outcomesbasedhealthcare

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, which OBH devised the original NHS data model for, over a decade ago. This has now become widely adopted across the NHS, including nationally, covering over 60 million people registered to a GP in England.

These subsegments and conditions have all been translated so that longitudinal registers going back to 2016 are derived from existing national and local data sources, enabling baselining and benchmarking for every geographical level in the NHS. In total the Segmentation Dataset currently provides 60 subsegments/conditions retrospectively over a period of at least 3 years, but nationally at least 8 years, as well as prospectively. This forms the basis of longitudinal, whole population condition registers, allowing for a range of population health insights and analyses that simply aren't possible with traditional 'snapshot' data approaches.

The national version of this dataset has been validated and utilised in many peer-reviewed studies and evaluations, including 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 national version of this 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 **standardised HEALTHSPAN®** measure, this is the first and currently only whole population metric in the world to objectively measure how well the population is at both a national and local level in a consistent manner, reflecting the success of national and local primary prevention efforts. This metric is in use by the NHSOF.

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, national SUS data, other national datasets (such as community, mental health, maternity), the National Segmentation Dataset, and more. For example, the national version of the dataset combines 14 different national data sources, spanning both secondary and primary care data, going back as far as 2008. These are often provider-specific, and event-based. After the data transformation process, which includes standardisation of source datasets and a series of complex business and clinically validated logic, it produces a Segmentation Dataset which includes unified information from across the multiple ingested data sources. The Segmentation Dataset is a person-centred dataset which 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 10 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:** 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. The National Segmentation Dataset is currently in use by NHS England in Azure (UDAL) and Palantir Foundry (FDP) environments. Because it follows relational principles, the dataset can be deployed in traditional database systems, or within distributed big data infrastructures. Users interact with the data via a range of interfaces including code-based tools such as Databricks (UDAL), SSMS (UDAL) and Code Workbooks/Code Repositories (Foundry), as well as point-and-click tools such as Contour (Foundry), Tableau (UDAL) .

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; for example the National Segmentation Dataset is frequently linked to SUS to calculate A&E attendances by health state, segment or condition/s. In addition, it can be linked to other segmentation models or risk scores for different purposes.

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, there are a number of different versions of the Segmentation Dataset available depending on the 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.

This section provides an overview of the different product versions of the **OBH Segmentation Dataset** available:

A. National Segmentation Dataset

The **National** Segmentation Dataset:

- features a 8+ year longitudinal period of data, with monthly granularity. It is refreshed regularly with the latest data. It includes 6 core segments, and 60 subsegments/conditions/risk factors.

- is derived from numerous data sources available nationally, including historic versions of the data sources often going back at least 10 years. These data sources include Master Patient Index, SUS, ECDS, Community Services Dataset, Mental Health Services Datasets, and a number of other sources. The OBH Segmentation Engine (an advanced data analytics pipeline) is 'run' on these source datasets on NHS England infrastructure regularly, and the National Segmentation Dataset is made available to analysts and developers.

- is available to ICBs, and other NHS organisations for their GP registered populations, subject to satisfying information governance requirements.

Upgraded National Segmentation Dataset

The National Segmentation Dataset can be 'upgraded' with local data, for example data from GP practices, on local infrastructure (including on a local FDP instance), subject to information governance. Note that Segmentation Dataset upgrades are only provided to local NHS organisations.

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

- ● **Basic Snapshot Upgrade**
  A snapshot Segmentation Dataset, with integrated local data. Suitable for basic population health analyses at a specific point in time.

  Technical implementation: OBH provide template schemas and integration methods. Local data sources are provided by the client. Integration and validation/testing of the Basic Snapshot upgrade

is performed by the client with OBH support.

- **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.

Overview of Associated Products

Depending on whether the National Segmentation Dataset or the Local Segmentation Engine and Dataset is selected (see table below), 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 ICS strategy 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 14 offer: "**OBH Outcomes Platform**".

