EAVE II Data Dictionary

EAVE II Analysts

22 January, 2021

Contents

1 GP Data Demographics

1.1 Data information

Description	Baseline GP extract on all patients and their key demographics
Datafile	EAVE_demographics.rds
Type	Cleaned
Data sources	GP data
Scripts	NA
Update frequency	Bi-monthly to 3 months
Update timestamps	NA
Analyses used	NA

1.2 Variable information

Name	Label	Description	Type	Data source	Values	Derived variable	Derivation hierarchy
EAVE_LINKNO	EAVE Link Number	Randomly assigned study index number	character	GP data	EAVE(7)digits (e.g. EAVE1234567)	NA	NA
ageYear	Age	Age in years	integer	GP data	Min= 0, Max= 120	NA	NA
Sex	Sex	Sex at birth	character	GP data	F, M	NA	NA
simd2020_sc_quintile	SIMD Quintile	Scottish Index of Multiple Deprivation (SIMD) Quintile (2020)	integer	GP data	1-5	NA	NA
ur6_2016	Urban/Rural classification	Urban/Rural 6-fold classification (2016 classification)	integer	GP data	1-6	NA	NA
ur6_2016_name	Urban/Rural classification name	Urban/Rural 6-fold classification (2016 classification)	character	GP data	1 Large Urban Areas, 2 Other urban Areas, 3 Accessible Small Towns, 4 Remote Small Towns, 5 Accessible Rural, 6 Remote Rural	NA	NA
hb2019	NHS Health Board Code	Codes of 2019 NHS Health Boards	character	GP data	S08000015, S08000016, S08000017, S08000019, S08000020, S08000022, S08000024, S08000025, S08000026, S08000028, S08000029, S08000030, S08000031, S08000032	NA	NA
hb2019name	NHS Health Board Name	Names of 2019 NHS Health Boards	character	GP data	NHS Ayrshire and Arran, NHS Borders, NHS Dumfries and Galloway, NHS Fife, NHS Forth Valley, NHS Grampian, NHS Greater Glasgow and Clyde, NHS Highland, NHS Lanarkshire, NHS Lothian, NHS Orkney, NHS Shetland, NHS Tayside, NHS Western Isles	NA	NA

2 GP Diagnostics

2.1 Data information

Description	Baseline GP extract of all EAVE II risk groups
Datafile	cohort_diags.rds
Type	Cleaned
Data sources	GP data
Scripts	NA
Update frequency	Bi-monthly to 3 months
Update timestamps	2020-09-11, 2020-09-15
Analyses used	NA

2.2 Variable information

Name	Label	Description	Type	Data source	Values	Derived variable	Derivation hierarchy
EAVE_LINKNO	EAVE Link Number	Randomly assigned study index number	character	GP data	EAVE(7)digits (e.g. EAVE1234567)	NA	NA
diag	Diagnosis (GP cluster)	Medical diagnosis, clinical indicators and other health-related characteristics	character	GP data	See Supplementary Information - EAVEII_dd_gp_clusters.xlsx	NA	NA

3 ECOSS

3.1 Data information

Description	Laboratory data on COVID-19 testing
Datafile	ECOSSdeduped_linked.rds
Type	Cleaned
Data sources	ECOSS
Scripts	NA
Update frequency	Weekly to fortnightly
Update timestamps	2020-12-08, 2020-12-15, 2020-12-22, 2020-12-29, 2021-01-05, 2021-01-19
Analyses used	NA

