

OBH® Segmentation Engine and Data Products

Cloud Software Service Description

OBH® Segmentation Engine and Data Products – Service Description

Providing a comprehensive backbone for your population health planning and analysis

Population segmentation aims to categorise the population according to health status, health care needs and priorities. This approach recognises that groups of people share characteristics that influence the way they interact with health and care services. To optimise health outcomes, service user experience, efficiency and care costs, care delivery systems should respond to the needs of different population segments in different ways. It is therefore essential to understand health and care population segments and their characteristics in detail.

Segmentation around population characteristics, rather than provider characteristics, can be challenging. Health systems have historically been organised largely around provider characteristics (e.g. clinical specialties like cardiology), rather than population characteristics (e.g. people living with frailty). This is increasingly unsustainable both financially and in terms of improving people's outcomes. Care systems organised around people, rather than providers, offer potentially the only sustainable long-term solution. Accurate longitudinal population segmentation data is vital to this.

OBH's Segmentation Engine, Segmentation Dataset and associated data products provide an essential, core backbone for population health management, service planning, design and evaluation, case finding and risk stratification, for any local NHS organisation, and their health and care partners. The OBH Segmentation Engine and Dataset are available to support both secondary uses and specific direct care, cohorting, risk stratification and case finding applications (always within and subject to applicable information governance requirements).

Using person-centred data for population health is a key enabler to improving health outcomes across an entire population or cohorts of people, rather than focussing solely on one person at a time. These data products enable NHS organisations to determine cohorts for targeted interventions, and to achieve goals such as:

- keeping people healthy by providing preventive care services and promoting healthy behaviours and lifestyles,
- risk stratification and reducing health risk by identifying and caring for people with emerging risk factors or early stages of disease,
- coordinating care for people with long term conditions,
- case finding to identify undiagnosed conditions and/or undertreatment and care gaps, and
- allocating resources effectively across the population.

OBH's methodology is based on a data driven and evidence-based implementation of the internationally accredited 'Bridges to Health' segmentation model (Lynn et al 2007). Bridges to Health describes the core population segments as: Healthy / Generally Well, Long Term Conditions, Disability, Incurable Cancer, Organ Failure, and Frailty or Dementia. There are 60 subsegments that make up each of these segments, which include long term conditions such as diabetes, coronary heart disease, and asthma, amongst others, as well as subsegments such as severe heart failure, or severe frailty which are more relevant to the end phase of life segments. Most, if not all, modern health segmentation approaches derive originally from the foundational Bridges to Health methodology. OBH devised the original NHS data model for Bridges to Health over a decade ago, which was adopted across the NHS and has informed a wide range of national and local population health use cases and analyses.

These subsegments and conditions have all been created by OBH so that longitudinal registers go back at least 3 years (and subject to data availability, considerably further), as well as prospectively. This forms the basis of longitudinal, whole population condition registers, allowing for a range of population health insights and analyses, enabling like-for-like comparisons over time that simply aren't possible with traditional 'snapshot' data approaches.

National scale implementations of this segmentation architecture, methodology and output dataset(s) have been validated and utilised in leading peer-reviewed studies and evaluations. These include the largest ever study of multiple long-term conditions (MLTC), undertaken with a range of academic and NHS partner organisations[1], resulting in major peer-reviewed publications in world-class medical journals. For more information see:

- https://outcomesbasedhealthcare.com/publications/
- https://www.imperial.ac.uk/news/249937/multiple-longterm-conditions-affect-more-than/.

The OBH Segmentation Dataset includes data tables that are also a key source of 'sentinel metrics' that can measure key government health missions, including the shifts:

- **From Hospital to Community** - using OBH's segment-specific **Days Disrupted by Care (DDBC)** metric, this measures success in moving care closer to home, reducing reliance on hospitals and better co-ordination of care for people with Long Term Conditions.
- **From Sickness to Prevention** - using OBH's **HEALTHSPAN®** measure, this is the first and currently only whole population metric in the world to objectively measure how well the population is, reflecting the success of local primary prevention efforts and is available in both crude and standardised forms, to allow for meaningful comparison between and within geographies.