Detailed Product Descriptions

A. National Segmentation Dataset

The National Segmentation Dataset is derived from numerous data sources available nationally, including historic versions of the data sources often going back at least 10 years. These data sources include Master Patient Index, SUS, ECDS, Community Services Dataset, Mental Health Services Datasets, and a number of other sources). The **OBH Segmentation Engine** (an advanced data analytics pipeline) is 'run' on these source datasets on NHSE infrastructure regularly, and the National Segmentation Dataset is made available to analysts and developers.

The National Segmentation Dataset can also be made available to local NHS areas (such as ICBs) for their specific populations subject to information governance. The National Segmentation Dataset is refreshed regularly, and each refresh is made available to local NHS areas who have subscribed.

The National Segmentation Dataset features a 8+ year longitudinal period of data, with monthly granularity. It includes 6 core population segments (including people who are healthy/generally well), and 60 subsegments/conditions. Uniquely it captures the entire population during the entire 8+ year time period of the data, including those people who have subsequently died, people who move in and out of area, and births. It has been designed to enable analyses of either mutually exclusive or overlapping segments, depending on the precise purpose of the analysis being undertaken.

Local Upgrade Options of the National Segmentation Dataset
National Segmentation Dataset upgrades are only available to local NHS organisations. The National Segmentation Dataset can be upgraded with local data, for example data from GP practices, on local infrastructure, subject to information governance.

There are three upgrade options available for different complexities of data integration:
- Basic Snapshot
- Standard
- Advanced Longitudinal

Basic Snapshot Upgrade
A **single 'snapshot'** time point of the Segmentation Dataset (provided by NHSE) with integrated local data (such as GP data) to enhance the accuracy of the condition registers.

<u>Summary types of possible analyses:</u> Allows for a range of cross-sectional segment and subsegment prevalence analyses, including socio-demographic splits/analyses.

<u>Technical implementation and requirements:</u> OBH provide template schemas and integration methods. **Local data sources** are provided by the client. The local data requirement is a snapshot of the current population with a basic condition register of who has and does not have a long-term condition. Not all conditions in the Segmentation Dataset are necessary to be included for the upgrade, however additional interpretation is then required when performing further analysis on the Upgraded Segmentation Dataset. OBH provide clinical definitions for each of the conditions in the Segmentation Dataset which can be mapped onto local data condition registers. This will include support on which QOF conditions can be mapped. The **'snapshot date'** for the local data and the National Segmentation Dataset are required to match but this date can be defined by the client. **Integration and validation/testing** of the Basic Snapshot upgrade is performed by the client.

Standard Upgrade

A **longitudinal** Segmentation Dataset (provided by NHSE) with integrated local data (such as GP data) to enhance the accuracy of the condition registers.

Summary types of possible analyses: Allows for cross-sectional prevalence analysis, as well as key longitudinal population health analyses (with caveats depending on on completeness of the local data used).

Technical implementation and requirements: OBH provide template schemas and integration methods. **Local data sources** are provided by the client. The local data requirement is for a snapshot of the current population with a condition register **to include dates of diagnoses of these conditions**. Not all conditions in the Segmentation Dataset are necessary to be included for the upgrade, however additional interpretation is then required when performing further analysis on the Upgraded Segmentation Dataset. OBH provide clinical definitions for each of the conditions in the Segmentation Dataset which can be mapped onto local data condition registers. This will include support on which QOF conditions can be mapped. **Integration and validation/testing** of the Standard upgrade is performed by the client.

Advanced Longitudinal Upgrade

A longitudinal Segmentation Dataset (provided by NHSE) with integrated GP data at daily granularity. The integration is performed by OBH using GP data extracts based on OBH's clinically curated extract specification, and matching of condition definitions between datasets using SNOMED codes applied to GP data.

Summary types of possible analyses: All population health analyses that can be performed on longitudinal data are possible (see Appendix A for full list of types of analysis that can be performed), including any which are reliant on a daily level of dataset granularity (e.g. detailed understanding of patterns of condition sequencing).

Technical implementation and requirements: **Local data sources from GP practices** are extracted based on OBH specifications. All source data is tested and validated. **Integration** of the source data into the Segmentation Dataset is performed by OBH, so that all conditions are upgraded with local GP data. This enables every condition in the Segmentation Dataset to be upgraded consistently with local GP data. **Validation and testing** of the upgraded Segmentation Dataset is performed by OBH against QOF data and a wide range of other publicly available reference benchmarks and published data, to ensure quality, consistency and accuracy.

