# Appendices – Real-world evidence framework

Contents

Appendix 1 – Data Suitability Assessment Tool (DataSAT) 2

Appendix 2 – Reporting on methods used to minimise risk of bias 12

Appendix 3 – Reporting information for selected analytical methods 18

# Appendix 1 – Data Suitability Assessment Tool (DataSAT)

### Research question

Add the research question here.

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.  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 data curation. |
| Data sharing | Describe how others can access data. |
| 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, 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.

| 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. |

## DataSAT - case study

Please note that the reporting for this case study is based on publicly available information in [Wing et al. 2021](https://www.journalslibrary.nihr.ac.uk/hta/hta25510#/abs1-1).

### Research question

What is the effect of the long-acting beta-2 agonist and inhaled corticosteroid combination product fluticasone propionate plus salmeterol compared with no exposure or exposure to salmeterol only in people with COPD?

Data provenance

| Item | Response |
| --- | --- |
| Data sources | Clinical Practice Research Datalink (CPRD) GOLD: <https://cprd.com/primary-care>  Hospital episode statistics (HES) Admitted Patient Care data: <https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics> |
| Data linkage and data pooling | CPRD and HES are linked. Patients are identified in a centralised linkage algorithm done by NHS digital. This uses an 8-step deterministic linkage algorithm based on 4 identifiers: NHS number, sex, date of birth and postcode.  Linkage to HES data is possible for 75% of enrolled patients.  See page 6 in: <https://www.cprd.com/sites/default/files/Documentation_HES_APC_set21_v2.7.pdf> |
| Type of data source | HES = administrative records  CPRD = electronic health records |
| Purpose of data collection | Hospital Episode Statistics (HES) is derived from the Secondary Uses Service (SUS) data based on information submitted to NHS digital by healthcare providers. Data collection is primarily intended to support the reimbursement of hospitals for the provision of services in England.  CPRD collects anonymised patient data from a network of GP practices across the UK. Initially this data is collected during a patient’s time in primary care services. |
| Data collection | CPRD = demographics, clinical diagnoses (READ v2 or SNOMED-CT), tests (medcode or SNOMED-CT), prescriptions (prodcode) including dose, route of administration and duration. CPRD GOLD collects fully-coded patient electronic health records from GP practices using the Vision® software system. Data are recorded by health and care staff working within the vision software.  HES = diagnoses (ICD-10), procedures (OPCS-4), admission, discharge, type of care, basic demographics. HES data are collected during a patient's time at hospital and may be recorded during their interactions with health and care staff in the hospital and assembled by teams of clinical coders. |
| Care setting | HES = secondary care  CPRD = primary care |
| Geographical setting | HES = England  CPRD = a representative sample of UK general practices using Vision software. HES-linked CPRD data is available for England only. |
| Population coverage | CPRD GOLD has data for about 3 million currently registered people (around 4.74% of UK population). <https://www.cprd.com/sites/default/files/2021-06%20CPRD%20GOLD%20Release%20Notes.pdf>  HES data covers all NHS Clinical Commissioning Groups in England. |
| Time period of data | The CPRD-linked HES dataset covers from January 2000 to January 2017. |
| Data preparation | No details available for CPRD. However, general practices are included only after demonstrating their records are of research quality.  HES applies centralised processing before the data are released for research:  The rules that run during the processing of the HES data set. These are in place to improve the value and quality of the data and include rules that validate the data within certain fields, derive additional fields and values, remove records that are invalid or out of scope for the HES data set. |
| Data sharing | Details on accessing data are available from the [Clinical Practice Research Datalink](https://www.cprd.com/Data-access) |
| Data governance | CPRD is a centre of the MHRA, which is an executive agency of the Department of Health & Social Care (DHSC). DHSC is therefore the data controller for CPRD data.  HES data is controlled by the Health and Social Care Information Centre (also known as NHS Digital).  CPRD has received funding from the MHRA, Wellcome Trust, Medical Research Council, NIHR Health Technology Assessment programme, Innovative Medicines Initiative, UK Department of Health, Technology Strategy Board, Seventh Framework Programme EU, and various universities, contract research organisations and pharmaceutical companies.  HES data collection is mandated and funded by the UK Government.  Data protection and processing notice for CPRD: <https://www.cprd.com/data-protection-and-processing-notice>  Hospital episode statistics GDPR webpage: <https://digital.nhs.uk/about-nhs-digital/our-work/keeping-patient-data-safe/gdpr/gdpr-register/hospital-episode-statistics-gdpr> |
| Data specification | Fields in HES are derived from the NHS data model and the NHS data dictionary  <https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/hospital-episode-statistics-data-dictionary>  <https://www.datadictionary.nhs.uk/index.html>  CPRD data specification document (May 2021) <https://cprd.com/sites/default/files/CPRD%20GOLD%20Full%20Data%20Specification%20v2.4.pdf> |
| Data management plan and quality assurance methods | HES undertakes processing and data quality checks: <https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/the-processing-cycle-and-hes-data-quality>  No data quality assurance information was identified for CPRD GOLD. However, records from individual general practices are assessed and only included in CPRD after being deemed of research quality. |
| Other documents | None. |