3.2 Variable information

Name	Label	Description	Type	Data source	Values	Derived variable	Derivation hierarchy
EAVE_LINKNO	EAVE Link Number	Randomly assigned study index number	character	ECOSS	EAVE(7)digits (e.g. EAVE1234567)	NA	NA
ECOSSID	ECOSS Link Number	Identifier	integer	ECOSS	8 digits (e.g. 12345678)	NA	NA
Sex	Sex	Sex at birth	character	ECOSS	F, M, U (Unknown)	NA	NA
AgeYear	Age	Age in years	integer	ECOSS	Min= 0, Max= 130	NA	NA
HealthBoard	NHS Health Board	Name abbreviation of NHS Health Boards	character	ECOSS	AA, BR, DG, FF, FV, GC, GR, HG, LN, LO, OR, SH, TY, UK (Uknown), WI	NA	NA
PostCode	Postcode	Postcode	character	ECOSS	7 digits/letters	NA	NA
LabSpecimenNo	Number of laboratory specimen	Number of laboratory specimen	character	ECOSS	11 digits/letters	NA	NA
Specimendate	date of specimen	date of specimen collection	date	ECOSS	yyyy-mm-dd	NA	NA
OriginalOrganism	Original organism	Original organism - Coronavirus 2019 positive or negative	character	ECOSS	Coronavirus 2019-nCoV, Coronavirus 2019-nCoV negative	NA	NA
NCOV_RESULT	COVID-19 result	Novel Coronavirus result	character	ECOSS	Negative, Positive	NA	NA
result	Result	Result	integer	ECOSS	0,1	NA	NA

Name	Label	Description	Type	Data source	Values	Derived variable	Derivation hierarchy
SpecimenOrigin	Origin of specimen	Origin of specimen	character	ECOSS	Community, Drug Service, GP, GUM Clinic, H.M Prison, Hospital, Local Authority EH Dept, NHS- Other, Non-NHS-Other, Occupational Health, Private, Residential Care Homes	NA	NA
EcossdateReceived	ECOSS date	date record received by ECOSS	date	ECOSS	yyyy-mm-dd	NA	NA
SubLab	Sub lab	Sub lab ID	character	ECOSS	ABD:GHB, AYR:ARL, BOR:ARL, DEE:UNI, DMF:ROY, EDI:RVL, FIF:ARL, GIL:BAN, GLA:GJH, GLA:REV, INV:RAI, LAN:HAR, LAN:LAW, LAN:MON, NHS:COV, ORK:BAL, STI:ROY, WES:LES	NA	NA
SourceLab	Source lab	Source lab ID	character	ECOSS	ABD:ALB, ABD:GHB, ABD:ROY, AYR:ARL, BOR:ARL, DEE:UNI, DMF:ROY, DUN:ARL, EDI:ROY, EDI:RVL, EDI:SCH, EDI:UNI, EDI:WGH, FIF:ARL, GIL:BAN, GLA:GAR, GLA:GJH, GLA:NUF, GLA:REV, GLA:ROY, GLA:SCH, GLA:SCS, GLA:SGH, GLA:STR, GLA:UNI, GLA:VIC, GLA:WES, GLA:CLY, INV:RAI, LAN:ALL, LAN:HAR, LAN:LAW, LAN:MON, NHS:UKL, OBA:LAB, ORK:BAL, PAI:RAH, STI:ROY, WES:LES, WLO:BAN	NA	NA
CAREHOMEID	Care home ID	Care home identifier	character	ECOSS	NA	NA	NA
NRS.Reg.date	NRS registration date	National Records of Scotland registration date	date	ECOSS	yyyy-mm-dd	NA	NA
NRS.date.Death	NRS death date	National Records of Scotland death date	date	ECOSS	yyyy-mm-dd	NA	NA
death28	Death within 28 days	Death within 28 days of positive covid test	integer	ECOSS, NRS deaths	0,1	Yes	Derived from daily linkage of ECOSS and NRS deaths within PHS.