Refer to the tables below for comparison between key features and benefits, and requirements of the Basic Snapshot, Standard, and Advanced Longitudinal upgrade options.

Local Data Upgrade - Key Features

	Basic Snapshot Upgrade	Standard Upgrade	Advanced Longitudinal Upgrade
Time period of dataset (longitudinal or snapshot)	Single snapshot time point	3+ years	7+ years
Granularity (Datapoints available across the time period of the dataset)	N/A	Monthly	Daily
6 core segments available	Yes	Yes	Yes
Conditions (subsegments) available and upgraded with local data	Some conditions upgraded with local data (depending on local data availability)	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	Yes
Ability to perform types of analysis only possible on longitudinal data described in Appendix A (eg trend analysis, impact evaluation)	No	Yes (with caveats depending on on completeness of local data used)	Yes
Accounts for population movement into and out of area, births and deaths	N/A	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	No	Yes
Ability to perform condition prevalence (total population with condition) calculations	Yes - only snapshot	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	No	Yes (with caveats depending on completeness of local data used)	Yes

Local Data Upgrade - Requirements

	Basic Snapshot Upgrade	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)	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	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 at a single point in time for the current population	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	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 Segmentation Dataset	Agreements in place for the local data patient-level condition registers to be integrated/linked with the National Segmentation Dataset	Agreements in place for OBH to act as Processors for GP data integration with the National Segmentation Dataset

B. Local Segmentation Dataset

The Local Segmentation Dataset:

- features a 3 year longitudinal period of data initially, with daily granularity, which increases prospectively as and when it is refreshed regularly with the latest data. The segments (and subsegments) can be configured for 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.

Key features: All population health analysis that can be performed on longitudinal data are possible (see Appendix A for full list of types of analysis that can be performed).

Technical Implementation: **Local data sources from GP practices** are extracted based on OBH extract specifications which conform with data minimisation principles to only extract data required for the specific purpose. All source data is tested and validated prior to integration. **Integration and linking** of data from GP practices and SUS with the application of business and clinical logic as part of OBH's Segmentation Engine is performed by OBH to produce the Segmentation Dataset. **Validation and testing** of the Segmentation Dataset is performed by OBH against QOF data and other publicly available reference benchmarks to ensure quality and accuracy prior to release of the data back to the Client's own infrastructure. Information governance support is provided.

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 Segmentation Dataset	Selected outcomes from OBH Outcomes Data Library
● Upgraded - Basic Snapshot	Limited to National Segmentation Dataset offer
● 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

Example outcomes data available from the OBH Outcomes Data Library

The following is an example list of outcomes available for selection. These outcomes can typically be applied to the entire population, or to a specific and relevant segment or subsegment of the population by linking the outcomes data to the Segmentation Dataset.

Outcome Type	Example Outcome 'Categories' from OBH's Outcomes Data Library	Example outcomes once linked to the Segmentation Dataset
Mortality related outcomes	- Overall mortality rate - Premature mortality - Potential years of life lost - Average age of death - Mortality following admission for a specific condition - One year survival from cancer or a specific type of cancer	- Potential years of life lost in people with serious mental illness
Risk factor prevalence	- Smoking - Obesity	- Smoking prevalence in the healthy/generally well population
Activity-related outcomes	- Days disrupted by care - Readmission within 30 days of discharge - Emergency admission - Time spent at home	- Days disrupted by care in people with frailty or multiple long term conditions - Emergency admissions in people 18+ years who have a respiratory condition
Cohort-specific complication or exacerbation outcomes	- Stroke, heart attack, amputation, blindness, ESRF in diabetes - Acute exacerbation of COPD - Acute exacerbation of heart failure - Delirium in people with frailty or dementia	
End phases of life outcomes	- People dying in preferred place of death - Proportion of those expected to die who are on the palliative care register	- People with frailty or dementia dying in preferred place of death

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
- Common segment-specific diagnoses and procedures
- 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 ICS 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 14 offer due to its complexity and additional implementation requirements.

