## **Appendix 4 – Systematic Review Database Metadata**

The systematic review database consists of 4 worksheets. Fields that contain “nan” indicate the data is not applicable or available.

Worksheet names and descriptions are given in the table below:

|  |  |
| --- | --- |
| **Worksheet** | **Description** |
| **PC\_Screened\_GRPM** | A list of all articles and models screened for genetic risk prediction of pancreatic cancer. Model and article information is given by the columns listed below. |
| **PC\_Included\_GRPM** | A list of all articles and models included for genetic risk prediction of pancreatic cancer. Model and article information is given by the columns listed below. |
| **PC\_Meta-analysis\_Reviews** | Previous Meta-analysis and reviews identified for pancreatic cancer genetic risk prediction models |
| **PC\_GWAS\_References** | A list of GWAS studies used as references for genetic risk prediction models that use SNPs. See GWAS ID below. |

Column names and their description for **PC\_Screened\_GRPM** and **PC\_Included\_GRPM** are given below:

|  |  |
| --- | --- |
| **Column** | **Description** |
| Paper ID | Unique identifier for article (1, 2, 3 ...) |
| Paper Title | - |
| Journal | - |
| First Author | - |
| Year Published | - |
| Obtained from Literature Review/Meta-analysis | Was the model obtained from a previous literature review or meta-analysis (yes/no) |
| Literature Review/Meta-analysis obtained from | Paper title of literature review/meta-analysis the paper/model was obtained from |
| Year Published | - |
| Screened full text | Was the full text screened (yes/no) |
| Included in final systematic review | Was the model included in the final systematic review? |
| Reason excluded | Exclusion reason (Not outcome of interest/Not features of interest/Insufficient model details) |
| Consortium (s) | Consortium or consortiums the model was built on (UK Biobank, PanScan, …) |
| Inclusion criteria | Cohort selection inclusion criteria |
| Exclusion criteria | Cohort selection exclusion criteria |
| Population size | Total population included in case-control study |
| Cases | Number of cases |
| Controls | Number of controls |
| Cases Percentage Male | - |
| Cases Percentage Female | - |
| Controls Percentage Male | - |
| Controls Percentage Female | - |
| Cases Mean Age | - |
| Cases Standard Deviation | - |
| Controls Mean Age | - |
| Controls Standard Deviation | - |
| Ancestry | Cohort ancestry (European, East Asian, ...) |
| Model ID | Unique identifier for model extracted from article with first number referring to paper ID (2a, 2b) |
| Model Description | Was a risk score calculated? (Risk score and risk prediction/risk prediction) |
| Target | Model target |
| Matching Variables | Case-control matching variables |
| Adjusted Variables | Model adjusted variables |
| Model | Type of model used (Logistic regression) |
| Prediction time interval (Years) | What time range, until diagnosis, was the model built? (0 years suggests all data up to the target’s diagnosis was used in model construction, 5 years suggests data 5 years prior to diagnosis was used to construct the model) |
| Model type | Does the model incorporate genetic and clinical factors? (Genetic/Genetic and clinical) |
| Risk score calculation method | How was the risk score constructed? |
| Feature description | Brief description of the feature used in model construction |
| Model feature | - |
| Risk score features | For cases where a risk score was used as a model feature, what were the components (SNPs or Phenotypes) of the risk score? |
| Weight (Beta) | Weights or coefficients of risk score components |
| GWAS ID | For SNPs used in risk scores or model features, a GWAS ID is given. This refers to the relevant GWAS worksheet and gives the SNP GWAS origin. |
| P-value | Model feature significance |
| OR | Model feature odds ratio calculated from logistic regression coefficient |
| OR 95% LB | 95% confidence interval OR lower bound |
| OR 95% UP | 95% confidence interval OR upper bound |
| OR 5th vs 1st Quintile | Odds ratio of pancreatic cancer risk score, comparing high risk (5th  quintile) vs low risk (1st quintile) participants |
| OR 5th vs 1st Quintile LB | 95% confidence interval OR 5th vs 1st quintile lower bound |
| OR 5th vs 1st Quintile UB | 95% confidence interval OR 5th vs 1st quintile upper bound |
| AUC | Area under receiver operating characteristic curve score |
| AUC 95% LB | Area under receiver operating characteristic curve 95% confidence interval upper bound |
| AUC 95% UP | Area under receiver operating characteristic curve 95% confidence interval upper bound |
| Stated Strengths | Stated strengths of the article |
| Stated Limitations | Stated limitations of the article |

Column names and their description for **PC\_Meta-analysis\_Reviews** are given below:

|  |  |
| --- | --- |
| **Column** | **Description** |
| Paper ID | Unique identifier for article (R1, R2, R3 ...) |
| Paper Title | - |
| Journal | - |
| First Author | - |
| Year Published | - |
| Relevant Papers | Relevant articles obtained from literature review/meta-analysis |

Column names and their description for **PC\_GWAS\_References** are given below:

|  |  |
| --- | --- |
| **Column** | **Description** |
| GWAS ID | Unique identifier for GWAS (G1, G2, G3 ...) |
| Paper Title | - |
| Journal | - |
| First Author | - |
| Year Published | - |