# DataSAT assessment template

### Research question

Add the research question here.

### Data provenance

Please see [recommendations for reporting data provenance](https://www.nice.org.uk/corporate/ecd9/chapter/assessing-data-suitability#data-provenance).

|  |  |
| --- | --- |
| Item | Response |
| Data sources | For each contributing data source provide the name, version and date of data cut. Provide links to their websites, if available. |
| Data linkage and data pooling | Report which datasets were linked, how these were linked, and performance characteristics of the linkage. Note whether linkage was done by a third party (such as NHS Digital).  Clearly describe which data sources were pooled. |
| Type of data source | Describe the types of data source (for example, electronic health record, registry, audit, survey). |
| Purpose of data collection | Describe the main purpose of data collection (for example, clinical care, reimbursement, device safety, research study) |
| Data collection | Describe the main types of data collected (for example, clinical diagnoses, prescriptions, procedures, patient experience data), how data was recorded (for example, clinical coding systems, free text, remote monitoring, survey response), and who collects the data (for example, healthcare professional, self-reported, digital health technology). If the nature of data collection has changed during the data period (for instance, change in coding system or practices, data capture systems) describe the changes clearly. Any differences between data providers in how and what data were collected and its quality should be described.  If additional data collection was done for a research study please describe, including how the validity and consistency of data collection was assured (for example, training). |
| Care setting | State the setting of care for each dataset used (for example, primary care, secondary care, specialist health centres, social services, home use [for wearable devices, or self-reported data on apps or websites]). |
| Geographical setting | State the geographical coverage of the data sources. |
| Population coverage | State how much of the target population is represented by the dataset (for example, population representativeness or patient accrual). |
| Time period of data | State the time period covered by the data. |
| Data preparation | Provide details of whether raw data were accessed for analysis, or whether the data owner had undertaken any data preparation steps such as cleansing or transformation. Mention whether centralised transformation to a common data model was undertaken. Include links to any relevant information including common data model type and version number and details of mapping.  Full details of data preparation specific to addressing the research question is covered in the [section on reporting on data curation](https://www.nice.org.uk/corporate/ecd9/chapter/conduct-of-quantitative-real-world-evidence-studies#reporting-on-data-curation-and-analysis). |
| Data governance | Provide the details of the data controller and funding for each source. Describe the information governance processes for data access and use. |
| Data specification | Note whether a data specification document is available. This may include a data model, [data dictionary](https://www.nice.org.uk/Glossary?letter=D#Data%20dictionary), or both. |
| Data management plan and quality assurance methods | Note whether a data management plan, documentation of source quality assurance methods is available with links to relevant documents. |
| Other documents | Note whether any other documentation is available. Provide hyperlinks or citations to key publications, if available.  If the dataset is available from the HDRUK innovation gateway, provide the hyperlink to its profile on the HDRUK website. |

### Data quality

Details of data quality should be provided for key study variables including population eligibility criteria, outcomes, interventions or exposures, and covariates.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Study variable | Target concept | Operational definition | Quality dimension | How assessed | Assessment result |
| What type of variable (for example, population eligibility, outcome) | Define the target concept (for example, myocardial infarction [MI]) | Define operational definition. For example, MI defined by an ICD-10 code of I21 in the primary diagnosis position | Choose: accuracy or completeness | Describe how quality was assessed. Provide reference to previous validation studies if applicable. | Provide quantitative assessment of quality if available. For example, ‘positive predictive value 85% (75% to 95%)’ |

### Data relevance

Please see [recommendations for reporting data relevance](https://www.nice.org.uk/corporate/ecd9/chapter/assessing-data-suitability#data-relevance).

|  |  |
| --- | --- |
| Item | Response |
| Population | Describe the extent to which the analytical sample reflects the target population. This should consider any data exclusions (for example, because of missing data on key prognostic variables). |
| Care setting | Describe how well the care settings reflect routine care in the NHS. |
| Treatment pathway | Describe how the treatment pathways experienced by people in the data reflects routine care pathways in the NHS (including any diagnostic tests). |
| Availability of key study elements | Note how the dataset met the requirements of the research question in terms of availability of the necessary data variables including key population eligibility criteria, outcomes, intervention and covariates (including confounders and effect modifiers). |
| Study period | State the extent to which the time period covered by the data provides relevant information to decisions. This should cover any important changes to care pathways (including tests) or background changes in outcome rates. |
| Timing of measurements | Describe whether the timing of measurements meet the needs of the research question. |
| Follow up | Note how the follow-up period available in the dataset is sufficient for assessing the outcomes. |
| Sample size | Provide the sample size of the target population in the dataset and demonstrate that it is adequate to generate robust results. |