Key Features of the Segmentation Engine and Dataset

Data in, data out
OBH's Segmentation Engine can ingest and integrate a variety of different data sources, including primary care data, secondary care datasets (such as SUS) and other nationally consistent or locally held datasets (such as community, mental health, maternity). After the data transformation process, which includes standardisation of source datasets and a series of complex business and clinically validated logic, the Engine produces a Segmentation Dataset which unifies information across multiple sources into a person-centred dataset. The Segmentation Dataset includes each person registered to a GP practice in a given local area, and holds information on which population segment and subsegments (or long term conditions) an individual is allocated to.
A dynamic, longitudinal model to analyse trends
The **OBH Segmentation Dataset** is structured to provide a longitudinal and retrospective baseline view of the information it holds over a number of years, and prospectively thereafter following periodic refreshes of the data. This method accounts for people who move into and out of area, as well as births and deaths. Depending on the selected Segmentation Dataset product, this also allows for application of evidence-based logic that identifies when conditions may have resolved and/or gone into remission. The unique fully longitudinal time period of the Segmentation Dataset allows for the evaluation of interventions retrospectively and prospectively, as well as analysis of trends over time in population movement and disease progression. These are essential requirements for population health management approaches including segmentation, cohorting, risk stratification and case finding.
Condition registers monitored longitudinally
With the **OBH Segmentation Dataset**, it is simple for analysts to create a snapshot view at any single time point in order to get a view of condition registers at a population and person-level. This functionality goes far beyond just taking a 'snapshot' of the currently alive population and then looking back at prior dates of diagnoses for that current population, and is significantly more powerful and analytically versatile as a result. OBH's Segmentation Dataset also (uniquely) maps historical populations, and their health/condition profiles, to enable like-for-like comparison of cohorts over time, including outcomes measurement and baselining. Without this, only prospective outcomes measurement is possible, which is generally inadequate for retrospective service evaluation and/or outcomes based commissioning.

[1] Hafezparast N, Bragan Turner E, Dunbar-Rees R et al. Adapting the definition of multimorbidity - development of a locality-based consensus for selecting included Long Term Conditions. BMC Fam Pract 2021;22(1):124. doi: 10.1186/s12875-023-02134-1 https://pubmed.ncbi.nlm.nih.gov/37905525/

Identifying the 'healthy and generally well' cohort for targeted primary prevention

Uniquely, the **OBH Segmentation Dataset** covers an entire GP registered population for the whole dataset period, and therefore also includes the 'healthy or generally well' cohort register, and how this cohort acquires risk factors and/or conditions over time. With this information it is possible to risk stratify the healthy/well cohort, for those most at risk of progression to ill health. With the use of additional risk factor information, those most at risk of progressing to develop their first long term condition could be identified and targeted services for preventative care can be designed and implemented, an example of case finding.

Developed based on evidence

The **OBH Segmentation Engine** transforms data using business and clinical logic and codes that are evidence-based, and derived from analysis of international and national best practice, guidelines and standards. **OBH** have spent over 11 years building and maintaining this database and codebase. OBH's clinical team undertake regular reviews of many hundreds of clinical guidelines and standards, and many thousands of clinical codes.

OBH's clinical team are widely recognised as experts in clinical informatics and PHM, combining years of front-line specialist and generalist clinical experience with training and qualifications across disciplines such as computer science, epidemiology/ public health and mathematical and statistical modelling.

When developing condition registers, OBH collaborates with external sub-specialists through OBH's Clinical Associates Network and NHS coding experts to understand clinical context and ensure accuracy across acute, community and GP settings. Extensive clinical code mapping between data sources ensures registers are source-agnostic, clinically robust, and easily expandable to new datasets.