4 Health Board Lookup

4.1 Data information

Description	Lookup of NHS Scotland Healthboards and the total population
Datafile	HB2019_pop_lookup.rds
Type	Lookup
Data sources	PHS lookup file
Scripts	NA
Update frequency	NA
Update timestamps	NA
Analyses used	NA

4.2 Variable information

Name	Label	Description	Type	Data source	Values	Derived variable	Derivation hierarchy
year	Year	Year of the lookup file used to determine the Health Boards	integer	Health board lookup	2019	NA	NA
hb2019	Health Board code 2019	Codes of NHS Health Boards in 2019	character	Health board lookup	S08000015, S08000016, S08000017, S08000019, S08000020, S08000022, S08000024, S08000025, S08000026, S08000028, S08000029, S08000030, S08000031, S08000032	NA	NA
hb2019name	Health Board name 2019	Names of NHS Health Boards in 2019	character	Health board lookup	NHS Ayrshire and Arran, NHS Borders, NHS Dumfries and Galloway, NHS Fife, NHS Forth Valley, NHS Grampian, NHS Greater Glasgow and Clyde, NHS Highland, NHS Lanarkshire, NHS Lothian, NHS Orkney, NHS Shetland, NHS Tayside, NHS Western Isles	NA	NA
hb2018	Health Board code 2018	Codes of NHS Health Boards in 2018	character	Health board lookup	S08000015, S08000016, S08000017, S08000019, S08000020, S08000022, S08000024, S08000025, S08000026, S08000028, S08000029, S08000030, S08000031, S08000032	NA	NA
hb2014	Health Board code 2014	Codes of NHS Health Boards in 2014	character	Health board lookup	S08000015, S08000016, S08000017, S08000019, S08000020, S08000022, S08000024, S08000025, S08000026, S08000028, S08000029, S08000030, S08000031, S08000033	NA	NA
age	Age	Age in years	integer	Health board lookup	Min= 0, Max= 90	NA	NA
sex	Sex (number)	Sex at birth	integer	Health board lookup	1, 2	NA	NA
sex_name	Sex	Sex at birth	character	Health board lookup	M, F	NA	NA

Name	Label	Description	Type	Data source	Values	Derived variable	Derivation hierarchy
pop	Population	Number of people by year, Health Board, age and sex	integer	Health board lookup	e.g. 1508	NA	NA

5 Severe cases

5.1 Data information

Description	EAVE II severe cases and outcomes
Datafile	cases_severe_dates.rds
Type	Derived
Data sources	GP data, ECOSS, NRS Deaths, RAPID, SICSAG
Scripts	NA
Update frequency	Weekly to fortnightly
Update timestamps	NA
Analyses used	NA

5.2 Variable information

Name	Label	Description	Type	Data source	Values	Derived variable	Derivation hierarchy
EAVE_LINKNO	EAVE Link Number	Randomly assigned study index number	character	GP data	EAVE(7)digits (e.g. EAVE1234567)	NA	NA
SPECIMENdate	Specimen date	Date of test	date	ECOSS	date in format yyyy-mm-dd	NA	NA
CURRENT_POSTCODE	Postcode	Postcode of residence	character	HPS case-control files	Postcode in pc7 format	NA	NA
is.case	is a covid case	Flag to indicate person is a covid case (+ve test and/or covid on death certificate)	logic	ECOSS, NRS deaths	TRUE/FALSE (should all be true)	NA	NA
DATE_OF_DEATH	Date of death	Date of patients death	date	NRS deaths	date in format yyyy-mm-dd	NA	NA
covid_ucod	COVID is underlying cause of death (COD)	Indicator or whether covid is underlying or primary COD on death certificate	logic	NRS deaths	0,1	NA	NA
covid_cod	Covid listed as any cause of death	Indicator or whether covid is listed as any COD on death certificate	logic	NRS deaths	0,1	NA	NA
dead28	Death within 28 days	Indicator of whether patient has died within 28 days of +ve covid Test	logic	ECOSS, NRS deaths	0,1	Yes	NA