Data quality

| Study variable | Target concept | Operational definition | Quality dimension | How assessed | Assessment result |
| --- | --- | --- | --- | --- | --- |
| Population | COPD | CPRD diagnostic (Read v2) codes for COPD (see codelist in supplementary material of [Quint et al. 2014](https://bmjopen.bmj.com/content/4/7/e005540)) | Accuracy | Previously published validation study comparing algorithms for identifying people with COPD with physician review questionnaire as gold standard ([Quint et al. 2014](https://pubmed.ncbi.nlm.nih.gov/25056980/)) | PPV: 87% (95% CI 78% to 92%) |
| Population | Disease severity | GOLD stage derived from spirometry measurements (see [codelist](https://datacompass.lshtm.ac.uk/id/eprint/1655/)) | Completeness | Proportion of patients with missing spirometry data | 20% |
| Intervention | Fluticasone propionate + salmeterol | CPRD prescribing record matching definition of drug treatment determined by [codelist](https://datacompass.lshtm.ac.uk/id/eprint/1655/) | Accuracy | CPRD prescribing data is expected to be highly accurate | n/a |
| Outcome | COPD exacerbation | Any of the following:   * + CPRD diagnostic (Read) code for LRTI or AECOPD   + A prescription of a COPD-specific antibiotic combined with OCS for 5–14 days   + A record (Read code) of 2 or more respiratory symptoms of AECOPD with a prescription of COPD-specific antibiotics and/or OCS on the same day.   See [codelist](https://doi.org/10.1371/journal.pone.0151357) | Accuracy | Previously published validation study comparing algorithms for identifying people with COPD exacerbations with physician review questionnaire as gold standard ([Rothnie et al. 2016](https://pubmed.ncbi.nlm.nih.gov/26959820/)) | PPV: 86% (95% CI 83% to 88%)  Sensitivity: 63% (95% CI 55% to 70%) |
| Outcome | All-cause mortality | Record in ONS mortality statistics (centrally linked to CPRD data) | Accuracy | ONS mortality records are the gold standard data for deaths | n/a |
| Covariate (confounder) | Alcohol intake | Reported directly in CPRD (closest to index date) | Completeness | Proportion of patients with missing data on alcohol intake | 30% |

Data relevance

| Item | Response |
| --- | --- |
| Population | Patients in CPRD have similar demographic characteristics to the wider UK population. Results from CPRD are generally expected to generalise to the wider eligible population.  Complete records analysis was done excluding records with missing data on socioeconomic status, alcohol consumption and BMI. All these variables had less than 5% of the data missing.  Around one-fifth of patients were excluded because they did not have spirometry measurements recorded in the CPRD. Those without measurements tend to have less contact with health services, which could impact on the generalisability of results. |
| Care setting | Appropriate. COPD drugs are typically administered in primary care (CPRD) while relevant events may be observed in primary or secondary care (CPRD or HES). |
| Treatment pathway | The data represents routine practice in the NHS. |
| Availability of key study elements | Sufficient data on exposures and outcomes are available. Although only prescribing and not dispensing data is available from CPRD this is expected to be a good proxy for dispensing.  No information was available on negative reversibility spirometry results which may be a key confounder.  Dosage information is limited in CPRD. |
| Study period | There have been no major changes to UK clinical practice for the management of COPD since the study period. |
| Timing of measurements | The longitudinal nature of the analysis allows for the research question to be answered. The date of entry is expected to reflect the actual timing of clinical events well. |
| Follow up | The average follow up of 2 years is sufficient for the primary outcome of COPD exacerbations to have occurred. |
| Sample size | The needed sample size for COPD exacerbations was estimated to be 600 per arm at 80% and 5% significance (see [Wing et al. 2021](https://www.journalslibrary.nihr.ac.uk/hta/hta25510#/abs1-1) for details). The actual sample size of about 2,500 per arm far exceeds this. |