Multiple applications and foundational use cases for PHM

The OBH Segmentation Dataset is a foundational data asset for PHM. It has a wide range of applications and use cases:

- **Cohorting:** There are in-built features to construct bespoke cohorts depending on use requirements
- **Segmentation:** Segmenting the population by the Bridges to Health model, or by constructing an alternative segmentation model based on the condition and risk factor features and attributes available
- **Disease/Condition Registers:** up to 60 condition and risk factor registers available longitudinally
- **Risk Stratification:** In-built risk scores, with more in continual development, as well as the ability to link to other datasets for the user to create new purpose-specific risk scores and/or link in alternative risk score outputs
- **Outcomes Measurement:** Central to the utility of this dataset are the critical longitudinal features that have been optimised specifically for cohorting to support person-centred outcomes measurement and service evaluation

Ease of use by data analysts

The **OBH Segmentation Dataset** has been designed as a 'dimensional data model', using a design approach for a database structure that has been optimised for PHM data analytics. Alongside a range of unique features, it consists of a central periodic snapshot 'Fact Table', connected to multiple 'Dimension' tables. The dimensional model is easy to understand and intuitive for analyst users. Information is grouped into coherent dimensions, making it easier to interpret and read. The model is denormalised to allow for query optimisation, and can be linked to other datasets for onward analysis.

Compatibility and Integration

The output Segmentation Dataset (and associated data products) is built as a relational data model, making it technology-agnostic and compatible with a wide range of platforms, including Azure, AWS, Palantir Foundry, Databricks, Snowflake and more.
Dataset Linkage: The Segmentation Dataset can be linked to different datasets or solutions (subject to approvals and IG) at different levels, but commonly it is linked using the Pseudo NHS number.
Query Languages: The dataset/s can also be queried with: Python, Pyspark, Spark SQL, SQL and R languages.

Analysis for health inequalities

The **OBH Segmentation Dataset** also includes data on ethnicity, deprivation (based on IMD using LSOA data), gender, and age to support population health evaluation and analysis on health inequalities at a person-level.

Validation, benchmarking, calibration and governance

OBH performs extensive testing on the data sources, and output Segmentation Dataset ahead of every release. This covers testing of prevalence and incidence, and benchmarking against hundreds of data sources. This comprehensive validation

ensures highest levels of data quality and accuracy and condition coverage, supporting both secondary care uses and direct care applications.

OBH regularly reviews regulatory standards, for example including recent external objective assessment against MHRA Software as a Medical Device guidelines, to ensure all our products align with current regulatory requirements and best practice. OBH's in-house team have considerable depth and expertise in information governance, including a qualified barrister/ clinician, who is a leading specialist in IG and ethics.

Overview of Segmentation Engine and Dataset Product Versions

Although the structure of the Segmentation Dataset tables are largely the same, OBH can produce different versions of the Segmentation Dataset depending on the selected geography, source data used, and the infrastructure or environment selected for processing. Depending on the version selected, the Segmentation Dataset enables and supports different 'types' of population health analyses.

A. National-Scale Segmentation Dataset

A National-Scale Segmentation Dataset is derived from a combination of nationally consistent and large-scale health and care data sources, assembled and maintained by OBH using standardised clinical definitions and methodologies. However, it does not include local data sources, such as GP data extracted direct from GP practices.

The OBH Segmentation Engine is run on secure, accredited cloud infrastructure to produce and refresh the National-Scale Segmentation Dataset, which can be made available to analysts and developers subject to applicable information governance requirements.

The Dataset features a multi-year longitudinal period of data with a defined refresh cadence. It includes six core population segments and up to 60 subsegments or conditions, capturing the entire registered population over the dataset period, including people who have died and those who move into and out of area.

Upgraded National-Scale Segmentation Dataset

The National-Scale Segmentation Dataset can be 'upgraded' with locally held data, for example data from GP practices, to enhance accuracy, granularity and analytical utility for local population health analysis. Local upgrades are available only to NHS organisations and apply to the Buyer's registered population, subject to information governance approvals.

There are two upgrade options available for different complexities of data integration:

- **Standard Upgrade**
  A longitudinal Segmentation Dataset with integrated local condition registers (including dates of condition diagnosis, where available). Suitable for some population health analyses such as prevalence trends which may also require interpretation, depending on local data used.