Name	Label	Description	Type	Data source	Values	Derived variable	Derivation hierarchy
inhosp	In hospital when test done	Indicator whether patient was already admitted to hospital at time of test	logic	ECOSS, RAPID	0,1	Yes	NA
icu	ICU admission	Indicator of whether patient was admitted to Intensive Care Unit (ICU)	logic	SICSAG	0,1	NA	NA
hdu	HDU admission	Indicator of whether patient was admitted to High Dependency Unit (HDU)	logic	SICSAG	0,1	NA	NA
RAPID	rapid flag	flag to indicate person in RAPID dataset	logic	RAPID	0,1	Yes	NA
ICU_date	ICU admission date	Date admitted to ICU	date	SICSAG	date in format yyyy-mm-dd	NA	NA
days_to_icu	Days to ICU	Number of days from test to admission to ICU	numeric	ECOSS, SICSAG	any +ve value	Yes	NA
admission.date	Hospital admission date	Date admitted to hospital	date	RAPID	date in format yyyy-mm-dd	NA	NA
days_2_hosp	Days to hospital	Number of days from test to admission to hospital	numeric	ECOSS, RAPID	any +ve value (if testedbefore admission, value changed to 0)	Yes	NA

6   Household identifier

6.1   Data information

Description	EAVE II Household from CHI registry
Datafile	HID_lookup_MarSept2020.rds
Type	Derived
Data sources	CHILI
Scripts	NA
Update frequency	6 months
Update timestamps	NA
Analyses used	NA

6.2   Variable information

Name	Label	Description	Type	Data source	Values	Derived variable	Derivation hierarchy
EAVE_LINKNO	EAVE Link Number	Randomly assigned study index number	character	GP data	EAVE(7)digits (e.g. EAVE1234567)	NA	NA
Date_address_changed.x	Address change date (March)	Date when address changed (March update)	numeric	CHILI	date as number yyyymmdd	NA	NA

Name	Label	Description	Type	Data source	Values	Derived variable	Derivation hierarchy
March_hid	Household ID (March)	Household ID number (March update)	numeric	CHILI	ID	NA	NA
March_hid_count	Number of ID's in household (March)	Number of people in household (March update)	numeric	CHILI	any +ve whole number	NA	NA
Date_address_changed.y	Address change data (September)	Date when address changed (September update)	numeric	CHILI	date as number yyyymmdd	NA	NA
March_transfer_out_date	Move or transfer date (March)	Date when resident moved out of scotland (March update)	date	CHILI	date	NA	NA
Sept_hid	Household ID (September)	Household ID number (September update)	numeric	CHILI	ID	NA	NA
Sept_hid_count	Number of ID's in household (September)	Number of people in household (September update)	numeric	CHILI	any +ve whole number	NA	NA
Sept_transfer_out_date	Move or transfer date (September)	Date when resident moved out of scotland (September update)	date	CHILI	date	NA	NA

7 QCOVID deduplicated

7.1 Data information

Description	EAVE II data for QCOVID validation analysis
Datafile	QCOVIDdeduped.rds
Type	Bespoke
Data sources	GP data, ECOSS, HES, RTDS
Scripts	NA
Update frequency	NA
Update timestamps	NA
Analyses used	NA

7.2 Variable information

Name	Label	Description	Type	Data source	Values	Derived variable	Derivation hierarchy
rowid	Row ID	UID for batch processor file	integer	n/a	n/a	NA	NA
sex	Sex	Sex	logic	GP data or Patient demographic service	0:female; 1:male	NA	NA