Appendix

A. Example types of analyses that can be performed on the longitudinal Segmentation Dataset

Types of analyses that can only specifically be undertaken with longitudinal retrospective population, prevalence and incidence data such as the Segmentation Dataset:

1. Trend analysis: Longitudinal data allows for the understanding of incidence trends over time (including retrospectively), identifying patterns, changes, and potential inflection points that may be associated with specific events, interventions, or policy changes. Snapshot data only provides a single point estimate, limiting the ability to assess trends, and only enables prospective analyses to be undertaken, which increases the time needed to evaluate whether interventions are having an impact, or not.

2. Risk factor analysis: With longitudinal data, analysts can investigate the associations between potential risk factors (either risk factors in the Segmentation Dataset or from datasets linked to the Segmentation Dataset) and the incidence of a condition over time, potentially identifying causal relationships, associations, or effect modifiers. Snapshot data limits the ability to establish temporality and causality.

3. Subgroup analysis: Longitudinal data which includes additional variables such as socio-demographic data (age, sex, ethnicity and deprivation) allows for stratified analyses to identify differences in prevalence and incidence trends across subgroups over time.

4. Impact evaluation: Either using the standalone Segmentation Dataset by identifying features of a cohort that defines a particular intervention, or combining longitudinal incidence data with data on people who have had an intervention, analysts can evaluate the impact of these interventions on incidence rates, both in the short-term and long-term. Evaluations can be used to decide which interventions to invest and continue with, or continue to roll-out. For example, the Segmentation Dataset may be joined to activity and expenditure data to evaluate trends associated with specific characteristics and features, and to support business cases and return on investment analyses. See Appendix B for a list of types of interventions that could be evaluated.

5. Life course epidemiology: Longitudinal data spanning multiple life stages can provide insights into the influence of early-life exposures, critical periods, and accumulation of risk factors on the incidence of conditions later in life.

6. Time-series analysis: With longitudinal data, various time-series methods can be applied to model and forecast future incidence rates based on historical patterns and identify any cyclical or seasonal components. For example, using the clinically curated features in the dataset, machine learning models can be used on longitudinal incidence and prevalence data to predict or model changes in future population health.

7. Age-period-cohort analysis: This type of analysis disentangles the effects of age, time period, and birth cohort on the incidence of a condition, allowing analysts to understand the relative contributions of these factors to the observed trends. This analysis requires long-term longitudinal data spanning multiple cohorts and time periods.

8. Survival analysis: For certain conditions, longitudinal incidence data can be used to estimate survival rates or disease progression over time using techniques such as Kaplan-Meier curves or Cox proportional hazards models. This analysis requires information on the timing of disease onset and follow-up data on outcomes which are either available in the Segmentation Dataset, or data that can be linked to the Segmentation Dataset.

9. Sophisticated risk stratification: Longitudinal data allows for case finding of "increasing risk" individuals, e.g. those at risk of experiencing future health deterioration e.g. movement between segments, acquiring certain conditions, multimorbidity progression, increasing frequency of emergency hospital admissions, long stays,

emergency department attendances, exacerbations. Risk of poor outcomes can incorporate both a person's current characteristics, and their recent health trajectory, to refine risk predictions.

B. Example types of interventions that can be evaluated with the Segmentation Dataset (a retrospective longitudinal dataset)

- Medical interventions (pharmaceutical, surgical, diagnostic)
- Public health interventions (prevention, vaccination or screening programmes)
- Health system interventions (access to care, care coordination, quality improvement initiatives)
- Lifestyle interventions
- Supportive interventions (patient education, psychological support)
- Environmental interventions

Evaluation of interventions may require linkage of another dataset to the Segmentation Dataset. Alternatively, if a discrete intervention has been applied to a precise known cohort based on features at a fixed time, it could potentially be evaluated retrospectively in the absence of this. A number of high impact peer-reviewed publications have demonstrated the unique ability of the Segmentation Dataset to support evaluation of national and/or local interventions such as these.