  Technical implementation: OBH provide template schemas and integration methods. Local data sources are provided by the client - this could be for only some of the conditions included in the Segmentation Dataset. Dates of diagnosis need to be provided to incorporate the longitudinal element of this upgrade option. Integration and validation/testing of the Standard upgrade is performed by the client with OBH support.

- **Advanced Longitudinal Upgrade**
  A longitudinal Segmentation Dataset with integrated longitudinal GP data, at daily granularity. Suitable for advanced population health analyses such as longitudinal prevalence, incidence, activity, expenditure and outcomes trend analysis.

  Technical implementation: OBH use specific data extracts from GP practices and integrate this data into the Segmentation Dataset. All integration, validation and testing of the Segmentation Dataset is

performed by OBH. This enables every condition in the Segmentation Dataset to be upgraded consistently with local GP data.

B. Local Segmentation Dataset

The 'stand alone' Local Segmentation Dataset:

- features a 3 year longitudinal period of data initially, with daily granularity, which progressively builds prospectively over time. It is refreshed regularly with the latest data. The segments (and subsegments) can be configured for specific local population needs.

- is derived from data sources available locally (SUS data and GP data). OBH 'run' the Segmentation Engine (an advanced data analytics pipeline) on the SUS and GP data on OBH's Microsoft Azure cloud infrastructure. It is tested and validated against benchmark data. The Segmentation Dataset is delivered back to the Client's own infrastructure for onward analysis and use.

Local Data Upgrade - Key Features

	Standard Upgrade	Advanced Longitudinal Upgrade
Time period of dataset (longitudinal or snapshot)	3+ years	7+ years
Granularity (Datapoints available across the time period of the dataset)	Monthly	Daily
6 core segments available	Yes	Yes
Conditions (subsegments) available and upgraded with local data	Some conditions upgraded with local data only for the current population (depending on local data availability)	All 60 conditions upgraded with local GP data
Socio-demographic data available (age, gender, ethnicity)	Yes	Yes
Ability to perform types of analysis only possible on longitudinal data (eg trend analysis, impact evaluation)	Yes (with caveats depending on on completeness of local data used)	Yes
Accounts for population movement into and out of area, births and deaths	Yes (but only current population will have conditions upgraded with local data)	Yes
Includes feature which captures time when a condition resolves or remits following an initial diagnosis	No	Yes
Ability to perform condition prevalence (total population with condition) calculations	Yes - snapshot and changes over time (with caveats depending on completeness of local data used)	Yes - snapshot and retrospective changes over time

Ability to perform condition incidence (new diagnostic cases of conditions) calculations	Yes (with caveats depending on completeness of local data used)	Yes

Local Data Upgrade - Requirements

	Standard Upgrade	Advanced Longitudinal Upgrade
Technical implementation of local data upgrade	Client-led (template schemas, condition definitions to enable mapping and integration methods with business and clinical logic provided)	OBH-led (includes longitudinal GP data extractions, integration and linkage of local data and QA/validation)
Validation of Segmentation Dataset	Client-led with OBH supporting tools/materials	OBH performs extensive validation against established benchmarks such as QOF and Audits
Local data requirement	Patient-level condition registers with a date of diagnosis for each condition for the current population	OBH extract specifications from GP practices to integrate directly into the Segmentation Dataset. This covers the entire population over the period of the Segmentation Dataset, including births, deaths and movements in and out of area.
Linkage and integration mechanism of local data	Use of pseudonymised NHS number	Use of pseudonymised NHS number
Information governance	Agreements in place for the local data patient-level condition registers to be integrated/linked with the National-Scale Segmentation Dataset	Agreements in place for OBH to act as Processors for GP data integration with the National-Scale Segmentation Dataset

Overview of Associated Products

The following associated products are available.