Name	Label	Description	Type	Data source	Values	Derived variable	Derivation hierarchy
age	Age	Age	integer	GP data or Patient demographic service	19-100	NA	NA
Q_RX_IMMUNO	Prescribed immunosuppressants indicator	Have you been prescribed immunosuppressants prescribed by your GP	logic	GP data	0:false; 1:true	NA	NA
Q_RX_LABA	Anti-leukotriene or long acting beta2-agonists (LABA) indicator	Are you taking anti-leukotriene or long acting beta2-agonists (LABA)?	logic	GP data	0:false; 1:true	NA	NA
Q_RX_PRED	Oral steroid indicator	Have you been prescribed oral steroids by your GP in the last 6 months?	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_AF	Atrial fibrillation indicator	Do you have atrial fibrillation?	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_CCF	Heart failure indicator	Do you have heart failure?	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_asthma	Asthma indicator	Do you have asthma?	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_blood_cancer	Cancer of the blood or bone marrow and are at any stage of treatment indicator	Have you a cancer of the blood or bone marrow such as leukaemia, myelodysplastic syndromes, lymphoma or myeloma and are at any stage of treatment?	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_cerebralpalsy	Cerebral palsy indicator	Do you have cerebral palsy?	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_chd	Coronary heart disease indicator	Do you have coronary heart disease?	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_cirrhosis	Cirrhosis of the liver indicator	Do you have cirrhosis of the liver?	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_Congen_HD	Congenital heart disease or had surgery for it in the past indicator	Do you have congenital heart disease or have you had surgery for it in the past?	logic	GP data or HES (OPCS)	0:false; 1:true	NA	NA
Q_DIAG_COPD	Chronic obstructive pulmonary disease (COPD) indicator	Do you have chronic obstructive pulmonary disease (COPD)?	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_dementia	Dementia indicator	Do you have dementia?	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_epilepsy	Epilepsy indicator	Do you have epilepsy?	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_fracture	A prior fracture of hip, wrist, spine or humerus indicator	Have you had a prior fracture of hip, wrist, spine or humerus?	logic	GP data	0:false; 1:true	NA	NA

Name	Label	Description	Type	Data source	Values	Derived variable	Derivation hierarchy
Q_DIAG_neurorare	Motor neurone disease, multiple sclerosis, myaesthenia, or Huntingtons's Chorea indicator	Do you have motor neurone disease, multiple sclerosis, myaesthenia, or Huntingtons's Chorea?	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_parkinsons	Parkinson's disease indicator	Do you have Parkinson's disease?	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_pulm_hyper	Pulmonary hypertension or pulmonary fibrosis indicator	Do you have pulmonary hypertension or pulmonary fibrosis?	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_pulm_rare	Cystic fibrosis or bronchiectasis or alveolitis indicator	Do you have cystic fibrosis or bronchiectasis or alveolitis?	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_pvd	Peripheral vascular disease indicator	Do you have peripheral vascular disease?	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_ra_sle	Rheumatoid arthritis or SLE indicator	Do you have rheumatoid arthritis or SLE?	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_respcancer	Lung or oral cancer indicator	Do you have lung or oral cancer?	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_SEV_MENT_ILL	Severe mental illness indicator	Do you have severe mental illness?	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_sickle_cell_	Sickle cell disease or severe combined immune deficiency syndromes indicator	Do you have sickle cell disease or severe combined immune deficiency syndromes?	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_stroke	A stroke or TIA indicator	Have you had a stroke or TIA?	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_DIABETES_1	Diabetes indicator	Type 1 diabetes	logic	GP data	0:none; 1:Type 1; 2:Type 2	NA	NA
Q_DIAG_DIABETES_2	Diabetes indicator	type 2 diabetes	logic	GP data	0:false; 1:true	NA	NA
Q_DIAG_VTE	A thrombosis or pulmonary embolus indicator	Have you had a thrombosis or pulmonary embolus?	logic	GP data	0:false; 1:true	NA	NA
BMI	BMI	Body Mass Index	numeric	GP data	0-40	NA	NA
ethnicity	Ethnic group	What is your ethnic group?	integer	GP data	0-17	NA	NA
homecat	Housing category	what is your housing category - care home or homeless or neither?	integer	GP data	0: neither; 1: care home; 2: homeless	NA	NA
learncat	Learning disability or Down's Syndrome indicator	Do you have a learning disability or Down's Syndrome?	integer	GP data	0: neither; 1: learning disability; 2: Down's	NA	NA

Name	Label	Description	Type	Data source	Values	Derived variable	Derivation hierarchy
p__marrow6	Bone marrow or stem cell transplant in the last 6 months indicator	Have you had a bone marrow or stem cell transplant in the last 6 months?	logic	HES (OPCS)	0:false; 1:true	NA	NA
p__radio6	Radiotherapy in the last 6 months indicator	Have you had radiotherapy in the last 6 months?	logic	Radiotherapy Data Set (RTDS) or HES (OPCS)	0:false; 1:true	NA	NA
p__solidtransplant	Solid organ transplant indicator	Have you had a solid organ transplant (lung, liver, somach, pancreas, spleen, heart or thymus)?	logic	HES	0:false; 1:true	NA	NA
renalcat	Kidney disease indicator	Do you have kidney disease?	integer	GP data, HES (OPCS)	0-5	NA	NA
Postcode	Postcode	What is your postcode?	numeric	QCovid Postcode Table	double (nullable)	NA	NA
town	Town	Townsend score	character	QCovid Postcode Table	string (nullable)	Yes	NA