# Appendix 2 – Reporting on methods used to minimise risk of bias

Form for reporting on methods used to minimise risk of bias

| Type of bias | How bias was addressed or assessed |
| --- | --- |
| Selection bias at study entry | Selection bias at study entry can arise for several reasons including selection of patients based on eligibility criteria related to the exposure and outcome, or from deviations between time zero, the date the patient meets eligibility criteria, and the date treatment is assigned. Common types of time-related bias are prevalent-user bias, lead time bias, immortal time bias and depletion of susceptibles. Discuss the potential for selection bias at study entry and how this was addressed or investigated through study design, statistical analysis or sensitivity analysis. |
| Selection bias at study exit | A common cause of selection bias because of how individuals exit a study is informative censoring. This may be because of loss to follow up or the occurrence of censoring events. Discuss the possibility of informative censoring and how this was addressed in the analysis. |
| Addressing confounding | Describe the risk of confounding from unmeasured (or unknown) confounders, poorly measured confounders, or time-varying confounding. This should be informed by a systematic identification of potential confounders, clear causal assumptions including the possibility of time-varying confounding, and differences in baseline characteristics between comparison groups. Show how you dealt with any identified risk of confounding through study design (such as selection of a suitable active comparator) and analysis (using an appropriate statistical model, accounting for time-varying confounding). If possible, provide empirical data on the balance of baseline characteristics after adjustment.  If concerns remain about residual confounding, show its impact on results has been assessed using sensitivity or bias analysis.  Confirm that no covariates were inappropriately adjusted to induce bias. For example, show that no covariates on the causal pathway between interventions and outcomes were adjusted for (overadjustment). This may result from the use of covariates measured after the index date. Avoid adjustment for colliders or instruments. This can be informed by causal diagrams. |
| Detection bias | Describe the potential for detection bias resulting from differences in healthcare practices across comparison groups (for example, because of differential frequency or intensity of follow up, or different tests) or length of follow up.  Describe how these have been dealt with through study design (for example, use of comparator with similar follow up) or analysis (for example, adjustment for healthcare use before index date). |
| Measurement error and misclassification | Describe the potential for bias from measurement error or misclassification (this should be formed by assessment of data suitability). Consider which variables are inaccurate, whether this is random or systematic, and how it differs across comparison groups.  Show you addressed risks of bias through statistical analysis (for example, by incorporating external data or calibration) or assessed its impact on results using sensitivity or bias analysis. |
| Missing data | Describe the potential for bias from missing data (this should be formed by assessment of data suitability). Consider which variables have missing data, whether this is random or systematic, and how it differs across comparison groups.  Show how you have addressed risks of bias using statistical methods (such as multiple imputation) and demonstrating their validity. If missingness may not be explainable by observed variables or has unknown mechanisms, sensitivity or bias analysis can be used to explore the impact of different missing ‘not at random’ assumptions. |
| Reverse causation | Describe the risk of reverse causation between the intervention and the outcome arising from causal relationships between variables, time lag between recording of data on interventions and outcomes, or care pathways. Describe how risk of reverse causation was addressed through study design (for example, induction periods or longitudinal follow up), analysis (for example, instrumental variables), or assessed through sensitivity analysis. |

## Methods reporting – case study 1

Please note that the reporting for this case study is based on publicly available information in [Fu et al. 2021](https://pubmed.ncbi.nlm.nih.gov/34844936/).

The study assesses the impact of initiating dialysis at different estimated glomerular filtration rates (eGFR) on cardiovascular events and survival in people with advanced chronic kidney disease. The study used data from the Swedish Renal Registry.