Outcomes Data	Designed, tested and validated over 12 years working successfully with local NHS systems, OBH provides Outcomes Data specifically designed so that it can be linked to the Segmentation Dataset for outcome measurement purposes. Outcome 'categories' are available to be selected from OBH's Outcomes Data Library. These outcome categories provide the pre-configured and quality assured 'numerator' component of the outcome measures. Each 'outcome category' selected can be applied to different cohorts of interest from the OBH Segmentation Dataset and can therefore be used flexibly by analysts to measure multiple outcomes, according to local priorities. This significantly reduces testing, development, QA/validation and clinical assurance time for NHS analyst teams and significantly speeds up the process of developing a local outcomes framework, populated with high quality and consistent outcomes baseline data. Example outcome categories include mortality, complications of long term conditions, or adverse events. In addition, activity measures can also be selected such as emergency attendances, hospital admissions, or outpatient appointments, as well as more complex measures involving readmission or time spent at home.
HEALTHSPAN® Data	Population-level HEALTHSPAN, is an objective, population-level measure of the amount of time individuals spend in good health, usually described as a proportion of their overall lifespan. Derived from **OBH's Segmentation Dataset**, HEALTHSPAN data includes aggregate data for crude HEALTHSPAN and Lifespan.
Segmentation Analytics Static Report	Derived from **OBH's Segmentation Dataset**, **OBH** are able to perform a range of standardised and advanced analytics to provide a comprehensive report (usually in PDF format) which summarises a local population's segments and subsegments, in clinical, public health, activity and expenditure terms. This analysis is typically performed as a one-off or periodic 'deep dive', for example in support of a local JSNA, Health Needs Assessment, to inform ICB strategy and planning and/or other Population Health Management uses. It also includes **HEALTHSPAN®** analytics.
Segmentation Analytics Dashboard	Derived from **OBH's Segmentation Dataset**, **OBH** include a range of standardised and advanced analytics which summarises a local population's segments and subsegments, in clinical, public health, activity and expenditure terms in a Segmentation Dashboard and is updated on a quarterly basis. Includes **HEALTHSPAN®** analytics.
Outcomes Platform	More information via dedicated G-Cloud 15 offer: "**OBH Outcomes Platform**".

Associated Products

Outcomes Data

OBH provides Outcomes Data specifically designed so that it can be linked to the Segmentation Dataset for outcome measurement purposes.

Outcome 'categories' are available to be selected from OBH's Outcomes Data Library (see table below for examples). These outcome categories provide the pre-configured and quality assured 'numerator' component of the outcome measures, for example 'days disrupted by care'. This can then be used to configure the outcome based on local population needs, and is then applied to the relevant cohort/segment/subsegment population (ie. the 'denominator' or population of interest). For example, 'days disrupted by care in people with frailty' – in this example, the category is 'days disrupted by care'. Each 'outcome category' selected can be applied to different cohorts of interest from the OBH Segmentation Dataset and can therefore be used flexibly by analysts to measure multiple outcomes, according to local priorities. This significantly reduces testing, development, QA/validation and clinical assurance time for NHS analyst teams and significantly speeds up the process of developing a local outcomes framework, populated with high quality and consistent outcomes baseline data.

Other outcome categories include mortality, complications of long term conditions, or adverse events. In addition, activity measures can also be selected such as emergency attendances, hospital admissions, or outpatient appointments, as well as more complex measures involving readmission or time spent at home.

The Outcomes Data products are person-level and structured such that they can be easily linked to the selected Segmentation Dataset at patient level to enable flexible application to the desired cohort and/or condition.

The following table shows outcomes data availability depending on the selected Segmentation Dataset:

	Outcomes Availability
National-Scale Segmentation Dataset	Selected outcomes from OBH Outcomes Data Library
• Upgraded - Standard	Selected outcomes from OBH Outcomes Data Library
• Upgraded - Advanced Longitudinal	All outcomes from OBH Outcomes Data Library
Local Segmentation Dataset (SUS + GP)	All outcomes from OBH Outcomes Data Library

HEALTHSPAN Data

Population-level HEALTHSPAN, is an objective, population-level measure of the amount of time individuals spend in good health, usually described as a proportion of their overall lifespan. It measures when people develop their first significant long term health condition, and the average (median, or mean) age at which this occurs within a population. It is analysed in the context of population-level Lifespan, and is also represented as a ratio of HEALTHSPAN to Lifespan.

