Checklist of items that should be included in reports of mIDA studies

|  |  |  |  |
| --- | --- | --- | --- |
|  | Item No | Recommendation | Page  No |
| **Study ID** | 1 | Indicate the study identifier using either PubMed ID or DOI |  |
| Objectives | | | |
| Background/rationale | 2 | Explain the scientific background and rationale for the study being reported in one or two sentences |  |
| Prespecified hypotheses: | 3 | State prespecified hypotheses in on or two sentences |  |
| Study design: data sources selection & variables selection & data integration | | | |
| Data source | 4a | Describe the time coverage |  |
| 4b | Describe the geographic coverage |  |
| 4c | Describe the sample size |  |
| 4d | Describe the demographic distribution |  |
| 4e | Describe the Cohort criteria |  |
| Study design dependent variables | 5a | State the variable type (e.g., primary outcome variable, secondary outcome variable) |  |
| 5b | State the data source of dependent variable |  |
| 5c | State the data type (e.g., numerical, categorical) of dependent variable |  |
| 5d | State descriptive statistics (e.g., min, max. median, value range, percentile) of dependent variable |  |
| 5e | State the NIMHD domain and levels of dependent variable |  |
| Study design independent variable | 6a | State the variable type (e.g., primary predictor, secondary predictor) |  |
| 6b | State the variables type (e.g., numerical, categorical) of independent variable |  |
| 6c | State the data source of independent variable |  |
| 6d | State descriptive statistics (e.g., min, max. median, value range, percentile) of independent variable |  |
| 6e | State the NIMHD domain and levels of independent variable |  |
| Study design controlled variables | 7a | State the variables type (e.g., numerical, categorical) of controlled variable |  |
| 7b | State the data source of controlled variable |  |
| 7c | State descriptive statistics (e.g., min, max. median, value range, percentile) of controlled variable |  |
| 7d | State the NIMHD domain and levels of controlled variable |  |
| Missing data | 8a | For each data source, describe whether required or expected data that is not present |  |
| 8b | For each variable, describe method of how to handle missing data |  |
| 8c | For each variable, describe the missing rate |  |
| **Data integration** | | | |
| Data processing | 9a | Data extraction: for each variable, describe how to process the raw data source to extract the variable |  |
| 9b | Data cleaning: for each variable, describe the method used to detect and correct (or remove) the incorrect records, missing values or outliers |  |
| Integration strategy | 10 | Describe the integration strategy for each variable:1) Integrate with variables from same level, 2) Integrate with variables from different levels, and 3) Creation of additional computed elements |  |
| Integration algorithms | 11 | For each variable, describe the algorithm used to integrate it with variables from other data sources |  |
| Variable validation | 12 | For each variable, describe data validation rule for the selected variable. Rule should identify both the variable and the validation algorithms |  |