Example of completed methods reporting tool based on [Fu et al. 2021](https://pubmed.ncbi.nlm.nih.gov/34844936/).

| Type of bias | How bias was addressed or assessed |
| --- | --- |
| Selection bias at study entry | Previous observational studies of the effects of the timing of dialysis initiation are at high risk of lead time and immortal time bias resulting from non-alignment of the time at which eligibility criteria were met, treatment assignment, and start of follow-up. The study emulated a target trial informed by the IDEAL trial. To avoid issues with misspecification of time zero, the study used the cloning, censoring, and weighting method. Patients are cloned and assigned to each treatment according to eGFR (one of 15 treatment strategies in the base case) and are censored once they deviated from a given treatment strategy. The approach was validated by replicating results from the IDEAL trial over the range of EGR values seen in the trial.  Selection bias due to the choice of population was not an issue in this population-based study. |
| Selection bias at study exit | Selection bias can be induced by the censoring when patients stop adhering to the ‘treatment strategy’ if this is related to patient characteristics. Inverse probability of censoring weights were estimated using baseline and time-varying confounders to address censoring-induced selection bias.  Loss to follow-up is very low. |
| Addressing confounding | The outcome model adjusted for baseline measurements including demographics, laboratory measurements, prior treatment and hospitalisations. Time-varying confounders were adjusted for in censoring weights including current and previous measurements of eGFR.  Data was not available on other potentially important confounders including muscle mass stores, uraemic symptoms, volume status, quality of life, or physical activity, and data was only available for subset of the cohort on urine albumin-creatinine ratio and plasma potassium. To assess the possibility of residual confounding, the study did the following sensitivity analyses:  • adjusted for urine albumin-creatinine ratio and plasma potassium in the subset of patients with measurements and observed no impact on results  • replicated the results of the IDEAL trial over the eGFR separation observed in the trial. |
| Detection bias | Outcomes included five-year all-cause mortality and major adverse cardiovascular events (composite of cardiovascular death, non-fatal myocardial infarction, or non-fatal stroke). These are likely to be accurately observed regardless of small differences in level of surveillance, for example, resulting from earlier dialysis treatment. |
| Measurement error and misclassification | Timeliness and accuracy of variables extracted from the Swedish Renal Registry have previously been demonstrated. In particular, cardiovascular comorbidities have a very high positive predictive value, generally between 85-95%.  eGFR was calculated with the Chronic Kidney Disease Epidemiology equation from routine plasma creatinine measurements. This has been shown to be accurate to within 30% of measured glomerular filtration rate 85% of the time. |
| Missing data | Data on initiation of dialysis and key outcomes are thought to be complete. Data on mandatory items such as eGFR is also very high.  For non-mandatory data items in the registry, missingness was greater. For example, body mass index was missing in 26% of patients, urinary albumin to creatinine ratio (UACR) in 44%, and potassium in 29%. This was assumed to be missing completely at random and determined by the preferences of the attending physician. Sensitivity analysis in the subset of people with data available showed had no impact on results. |
| Reverse causation | Reverse causation is not expected to be a problem in this analysis. |

## Methods reporting – case study 2

Please note that the reporting for this case study is based on publicly available information in [Wilkinson et al. 2021](https://pubmed.ncbi.nlm.nih.gov/34618040/).

The study estimates the comparative effectiveness of alectinib versus ceritinib on survival in people with ALK-positive non-small cell lung cancer. The study uses real-world data on ceritinib from Flatiron Health to form an external control to patients receiving alecitnib in phase 2 trials.

Example of completed methods reporting tool based on [Wilkinson et al. 2021](https://pubmed.ncbi.nlm.nih.gov/34618040/).