Derived from OBH's Segmentation Dataset, HEALTHSPAN data includes aggregate data on crude HEALTHSPAN and crude Lifespan:

- Aggregate data (5th, 25th, 75th and 95th percentile, and median age values) for crude HEALTHSPAN, by gender, and sub-ICB at its lowest geographic level.
- Aggregate data (5th, 25th, 75th and 95th percentile, and median age values) for crude Lifespan, by gender, and sub-ICB at its lowest geographic level.

Data is broken down into:

- Rolling 12 month data (to align with the end of each financial quarter)

- Quarterly data

Segmentation Analytics Dashboard

Derived from **OBH**'s Segmentation Dataset, **OBH** are able to perform a range of standardised and advanced analytics, accessed via an online dashboard, which summarises a local population's segments and subsegments, in clinical, public health, activity and expenditure terms. This is updated quarterly to enable monitoring of trends. The main domains available are set out below, with some illustrative examples:

A. Whole Segmentation Model: Analysis of the model as a whole population

This set of analyses explores the health status of people across the whole segmentation model and changes over time.

Illustrative examples of analyses produced are:

- Relative sizes of each population segment
- Number of people in multiple segments
- How the overall number of segments per individual changes over time
- How the proportion of people in each segment differs by socio-demographic and health inequalities factors such as deprivation, and ethnicity
- How people move between and within segments dynamically over time
- Number of hospital contacts by individual segment or segment combination, including outpatient appointments, hospital admissions, average length of stay (and total days disrupted by care), and A&E attendances
- Segment-specific analyses of contacts with GP surgeries and other clinical resources

B. Generic Segment Profiles: Segment-specific analysis applicable to all segments

There are analyses that can be performed for each segment and shown longitudinally using monthly and annual snapshots to understand trends.

Some illustrative examples of segment-specific analyses include:

- Age distribution of people in each segment at specific timeframes
- Gender, ethnicity, deprivation and risk factor profiles of each segment
- Number of hospital contacts including outpatient appointments, hospital admissions, average length of stay (and total days disrupted by care), and A&E attendances
- Number of GP contacts (per person), if GP data available
- Analysis related to segment overlaps and turnover, how people move between segments, and towards to the end of life
- Co-morbidity profiles, and progression rates to multiple long term conditions
- Capitated activity and expenditure associated with each segment

C. Specific Segment Profiles: Analysis specific to each segment

For each population segment, there are a number of specific analyses that can be performed that only apply to the specific population segment.

Some illustrative examples include:

- for people who are in the 'healthy/generally well' segment, the age at which people move to any other 'non-healthy' segment in the given year (i.e. **HEALTHSPAN®**). This can be analysed in the context of lifespan trends (i.e. **HEALTHSPAN**:**LIFESPAN®** ratio), or the **HEALTHSPAN GAP®** between most and least deprived
- for people who are in the 'LTC' segment, how multi-morbidity (defined here by the number of LTCs) changes with age, health inequalities, and risk factor profiles
- for people in the 'LTC' segment, numbers of people with single and multiple LTCs, and how they progress between the two

- for people for have died, proportion of people on the palliative care register vs not on the palliative care register
- for people in the 'frailty' segment, numbers of people in each frailty group , and how they change over time and by age

Segmentation Analytics Static Report

Containing similar types of analyses described above for the Segmentation Analytics Dashboard, this takes the form of a comprehensive report (usually in PDF format, suitable for inclusion in Board papers for example) which summarises a local population's segments and subsegments, in clinical, public health, activity and expenditure terms. This analysis is typically performed as a one-off or periodic 'deep dive', for example in support of a local JSNA, Health Needs Assessment, to inform strategic commissioning strategy and/or other population health management uses. This report helps support areas who need further insights around their populations before deciding on which of **OBH**'s cloud-based offerings better suits their needs. The report's length and complexity will be jointly decided between **OBH** and the client.

OBH Outcomes Platform

While the **OBH Outcomes Platform** draws on the power of the **OBH Segmentation Engine and Dataset**, it is set out as a separate standalone G-Cloud 15 offer, due to its complexity and additional implementation requirements.