last-year/

NHS Health Research Authority. (2021). Applying for research ethics approval. hra.nhs.uk/approvals-amendments/what-approvals-do-i-need/research-ethics-committee-review/applying-research-ethics-committee

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453.

Observational Health Data Sciences and Informatics. (n.d.). Common Data Model. ohdsi.github.io/CommonDataModel/

Open Data Institute. (n.d.). What is open data?. theodi.org/news-and-events/blog/what-is-open-data/

Public Health England. (2020). Health inequalities: A briefing on health inequalities in England. assets.publishing.service.gov.uk/media/615213efe90e077a2db2e804/health_inequalities_briefing.pdf

Public Health England. (2020). National Child Measurement Programme. gov.uk/government/collections/national-child-measurement-programme

Public Health England. (2021). Our data and analysis. gov.uk/guidance/phe-data-and-analysis-tools

Rasooly, I. R., Lu, M., Morse, K. E., Zheng, N. S., Budnitz, D. S., & Leonard, C. E. (2024). The FDA Adverse Event Reporting System: A descriptive analysis of use with real-world data. Frontiers in Pharmacology, 15, Article 11075801. pmc.ncbi.nlm.nih.gov/articles/PMC11075801/

Rutter, H. (2012). The single most important intervention to tackle obesity... International Journal of Public Health, 62(7), 859-860.

Szklo, M., & Nieto, F. J. (2014). Epidemiology: Beyond the basics. Jones &Bartlett Publishers.

The King's Fund. (2024, October). Public health and population health: Leading together. assets.kingsfund.org.uk/f/256914/x/d90fd1c682/public_health_population_health_leading_together_report_oct2024.pdf

UK Biobank. (2021). About us. ukbiobank.ac.uk

UK Data Service. (2021). ukdataservice.ac.uk/about

UK Health Security Agency. (n.d.). UKHSA data strategy. gov.uk/government/publications/ukhsa-data-strategy/ukhsa-data-strategy#challenges

West, R., Raw, M., McNeill, A., Stead, L., Aveyard, P., Bitton, J., ... & Borland, R. (2015). Health-care interventions to promote and assist tobacco cessation: a review of the efficacy, effectiveness, and affordability for use in national guideline development. Addiction, 110(9), 1388-1403.

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. Springer.

World Medical Association. (2013). Declaration of Helsinki: Ethical principles for medical research involving human subjects. JAMA, 310(20), 2191-2194.

About CGI

Insights you can act on

Founded in 1976, CGI is among the largest IT and business consulting services firms in the world.

We are insights-driven and outcomes-focused to help accelerate returns on your investments. Across hundreds of locations worldwide, we provide comprehensive, scalable and sustainable IT and business consulting services that are informed globally and delivered locally.

For more information

- Visit cgi.com/uk
- Email us at enquiry.uk@cgi.com