| Type of bias | How bias was addressed or assessed |
| --- | --- |
| Selection bias at study entry | The study compared people enrolled in phase 2 trials assigned alectinib against patients from routine care in the US initiating ceritinib. Several steps were taken to minimise the risk of selection bias:  Matching inclusion criteria in the real-world data to the population included in the trial   * Excluding additional patients from the trial with prior lines of therapy not observed in the real-world data * Using real-world data over a similar time-period to that covered in the trial * Using a new-user, active comparator design   To help demonstrate the validity of the approach, the comparison was repeated using only real-world data and similar results were found. |
| Selection bias at study exit | This was an as-started analysis with limited loss to follow-up. Censoring is not thought to be informative. |
| Addressing confounding | Key prognostic variables were prospectively identified by a systematic review.  Key known confounders were captured in the data albeit with limitations. See below for information on addressing missing data and misclassification of key confounders.  Observed confounders measured at or before baseline were used to estimate propensity scores. Estimation used the inverse probability of treatment weights method. There was no evidence of large differences in covariate patterns between treatment groups after adjustment (standardised mean difference was less than 0.1 for all variables).  In sensitivity analysis, adjustment for additional variables did not change results.  Quantitative bias analysis was used to assess how strong a confounding effect an unknown confounder would need to have to eliminate the estimated treatment effect. The estimated e-value was 2.4 which would require a level of confounder-mortality and confounder-treatment association substantially higher than seen for any measured confounders. |
| Detection bias | The outcome of mortality was not thought to be subject to detection bias. |
| Measurement error and misclassification | Data on mortality is sufficiently well captured in the real-world data with sensitivity of 91% and specificity of 96%.  There were concerns that central nervous system metastases were misclassified (underreported) in the real-world data due to limited surveillance. A sensitivity analysis found that the prevalence in the real-world data would have to be 40% larger to eliminate the estimated treatment effect. |
| Missing data | Missing data on baseline performance status (ECOG score) was high in the real-world data (32%) and this is a key prognostic variable. The main analysis assumed data was missing completely at random in a complete case analysis.  Because this assumption was expected to be invalid, sensitivity analysis was performed using multiple imputation assuming data was missing at random. Results were consistent with the complete case analysis.  Quantitative bias analysis was performed to address remaining concerns about missing not at random data, where ECOG scores are worse than expected by the imputation model. Using threshold analysis the study conclusions remained similar under any reasonable assumptions about the ECOG scores in those with missing values. |
| Reverse causation | Reverse causation is not expected to be a problem in this analysis. |

# Appendix 3 – Reporting information for selected analytical methods

Guide to reporting on selected analytical methods

| Method | Description | Reporting information |
| --- | --- | --- |
| Direct or indirect standardisation | Methods to increase comparability of exposure groups in terms of selected covariates | * Standard reference population (description) * Covariates used for standardisation |
| Stratification | Restricting comparison to select population groups | * How weights are calculated * Whether and how weights are trimmed, truncated or stabilised * Tables for unweighted and weighted population characteristics * Mean and distribution of weights |
| Matching | Matching individuals with the same or similar characteristics | * Covariates used for matching * Matching ratio * Number matched and number excluded |
| Propensity score (general) | Estimate of probability of receiving a particular intervention; range of methods available (below) | * Model used to estimate propensity scores (such as logistic or multinomial) * Covariates used and how they were included in the model * Propensity score distribution before and after adjustments (for example, pre and post-matching) * N/% contributing to matched, trimmed, truncated or weighted analyses * Diagnostic checks for any statistical analysis done * See [Tazare et al. 2022](https://pubmed.ncbi.nlm.nih.gov/35092316/) for reporting of high-dimensional propensity score models |
| Propensity score (stratification) | Patients grouped into strata (for example, deciles) based on propensity score and stratum-specific effects aggregated | * How strata are defined * Trimming and whether applied before or after strata defined * Tables for stratified population characteristics |
| Propensity score (weighting) | Weights attached to individuals based on inverse of propensity scores | * How weights are calculated * Whether and how weights are trimmed, truncated or stabilised * Tables for unweighted and weighted population characteristics * Mean and distribution of weights |
| Propensity score (matching) | Matches individuals with similar propensity scores | * Matching algorithm used including caliper and scale * Matching ratio (such as fixed 1:1 or variable 1:5) * Tables for unmatched and matched population characteristics |
| Multivariable regression adjustment (includes using propensity scores) | Statistical models comparing outcomes as a function of the intervention and covariates | * Type of model (such as linear regression or Poisson) * Covariates used and how they were included * Diagnostic checks |
| Instrumental variable regression | Use of an exogenous source of variation that influences exposure but with no independent effect on the outcome as an instrument | * Model used (for example, 2-stage least squares) * Diagnostic checks * Table with distribution of population characteristics across levels of instrument * Table with distribution of outcomes across levels of instruments * Strength of association between instrument and exposure * Results of falsification tests |