8   Sgene linked

8.1   Data information

Description	EAVE II data for new COVID-19 variant
Datafile	Sgene_linked.rds
Type	Bespoke
Data sources	GP data, ECOSS, RAPID, NRS deaths
Scripts	NA
Update frequency	Weekly to fortnightly
Update timestamps	2020-12-30, 2021-01-05, 2021-01-14
Analyses used	NA

8.2   Variable information

Name	Label	Description	Type	Data source	Values	Derived variable	Derivation hierarchy
EAVE_LINKNO	EAVE Link Number	Randomly assigned study index number	character	GP data	EAVE(7)digits (e.g. EAVE1234567)	NA	NA
ECOSSID	ECOSS ID	ECOSS ID	numeric	ECOSS	ECOSSID digits	NA	NA
age_year	Age	Age in years	numeric	ECOSS	0-100	NA	NA
sex	Sex	Sex	character	ECOSS	M/F	NA	NA
SpecimenDate	Specimen Date	Date sample taken	date	ECOSS	date in format yyyy-mm-dd	NA	NA
SubLab	Sub Lab	location sample processed	character	ECOSS	Lab ID	NA	NA
result	COVID result	result of covid test	numeric	ECOSS	1 (only +ve tests in this file)	NA	NA

Name	Label	Description	Type	Data source	Values	Derived variable	Derivation hierarchy
NRS.Date.Death	Date of NRS death record	Date of death	date	NRS deaths	date in format yyyy-mm-dd	NA	NA
death28	Death within 28 days of positive test	Died within 28 days of +ve test	logic	ECOSS, NRS deaths	0,1	Yes	NA
simd2020	SIMD Quintile 2020	Quintile of scottish index of multiple deprivation (from postcode of residence)	numeric	ECOSS	1-5	NA	NA
admission_date	Admission to hospital date	Day admitted to hospital	date	RAPID	dates	NA	NA
discharge_date	Discharge from hospital date	Day discharged from hospital	date	RAPID	dates	NA	NA
covid_adm	COVID admission	Flag for covid admission	logic	ECOSS, RAPID	0,1	Yes	NA
time_test_covid_adm	Time from positive test to admission	Time (days) from +ve test to admission	numeric	ECOSS, RAPID	0-14	Yes	NA
covid_dth	COVID death	Covid death (defined as death within 28 days of +ve test ?check if also includes covid on death certificate >28 days	logic	ECOSS, RAPID	0,1	Yes	NA
time_test_covid_dth	Time from poistive test to death	Time (days) from +ve test to death	numeric	ECOSS, RAPID	0-28	Yes	NA
days	Days of...	Days of ?	numeric	ECOSS, RAPID	0+	Yes	NA
true_s_gene_dropout	True S Gene Dropout	Whether S gene positive, S gene negative or weak Sgene positive on PCR test.	character	??ECOSS - need to check with Chris	Positive S Gene/True S Gene Dropout/Weak Positive/Unknown	NA	NA
new	New patient since last linkage	Flag for new compared to last linkage of gene dropouts	logic	ECOSS, RAPID	0,1	Yes	NA

9 Supplementary Information

All supplementary information are stored in the Supplementary GitHub folder. This contains the following:

Name	Description	File
GP Cluster ReadCodes	Links each GP risk group cluster to the GP ReadCodes within the GP IT system	EAVEII_dd_gp_clusters.xlsx
Data source external information	Provides extra information and links on linked data sources	EAVEII_dd_external_info.xlsx